5-11-2023

Excuse Me, Something Is Unfair! - Implications of Perceived Fairness of Service Robots

Sebastian Thomas Büttner
University of Duisburg-Essen, sebastian.buettner.ude@gmail.com

Cora Maria Sourkounis
Leibniz University Hannover, cora.maria.sourkounis@tu-clausthal.de

Jan Christoph Gutzmann
University of Duisburg-Essen, jan.gutzmann.ude@gmail.com

Michael Prilla
University of Duisburg-Essen, michael.prilla@uni-due.de

EXCUSE ME, SOMETHING IS UNFAIR! – IMPLICATIONS OF PERCEIVED FAIRNESS OF SERVICE ROBOTS

Research Paper

Sebastian Thomas Büttner, University of Duisburg-Essen, Germany, sebastian.thomas.buettner@uni-due.de

Cora Maria Sourkounis, Leibniz University Hannover, Germany, sourkounis@match.uni-hannover.de

Jan Christoph Gutzmann, University of Duisburg-Essen, Germany, jan.gutzmann@uni-due.de

Michael Prilla, University of Duisburg-Essen, Germany, michael.prilla@uni-due.de

Abstract

Fairness is an important aspect for individuals and teams. This also applies to human-robot interaction (HRI). Especially if intelligent robots provide services to multiple humans, humans may feel treated unfairly by robots. Most work in this area deals with the aspects of fair algorithms, task allocation, and decision support. This work focuses on a different, yet little-explored perspective, which looks at fairness in HRI from a human-centered perspective in human-robot teams. We present an experiment in which a service robot was responsible for distributing resources among competing team members – in one condition with an efficient (but rather unfair) distribution algorithm and in another condition with an algorithm that was less efficient but could be considered fair. We investigated how the different strategies of distribution influence the perceived fairness and how this perceived fairness influenced the perception of the robot as such. Our study shows that humans might perceive technically efficient algorithms executed by a robot as unfair, especially if humans personally experience negative consequences. Interestingly, this perceived unfairness had a negative impact on human perception of the robot as such, which should be considered in the design of future robots.

Keywords: Human-Robot Interaction, Fairness, Human-Robot Groups, Service Robot, Empirical Study.

1 Introduction

Fairness is an important aspect in human interaction. In cooperation and collaboration between humans, fairness is a decisive element that influences how well people can work together (Colquitt, Zapata-Phelan and Roberson, 2005). Fairness "is essential to a mutually satisfactory exchange between two parties" (Berry and Seiders, 2008) in business settings, it is critical to the work of people in companies as well as for the success of the company (Alexander and Ruderman, 1987), and it is a central element to our society (Colquitt, Zapata-Phelan and Roberson, 2005). The importance of fairness has been shown in research on human-robot teams, in which it was found that humans may have a different perception of fairness towards robots when compared to humans (Torta *et al.*, 2013; Nagataki *et al.*, 2019) and that the perception of unfairness in robotic team members may have influences on team performance (Claure *et al.*, 2020; Chang and Thomaz, 2021) as well as on the relationship between team members (Jung *et al.*, 2020). It has also been found that people want to be treated fairly by robots (Chang and Thomaz, 2021).

There are different reasons why people may perceive robots and other AI-based agents to act unfairly. Only in very few situations, the reason is that the robot was intentionally built to be unfair, e.g. to spy on companies or people (Hartzog, 2014). In most cases, there are other reasons: Fair decision-making of algorithms may be undermined by bad data (Lee, Madotto and Fung, 2019) or algorithms may simply be unfair without bad intentions (Wang, Harper and Zhu, 2020). Even if this is not the case, an agent's behavior may seem unfair if it interprets subtle differences in requests or situations and therefore reacts differently on these (Følstad and Taylor, 2020), or if it provides different service levels to people based on the status or license level of the respective person (Radhakrishnan and Gupta, 2020). Sometimes, unfairness is also mostly a perception of people (Lee and Baykal, 2017). It is common sense that unfairness in robots and other AI-based agents must not occur (Abdul *et al.*, 2018; Keyes, 2018; Holstein *et al.*, 2019), but there is also a lot of work left to achieve this goal.

Besides other areas, the perception of fairness in Human Robot Interaction (HRI) is especially important for service and delivery robots, as the perception of unfair service creates negative reactions in those being served (Seiders and Berry, 1998), and as these negative reactions often linger and impede the potential recovery of satisfaction with the service (Ok, Back and Shanklin, 2006; Siu, Zhang and Yau, 2013). Fairness in service and delivery robots has many practical applications. Consider, for example, a robot serving customers in a restaurant, which is a popular and very recent application area of service robots (Lee, Lin and Shih, 2018; Berezina, Ciftci and Cobanoglu, 2019; Byrd *et al.*, 2021). Customers might expect to be served in the same sequence in which they had placed an order. Service robots, however, may use ideal path planning to deliver them as soon as possible. This may result in the perception of unfair treatment. There are similar scenarios such as the delivery of goods (Lee *et al.*, 2012), robots allocating resources (Claure *et al.*, 2020; Jung *et al.*, 2020), rescue robots (Brandao *et al.*, 2020), robots playing with humans (Short *et al.*, 2010), and others. In all of these scenarios, fairness of robots is an issue.

Fairness in robotics and AI in general may be tackled from different perspectives. As Auernhammer (2020) lays out using a famous differentiation by Winograd (1996), there are two main perspectives in research on human interaction with AI, robotics, and other autonomous machinery: the *rationalistic* perspective, which takes a mainly technical approach and tackles challenges in AI from an algorithmic and data-centered view, and the *design* perspective, which takes a humanistic stance and looks at human interaction with technology, seeking solutions that "improve human conditions" (Auernhammer, 2020). A lot of research is done on the former, the rationalistic perspective. In the context of fairness, this includes work on fairness in algorithms (Celis *et al.*, 2019; Wang, Harper and Zhu, 2020), and on using data that enables fair decisions of AI-based agents (Pastaltzidis *et al.*, 2022; Ruf and Detyniecki, 2022). The design perspective, while not regarded as much as the rationalistic perspective, is as important as focusing on technical aspects, and has gained attention under terms such as "human-centered AI" recently (Riedl, 2019a; Shneiderman, 2022). Besides the focus on human interaction with technology, this perspective also demands a "sociocultural understanding" (Riedl, 2019a) of the context robots and other AI are embedded in. In this paper, we look into fairness in HRI from this human-centered perspective; we present a study in which we look into human perception of fairness in the behavior of service robots.

Our work is driven by the question of which *impact the perceived unfairness of a service robot has on humans in social situations*. For this, we conducted an experiment with 34 participants, in which a service robot distributed resources to three competitors. Our results show that humans perceive robots to be unfair in certain constellations, and this led to changes in other perceptions towards the robot.

2 Related Work

Fairness has been defined as "a global perception of appropriateness" (Colquitt and Rodell, 2015) of how an individual is treated by other individuals or organizations. As such, fairness is subjective, and the individual perception of whether a situation is perceived as fair or unfair has individual (Colquitt *et al.*, 2018) and cultural influences (Blake *et al.*, 2015). Fairness or more specifically responses to fairness or unfairness seem to be deeply rooted in human behavior. Theories relating to fairness often explain the purpose of fairness with the recognition of the value of others in cooperative interactions (Brosnan, 2013). In human-to-human cooperation, it probably is more beneficial to stop cooperation, if the outcomes are regularly distributed unfairly among humans (Fehr and Schmidt, 1999; Brosnan, 2013). Consequently, the perception of unfairness causes strong negative emotions in humans and non-human primates (Brosnan and de Waal, 2014).

According to current theories of fairness, fairness can be divided into four main dimensions. In a larger social context, these are *interactional* and *informational* fairness (Alexander and Ruderman, 1987; Greenberg, 1993), and in the context of organizations, these are *procedural* and *distributive* fairness. Interactional fairness involves how rules of (fair) decision-making are applied in a specific case (Tyler and Bies, 1990). Importantly, the perception of interactional fairness includes all aspects of the interpersonal relationship between the decision maker and the addressee of the decision (Greenberg, 1993). Informational fairness describes whether decision-makers explain the decision adequately or not (Greenberg, 1993; Marcinkowski and Starke, 2019). Procedural fairness is about the rules of how the distribution of goods or information is managed (Alexander and Ruderman, 1987), which includes criteria such as consistency, neutrality, accuracy, revisability, ethics, and representativeness (Leventhal, 1980). Distributive fairness is about the equal distribution of resources and thus the outcome of decision-making processes (Alexander and Ruderman, 1987; Yean and Yusof, 2016).

A lot of design and empirical work in ***fairness perceptions towards algorithms and AI*** is based on interpersonal fairness concepts (Starke *et al.*, 2021). This is closely related to paradigms such as “Computers-Are-Social-Actors” (CASA) by Nass et al. (1994), which describes that humans may perceive interactions with algorithms and AI as social interactions. However, there is also empirical work available showing that people use different criteria to assess the fairness of an algorithm and a human decision-maker (Dietvorst, Simmons and Massey, 2015). This can be seen in work that uses ultimatum games, which are popular to investigate human justice and fairness perceptions: Nagataki et al. (2019) and Torta et al. (2013) found that in such games, people more often declined unfair offers from computers and robots than from humans. Perceptions of (un)fairness in robots and AI can be problematic, as they can lead to a loss of trust in the corresponding system (Zhou *et al.*, 2021), which may in turn lead to reduced user experience, adoption, and acceptance of the system (Gulati, Sousa and Lamas, 2019). In the context of AI, humans show a tendency to be critical about algorithmic decisions. For example, Lee and Baykal (2017) found that mathematically proven fair algorithmic decisions are often perceived as less fair than decisions based on a human group discussion. It is notable that the decision made by the human group does not have to be mathematically fair, but that people honored the transparency of decision-making based on their knowledge of the group discussion as well as the possibility to intervene in the discussion. In follow-up work, Lee et al. (2018) found that humans may perceive algorithmic decisions as fairer if “mechanical” skills are needed (e.g., computing optimal resource allocation or scheduling), while human decisions were found to be fairer if “human” skills were needed (e.g., hiring people). Langer et al. (2022) studied fairness perceptions when people are (only) told that decisions were taken by a human, an algorithm, and AI or other computational mechanisms, and their results matched the findings of Lee et al. (2018).

This and other work emphasize the need to study fairness perceptions towards algorithms and AI, as it shows differences in such perception compared to interpersonal fairness. This is supported by Starke et al. (2021), who conducted a literature survey about fairness and algorithmic decision-making. They conclude that there is surprisingly little research on the consequences of perceived fairness and how to design for fair interaction.

The ***interaction of humans with robots*** is different from the interaction of humans with algorithms and AI in general. Due to their embodiment and physical appearance, dealing with robots may include richer experiences and interactions than with computers, other machines, or algorithms (Salem *et al.*, 2015). This is mirrored in findings showing that the perception of unfairness in robots may have influences on the performance of teams robots are involved in (Ötting *et al.*, 2017; Claure *et al.*, 2020; Chang and Thomaz, 2021), which includes influences on the relationship between team members (Jung *et al.*, 2020) and on the trust towards the robot (Ötting *et al.*, 2017).

Short et al. (2010) found that people trusted a robot less than if it was obviously cheating on them in a game, and that their intention to play with the robot diminished because of this. While this could have been expected, Jung et al. (2020) show that the unfair behavior of robots affects teams *and* individuals. The researchers used a tower construction task in which two people have to build a tower with wooden blocks, and in which a robot managed the allocation of the blocks. They found that the way in which resources are allocated (equally or unequally) had an influence on the interpersonal relationship between team members. This suggests that a lack of distributive fairness may impact interpersonal relationships. In a similar study, Claure et al. (2020) explored the equitable distribution of blocks for a collaborative Tetris game by a robot. Their results showed that considering fairness in the allocation of resources leads to better trust. Chang et al. (2021) found that people perceived robot support of two co-workers to be fairer if the robot devoted equal time to support both co-workers, and if the co-workers completed equal shares of the work. These studies show that the perception of robot fairness matters and that it depends on the behavior of the robot. However, for all studies mentioned, it needs to be emphasized that the team members were supposed to solve a task *together*.

Regarding the fairness of service robots, Lee et al. (2012) conducted a two-month field study using a service robot to distribute snacks in an office. Their study involved repeated interaction with the service robot (nine times on average for each participant), and it included personalized conversations with the robot about snack choices or apologies for mistakes. Besides other aspects, the authors observed indications of fairness around the distribution of snacks. For example, they found participants to feel envy and to be treated unfairly in comparison to others because they assumed that the robot's delivery order favored certain participants, while this was not the case. Seo and Lee (2021) investigated the use of service robots at restaurants. They found "that consumers perceived risk is high when they have more uncertainty about the malfunction, errors, or technical inaccuracies from unfamiliar serving or chef robots." They also emphasize that if the customer has more trust in the service robot, the perceived usefulness and the chance that the guest will visit the restaurant again is higher (Seo and Lee, 2021).

Brandao et al. (2020) investigated fairness of robot navigation in the context of rescue robots. They discuss the impact of context on the fairness of robot navigation, and they illustrate this by demonstrating the context of where to apply robots for rescue purposes in a city. They show that there are dilemmas such as picking rich districts with young and healthy over neighborhoods with elderly people. They end up with the conclusion that a compromise between efficiency and fairness has to be made for fairness-conscious navigation planning of rescue robots, and they recommend considering the situation and context when planning fair robot navigation. Hurtado et al. (2021) found that fair navigation planning by robots plays an important role for social acceptability in social environments. They define five factors for fair robot navigation: reaching "value alignment" between humans and the robot, considering "bias evaluation" to avoid discrimination, considering "deterrence" to cultural differences and to be respective with cultural habits, "non-maleficence", meaning that there is no opportunity to violate people by the robot navigation, and "shared benefit" to serve humans equal independent of the scenario.

While there is work available on the perceived fairness of robots interacting with humans, we found that, among the available studies, most describe different scenarios than ours, e.g., collaboration instead of competition. To the knowledge of the authors, there is only the work by Lee et al. (2012) providing insights into the fairness perception of service robots among people who are only *indirectly related* (via the distribution). While this study of service robots is not exclusively focused on fairness, Lee et al. observed that humans developed "feelings around fairness and the distribution of resources" (Lee *et al.*, 2012). Given the prevalence and utility of service robots in many situations, the absence of studies on fairness of these robots is a ***research gap*** to be addressed to inform the design of fair human-robot interactions in service robotics. Our work on this gap is driven by two central research questions:

RQ1: To which extent does efficient delivery of resources performed by a service robot lead to perceived unfairness?

RQ2: What impact does the perceived unfairness of service robots have on the design of human-robot interaction?

3 Method

To provide initial answers to the research questions mentioned above, we ran an experiment in which participants completed competitive puzzle tasks and were supported by a service robot delivering the required material. In the experiment, two participants and one agent from our team competed against each other. The perception of unfairness was supposed to be influenced by the sequence in which the robot distributed the material for the competition.

3.1 Experiment Task: the Puzzle Competition

We designed a competition in which participants had to complete multiple building-block puzzle tasks.

For every *single building-block task*, the participants received a 16 × 16 base plate, a set of black and white building blocks in a bag, and an instruction sheet consisting of a schematic image (see Figure 1, left). The goal of the task was to arrange the blocks on the base plate as shown in the instruction as quickly as possible. All puzzle tasks followed this principle, but the specific pattern changed. The pattern always had the size of 12 × 12 and had to be placed in the center of the base plate (see Figure 1, right).

The *competition* between the participants was designed as follows: all competitors received an identical set of four numbered manuals before the competition. Their goal was to complete each of the four tasks as quickly as possible. The competitor to finish the last task first was rewarded with a bonus (2 Euros), which was communicated to the competitors beforehand. The competition was turn-based, so there were four separate rounds in which the competitors were taking part. Each round consisted of a *building phase* and an *ordering phase* (material for the next round). In the first round the competitors already had the required material on the table. In this first round, all competitors started at the same time on command and tried to solve the puzzle as fast as possible (*building phase*). Once the task was completed, the competitors rang the call bell to trigger the service robot to come to their position, which allowed the competitors to place an order for the next round (*ordering phase*). It is noteworthy that this means that who finished the task first was also the person to order first for the next round. To place an order, the competitors had to scan a quick-response (QR) code. After all participants had ordered, the next round started. From the second round on, the material was provided by a service robot (see the description of the delivery process below). The competitors were asked to take the material as soon as the robot arrived and to start immediately with the task without waiting for the others. Thus, receiving material early gave competitors an advantage.

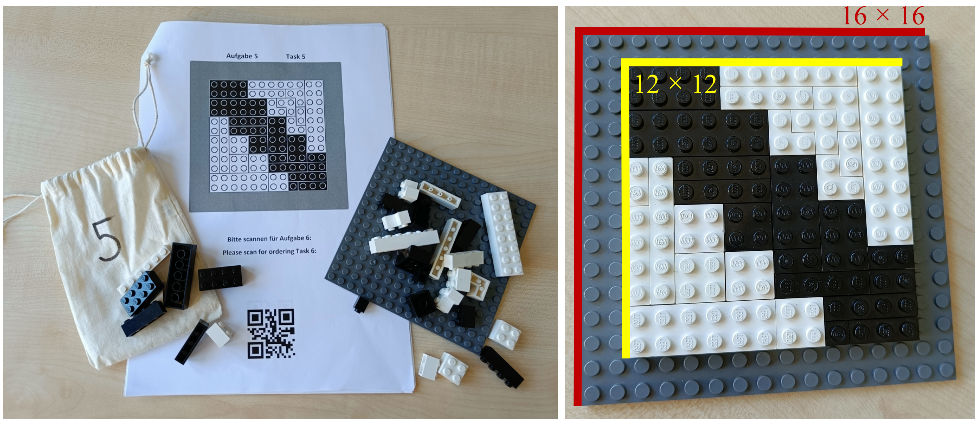

Figure 1. Material for one of the building-block tasks (left) and result of the task (right).

3.2 Design

As can be seen from the description of the experiment task, the sequence of delivery by the robot plays a crucial role for the individual chances of the competitors. We used this dependency to model (un)fairness in the **independent variable** *tour planning*, which had two levels: *user-centric* and *shortest path* (see example in Figure 2). In the condition *user-centric* the robot delivered the required material based on a queue following the FIFO principle: the participant who ordered first also received the parts first for the next task. We assumed that this way of tour planning was fair to users, as it took into account their individual success in the previous round, and therefore call it *user-centric*. In the second condition, which we named *shortest path*, the robot delivered parts to the participants in a fixed order according to their distance to the material warehouse, which is often applied in service robotics to save time and energy in delivery. Participants who sat close to this warehouse were served earlier than participants who sat further away from the warehouse, even if this conflicted with the order sequence. While this second condition reduced the robot's overall driving distance and delivery time, we considered it unfair.

Regarding the *shortest path* condition, we consider *seat position* as another independent and potentially conflating variable with the two levels *position 2* and *position 3* (position 1 was taken from the agent, see below). In the *shortest path* condition, *position 2* can be considered as a neutral position with a disadvantage over position 1 but with an advantage over *position 3*. *Position 3* had a larger disadvantage since parts were always delivered to it last. The variable *seat position* is a between-subject factor.

To compare the strategies, we ran two rounds of the study, in which the robot would apply one of the two strategies each. This was done following a **within-subject design**, in which each participant was exposed to both strategies. To prevent carry-over effects, we counterbalanced the order of strategies.

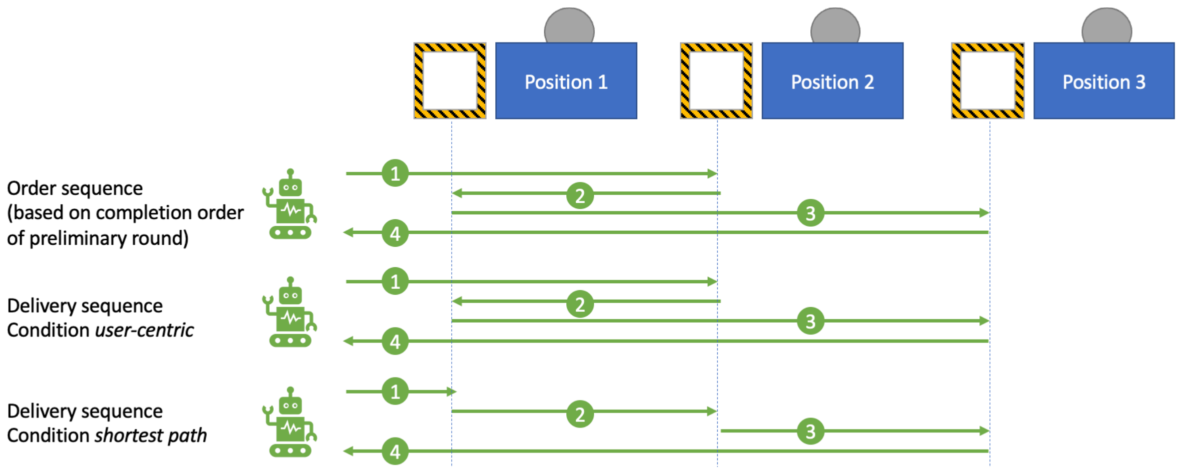

Figure 2. Example of the two delivery strategies. The completion order of the preliminary round (here: 2 ➔ 1 ➔ 3) influenced the delivery sequence only in condition user-centric.

As **dependent variables,** we measured variables relating to the perception of fairness of the situation and to the overall perception of the service robot.

In terms of the perception of fairness of the situation, we were questioning, whether the implications of the shortest path condition would be attributed to the robot or to their seat position. This contradiction is crucial, since participants who perceive a robot as a rational technical device might recognize a disadvantage due to the shortest path strategy but might attribute it to their (randomly assigned) position rather than to the robot. Therefore, we measured the two variables perceived *fairness of the robot* and perceived *fairness of the seat position*. To measure the intensity of perceived fairness, we included the two variables perceived *advantage*, and perceived *disadvantage*. The aforementioned variables were measured using a questionnaire with seven statements that had to be rated on a 7-point Likert scale.

The other set of variables relates to the perception of the service robot since we wanted to find out if a perceived (un)fairness has implications on the characteristics attributed to the robot. For this purpose, we measured the four variables *likeability* towards the robot, perceived *intelligence* of the robot, perceived *relationship* towards the robot, and *trust*. *Likeability* and *intelligence* were measured using the Godspeed questionnaire (Bartneck *et al.*, 2009) part III (*likeability*) and IV (*intelligence*).

Relationship was measured using an adaption of the relationship part of the Subjective Value Inventory (SVI) questionnaire (Curhan, Elfenbein and Xu, 2006). For *trust*, we asked the question “Overall, how much do you trust the robot?” that had to be rated on a seven-point scale ranging from “not at all” to “perfectly”.

3.3 The Role of the Task and the Agent for the Control of the Experiment

One challenge in the experiment was to control the course of the experiment. As there was no way to control the speed of task completion by the individual participants, different sequences could have occurred, which would have made the comparison between the experiments difficult. To be able to control the experiment, we applied two measures. First, we used tasks that took an untrained person between one and two minutes. Provided that the robot needed 2030 seconds to go from one position to the next, this made it very likely that the participants would finish the task in the sequence they were served. Second, among the three participants, we always included an agent with special instructions. This role was taken by a researcher, who was trained in executing the puzzle tasks quickly. In each round of the competition, the agent made sure that he was at least in one task the slowest and in one task the fastest to alter the sequence of ordering and consecutive serving. In particular, the agent aimed for the second position in task 1, for the first position in task 2, for the third position in task 3, and for the second position in task 4. Since the agent was trained in all puzzle tasks, this strategy worked well and the participants always experienced changes in the ranking within the rounds. Consequently, the participants could observe and perceive the implications of the delivery order.

3.4 Participants

We recruited 34 participants, all of whom were students from our university but had no prior knowledge about our research or the experiment. Out of the 34 participants, 21 identified as male, and 13 identified as female. The age of the participants ranged from 18 to 35 years ($M = 25.06$, $SD = 4.16$).

The participants received financial compensation for taking part in the experiment, which took about 60 minutes. They received a base compensation of 10 Euros (20 Euros for recurrent study participants). Additionally, participants could gain an additional bonus of up to 4 Euros (2 Euros per round and strategy applied) based on their performance in the experiment (see section 3.1).

3.5 Apparatus

For the study, we used two rooms connected by a door at our institute. One of the rooms was the material warehouse where the robot could pick up new material. The other room was the area in which the competition among the participants took place (see Figure 3).

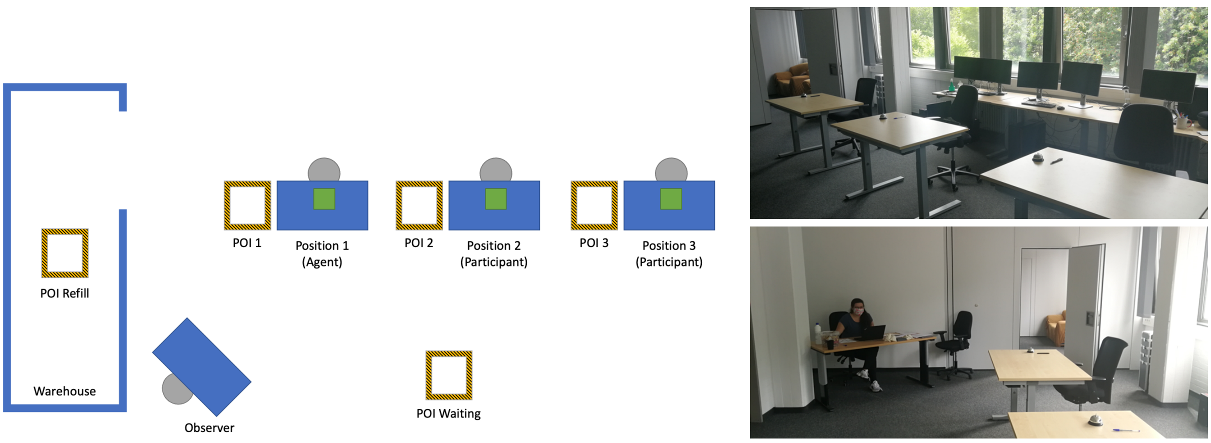

Figure 3. Left: the spatial arrangement of the seat positions. Top right: three tables of the competitors. Bottom right: position of the observer and door to the warehouse.

As can be seen in Figure 3, we placed three tables and chairs in the middle of the room. Position 1 was the table at which the agent was sitting during the experiment. Positions 2 and 3 were the seating positions of each of the two participants that took part in a run of the experiment. Each table was equipped with a call bell. Diagonally opposite to the participants, we placed an observer for the experiment (Figure 3 bottom right), who also controlled the robot. Controlling the robot was done in an unobtrusive manner (single touch on the touchpad) so that the participants did not notice.

For our experiment, we used the service robot *James* from Belgian manufacturer *ZoraBots* (see Figure 4). It is 80 cm high and weighs 17kg. Its built-in LIDAR sensor and simultaneous localization and mapping (SLAM) algorithms allow the robot to navigate through a defined area based on a prerecorded map that contains Points of interest (POI) and obstacles. For interacting with users, *James* is equipped with additional sensors like a microphone, speakers, a 10" capacitive touch display, and a button on his head. To enable *James* to deliver material to the competitors we attached a small self-made plastic backpack to its back (see Figure 4 middle).

For the graphical user interface, most of the time animated eyes were shown on *James's* touchscreen (see Figure 4 left). The eyes were blinking from time to time. The goal of this animation was to give the service robot a more human-like appearance. Only for scanning QR codes as part of ordering material this user interface was changed as shown in Figure 4 right. In this mode, the camera was activated, and the camera image was shown to the user to give visual feedback for scanning the code. In the *ordering phase*, the main interactions with users were done using the speech-output engine of the robot. After a competitor rang the bell, the robot drove to the competitor and asked "Hello, would you like to order?" After scanning the QR code the robot replied "Thanks, I will bring you task [number of round]."

Figure 4. *Service robot James with eye animation (left) and the attached backpack (centre). By scanning QR codes, participants could order new material (right).*

For the navigation of the service robot, we pre-programmed points of interest (POIs) in the lab and used a Wizard-of-Oz (WOz) approach to let the robot appear to be autonomous. We defined five POIs as shown in Figure 3. Using these POIs, the observer could send the service robot to the required POI by using its control software. Since this manual selection was unobtrusive and the routing itself was done automatically by the software, the service robot appeared to be completely autonomous.

3.6 Procedure

The participants registered for the experiment using an online form, which also collected demographic data. We created a meeting area in front of our lab for all participants. The agent also went to this meeting point and acted as a participant until the end of the experiment. When all three competitors were complete, we guided them to the lab. When entering the room, the moderator assigned seats to the participants and the agent. While the agent was always placed at position 1, the participants were randomly placed at position 2 and position 3. Before starting, participants filled in a written consent form adhering to the local data protection regulations. Then the robot *James* was introduced, and the task and competition were explained in detail to the participants. In the introduction, we told the participants that the service robot was able to deliver the material autonomously. We did not give any information about the delivery order. After this introduction, we started one practice round to make the participants familiar with the task. The participants received the instruction sheet and material for the practice task. We emphasized that this practice task was not part of the competition. The participants started on command and had to solve the task and ring the bell. Once they rang the bell, *James* drove to them in the order in which the participants completed the task to become familiar with the *ordering phase*. The participants were then asked to scan the QR code. Finally, the participants had the chance to ask questions. After the introduction, we ran two competitions: one competition under the condition *user-centric* and one under the condition *shortest path*, with both competitions consisting of four separate puzzle tasks. As mentioned above, we counter-balanced the order, so nine out of 17 sessions started with the *user-centric* condition and eight sessions started with *shortest path*. After each competition (four puzzle tasks), we first announced the winner of the round and then handed out the questionnaire to collect the participants' perceptions of fairness and the robot in the experiment. After both competitions were finished and all questionnaires were completed, we informed the participants about the actual background of the experiment, the manual control of the robot, and the role of the agent. We collected informal feedback on their perception of the robot and its "autonomy". The participants were asked not to share details of the experiment with others and they received their compensation.

4 Results

Here, we present the results in terms of the perception of fairness and in terms of perception of the robot.

4.1 Perception of Fairness

Looking at perceived fairness we look into the variables *fairness of the robot*, *fairness of the seat position*, *advantage*, and *disadvantage*. All of the variables were measured on a scale from 1 (low) to 7 (high). The results are presented in Figure 5 and Table I. The latter also contains the pairwise comparisons using the Wilcoxon signed-rank test.

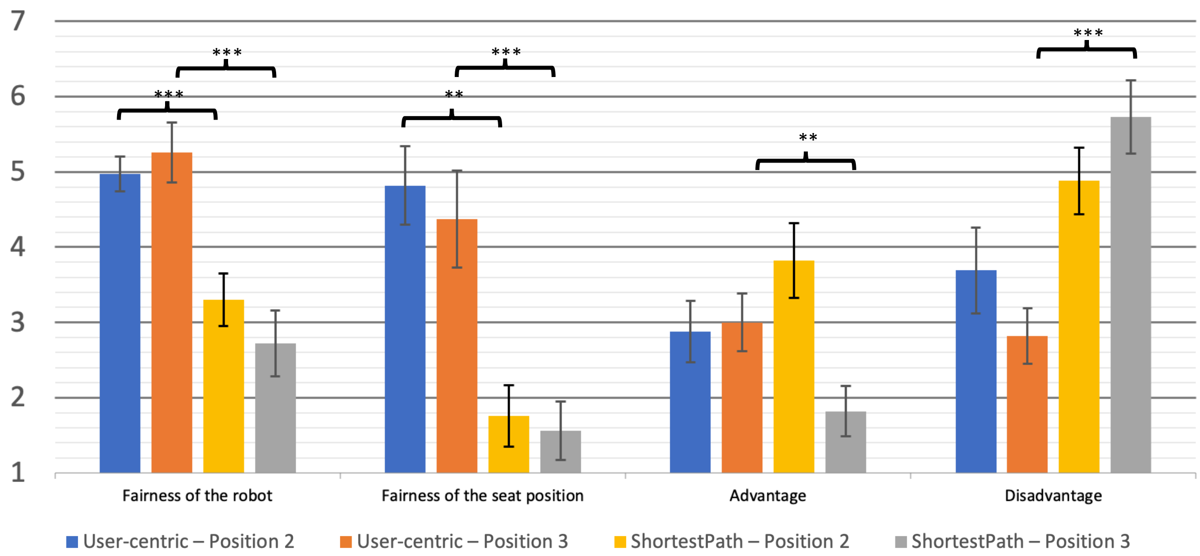

Figure 5. Participants' perception of fairness in the experiment. The scale of each variable reached from 1 (lowest value) to 7 (highest value). The error bars depict the standard error and significant differences are marked with asterisks.

For *fairness of the robot* we observed significant differences between the *user-centric* and *shortest path* conditions for both positions together as well as for the individual positions. Comparing the mean values of *position 2* and *3* shows that the differences between the rating of the *user-centric* condition and the *shortest path* condition are higher for *position 3*, which was the position that had the stronger disadvantage under the *shortest path* condition. However, comparing the data of *position 2* and *position 3* using the Mann-Whitney U test does neither show significant results between the seat positions in the *user-centric* condition ($U = 98.00$, $p = 0.170$) nor in the *shortest path* condition ($U = 89.00$, $p = 0.140$).

Again, for *fairness of the seating* we could observe significant differences between the *user-centric* and *shortest path* condition for both positions together and for the individual positions. Analyzing the data using the Mann-Whitney U test does not show a significant difference between *position 2* and *position 3* in the *user-centric* condition ($U = 124.00$, $p = 0.656$) or in the *shortest path* condition ($U = 123.00$, $p = 0.533$).

Perceived Fairness of the Robot						Perceived Fairness of the Seating					
		M	SD	Z	p			M	SD	Z	p
Both Positions	User-centric	5.12	1.30	-4.46	<.001	Both Positions	User-centric	4.61	2.34	-4.33	<.001
	Shortest Path	3.01	1.59				Shortest Path	1.67	1.59		
Position 2	User-centric	4.97	0.93	-3.19	<.001	Position 2	User-centric	4.82	2.16	-3.20	<.001
	Shortest Path	3.30	1.40				Shortest Path	1.76	1.68		
Position 3	User-centric	5.26	1.59	-3.13	<.001	Position 3	User-centric	4.37	2.58	-2.92	0.002
	Shortest Path	2.72	1.76				Shortest Path	1.56	1.55		
Perceived Advantage						Perceived Disadvantage					
		M	SD	Z	p			M	SD	Z	p
Both Positions	User-centric	2.94	2.82	-0.41	0.702	Both Positions	User-centric	3.24	1.90	-3.37	<.001
	Shortest Path	1.61	1.99				Shortest Path	5.29	1.85		
Position 2	User-centric	2.88	1.69	-1.79	0.086	Position 2	User-centric	3.69	2.27	-1.66	0.100
	Shortest Path	3.82	2.04				Shortest Path	4.88	1.78		
Position 3	User-centric	3.00	1.58	-2.62	0.008	Position 3	User-centric	2.82	1.43	-3.07	<.001
	Shortest Path	1.82	1.38				Shortest Path	5.73	1.87		

Table I *Results for the four variables fairness of the robot, fairness of the seating, perceived advantage, and perceived disadvantage for both positions and for each of the single positions. We ran pairwise comparisons using the Wilcoxon signed-rank test that resulted in the shown Z and p values.*

In the variable perceived *advantage*, we measured the perceived advantage over the competitors. Here we could observe significant differences for each of the positions, but not for the complete data set (see Table I). The Mann-Whitney U test revealed that there is a significant difference for the ratings between *position 2* and *position 3* in the *shortest path* condition ($U = 57.50$, $p = 0.002$) but not in the user-centric condition ($U = 137.00$, $p = 0.792$).

In a similar way, with the variable *disadvantage* we measured whether the participants perceived a disadvantage over the other competitors. This variable resulted in a significant difference when looking into the data of *position 3* as well as when considering both positions, but we could not show a significant difference when looking into the data of *position 2*. However, the Mann-Whitney U test did not show significant differences for the ratings between *positions 2* and *3* ($U = 106.50$, $p = 0.279$ for *user-centric* and $U = 77.50$, $p = 0.084$ for *shortest path*).

4.2 Perception of the Robot

Figure 6 and Table II show the results of the four robot-related variables *likeability*, *intelligence*, *relationship,* and *trust*, which we used to measure the perception of the robot under different conditions. Beyond the given results and analysis in Table II, we analyzed the differences between *positions 2* and *3* using the Mann-Whitney U test, which resulted in significant differences for the variable *likeability* in the *shortest path* condition ($U = 63.50$, $p = 0.026$) but not for the *user-centric* position and not for the other three variables.

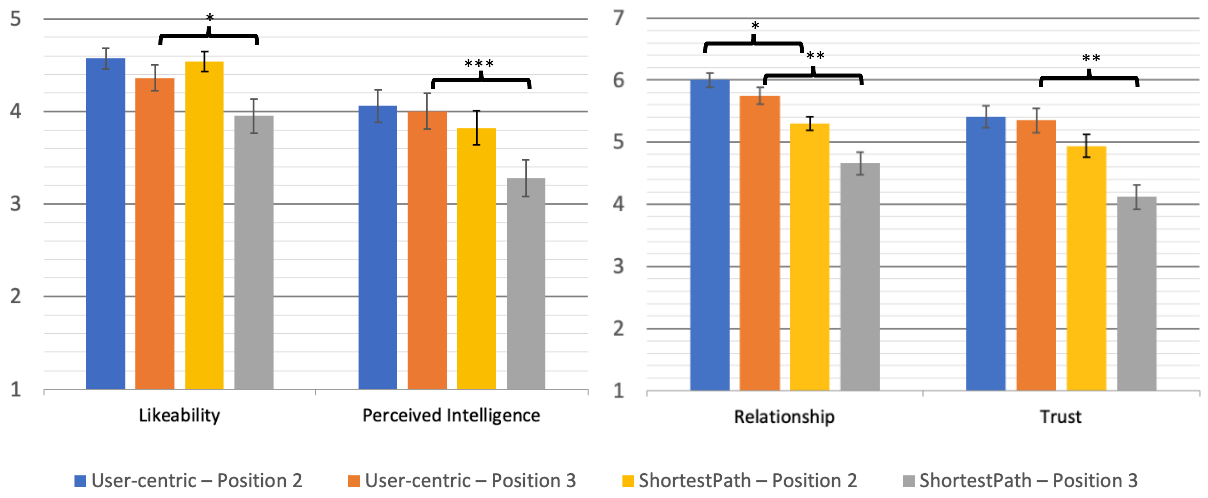

Figure 6. Participants' perception of the service robot. The scale reached from 1 (lowest value) to 5 (highest value), respectively from 1 (lowest value) to 7 (highest value). The error bars depict the standard error and significant differences are marked with asterisks.

Likeability						Intelligence					
		M	SD	Z	p			M	SD	Z	p
Both Positions	User-centric	4.47	0.51	-2.10	0.037	Both Positions	User-centric	4.03	0.72	-3.50	<.001
	Shortest Path	4.22	0.68				Shortest Path	3.55	0.80		
Position 2	User-centric	4.57	0.45	-0.14	0.992	Position 2	User-centric	4.06	0.70	-1.54	0.141
	Shortest Path	4.54	0.43				Shortest Path	3.82	0.74		
Position 3	User-centric	4.36	0.55	-2.53	0.012	Position 3	User-centric	4.00	0.77	-3.07	<.001
	Shortest Path	3.95	0.74				Shortest Path	3.28	0.79		
Relationship						**Trust**					
		M	SD	Z	p			M	SD	Z	p
Both Positions	User-centric	5.87	0.96	-3.78	<.001	Both Positions	User-centric	5.38	1.30	-3.01	0.002
	Shortest Path	5.32	1.27				Shortest Path	4.53	1.48		
Position 2	User-centric	6.00	0.75	-2.39	0.015	Position 2	User-centric	5.41	1.46	-1.49	0.166
	Shortest Path	5.30	1.15				Shortest Path	4.94	1.44		
Position 3	User-centric	5.75	1.13	-2.94	0.002	Position 3	User-centric	5.35	1.17	-2.70	0.006
	Shortest Path	4.66	1.10				Shortest Path	4.12	1.45		

Table II Results for the four variables likability, intelligence, relationship, and trust for both positions and for each of the single positions. We ran pairwise comparisons using the Wilcoxon signed-rank test that resulted in the shown Z and p values.

5 Discussion

Based on the results regarding the perception of fairness, we argue that the different conditions had a strong influence on the perception of fairness, which we expected beforehand since the participants actually had an objective disadvantage over the agent in the *shortest path* condition. The awareness of the participants that there actually was something unfair in this condition was also shown in the observations. Participants showed irritation about the delivery order and some participants even complained – interestingly in favor of their competitors: when they were delivered second even though they were last in the previous round, one participant drew attention to a supposed error: "*Excuse me, something is unfair! The robot should go to him first*". The data and the observations during the experiment show that seemingly optimal path planning can be perceived as unfair, even though it might not be intended. Apart from perceiving the robot as unfair in the shortest path condition, the position of the seat was considered unfair to the same extent. The low ratings of the seating position were independent of the particular position. The comparison between the variables *advantage* and *disadvantage* is very interesting. While the participants rejected the statement to have an advantage in the *user-centric* condition (which was objectively not the case), they had a neutral position towards this statement in the *shortest path* condition on *position 2*. Since this position actually had an advantage over *position 3*, but a disadvantage over *position 1*, this rating seems to be rational. However, when asked whether their competitors had an advantage, the participants in *position 2* mostly agreed in the *shortest path* condition, which is surprising, since one of the competitors had a clear advantage but the other a clear disadvantage, which would make a neutral rating more rational. We explain this observation with a self-serving bias, a cognitive bias that leads to the effect that humans tend to attribute positive outcomes to their own behavior but negative outcomes to external factors (Campbell and Sedikides, 1999). In a similar way, in our experiment, the advantage was rated neutral, but the advantage of the others was perceived clearly since this led to an attribution of poor performance to the other competitors. This observation holds for *position 2* only. In *position 3*, no one felt a personal advantaged but a huge advantage for the others was perceived, which fits the objective situation.

Given the data about the perception of the robot, we can conclude that the perceived (un)fairness had an impact on the perception of the service robot, especially if people personally perceived strong negative consequences. When looking at either the complete data or exclusively the data of *position 3*, the four related variables are rated significantly lower in the *shortest path* condition, which we attribute to the perceived unfairness. In this case, the participants *liked* the robot less, they perceived the quality of the *relationship* towards the robot as lower and they did not *trust* the robot as much. Additionally, they perceived it as less *intelligent*, which is interesting because they could also have considered the more efficient routing as unfair but "intelligent", which participants apparently did not perceive in this way. These effects seem to be dominated by the ratings of the participants in *position 3* since the differences in *position 2* are not significantly different except for the assessment of the *relationship*.

From the observations and short interviews after the sessions, it can be said that the participants perceived the robot to be autonomous. Some participants mentioned theories on how they thought the robot worked: most of them created a connection between the used call bells and the robot and assumed that the robot reacts to the sound of the ringing bells. If people perceived the robot as an autonomous agent, it is not surprising that they attribute unfairness to the robot. Given that aversion against unfairness is deeply rooted in human evolution (Brosnan and de Waal, 2014) and that people want to be served fairly by a service robot (Chang and Thomaz, 2021), it is not surprising that this perceived unfairness leads to disliking the robot and to the perception of lower relationship quality.

In our experiment, we intentionally selected two conditions that had the potential to influence the perceived fairness in the intended way – one condition, that is in line with human objectives (*user-centered*) and one condition that causes a conflict between efficiency and human objectives (*shortest path*). Of course, the results of the experiment might have been different if the mentioned conflict between robot and human objectives would not exist. Still, we argue that these two conditions are two extreme cases of realistic scenarios. If designers of robots focus on efficiency they might end up in scenarios, where they design service robots that are efficient in their task, e.g. in delivery, but cause negative emotions in humans as observed by Lee et al. (2012) and could in consequence lack in user acceptance. One could assume that humans understand that service robots are trimmed for efficiency and therefore exhibit behavior that can lead to advantages or disadvantages to humans in individual cases. However, our study provides empirical evidence that humans are sensitive to such unfair situations and attribute negative characteristics to the robot and trust it less. It is not the case that people see a robot as a rational technical device that just does its job. Given our data, we argue that reactions to the robot's behavior shown here are similar to the reaction to unfair treatment by a human.

As in any study, our findings are subject to certain ***limitations***: First of all, as we have laid out above, fairness is context-dependent. Therefore, the design of the task, the chosen service robot, and its particular user interface influence the results. Especially the perception of the robot as such might influence the ratings of the participants. We chose a small commercial service robot, which resembles many robots typically used in the service sector. As discussed above, the choice of specific tour planning strategies obviously has an influence on the results, because the impact of differences in the perceived fairness can only be measured if the differences are perceivable by the participants. Finally, our sample size might have been too small to reveal minor differences between the conditions.

6 Conclusion and Future Work

In this paper, we presented the results of an experiment in which a service robot was responsible for distributing resources among competitors. We investigated the influence of different distribution strategies on the perceived fairness of the robot as well as on the perception of the robot by humans.

Our results show that there might be a conflict between the technically ideal or efficient behavior of the robot and the fairness expectations of humans. The behavior of service robots may be perceived as unfair if a situation arises where individuals have personal advantages or disadvantages due to the used algorithm of the service robot. Our work also suggests that people are more sensitive to disadvantages in comparison to their own advantages, as our participants perceived the overall situation to be unfair although it gave them individual advantages. Furthermore, our results indicate that service robots, which treat humans in an unfair way, are perceived as less intelligent; they were re-liked less, or – in case of unfair treatment that has strong negative consequences as in *position 3* of our experiment – might be even disliked. The relationship with the robot was rated as less positive and people trusted the robot less. This suggests that humans attribute negative characteristics to a robot if it is perceived to act unfairly.

As these negative characteristics may have a strong influence on the acceptance and usage of service robots, we conclude that fairness needs to be considered in the design of future robots. While algorithms are often optimized in terms of efficiency, designers, and developers need to be aware that the optimization of efficiency might have severe consequences on the human perception of robots. Designing large-scale robotic services should therefore not strive for efficiency only. Instead designers and developers need to balance the trade-off between efficiency and humans' expectations of fairness and equal treatment. Finding this right balance will be an important success factor for future robotic services. While previous research has dealt with the question of how to implement "fair" algorithms regarding discrimination and data quality, our study shows the implications of a situation in which robots treat humans unfairly because it does not consider the social and group context they are in. This social context needs to be regarded in situations, in which a service robot interacts with a group of people to avoid the negative effects we found. This fits Riedl's requirement of taking the "socio-cultural context" into account when designing human-centered AI (Riedl, 2019b), and at the same time, it shows that there is work left to accomplish this goal.

Similar to the claim of Brandao et al. (2020) described above, ***future work*** should investigate in more detail how a trade-off between efficiency and fairness can be realized best. We assume that these design decisions are highly context depended, e. g. trust and likeability might be of higher relevance for service robots in the care sector than for service robots delivering food, and there might be different reactions of humans in collaborative or competitive situations. Given that some of our participants verbally expressed their perception of unfairness during our experiments, designers may also think about how to enable interventions of people in these situations, which have been discussed in the context of autonomous agents, e.g. in Schmidt and Herrmann (2017). Another question, that we leave out for future work, is the question of whether communicating the reason for unequal treatment can influence humans' perception of fairness and the robot. In our experiment, the different delivery strategies were not communicated to the participants. If the robot had explained that it chose a strategy that minimizes its ways, people might have the robot perceived differently.

Our work demonstrates and emphasizes the importance of considering the human perception of fairness as well as the social context in the future design of human-robot interactions with service robots. However, we are also aware that it only provides initial insights into this topic and that further work is needed. While our future work will be devoted to this, we also hope that the work presented here inspires other researchers to also investigate this important area.

References

Abdul, A. *et al.* (2018) 'Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC, Canada: Association for Computing Machinery (CHI '18), pp. 1–18. Available at: https://doi.org/10.1145/3173574.3174156.

Alexander, S. and Ruderman, M. (1987) 'The role of procedural and distributive justice in organizational behavior', *Social justice research*, 1(2), pp. 177–198.

Auernhammer, J. (2020) 'Human-centered AI: The role of Human-centered Design Research in the development of AI', *DRS Biennial Conference Series* [Preprint]. Available at: https://dl.designresearchsociety.org/drs-conference-papers/drs2020/researchpapers/89.

Bartneck, C. *et al.* (2009) 'Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots', *International Journal of Social Robotics*, 1(1), pp. 71–81. Available at: https://doi.org/10.1007/s12369-008-0001-3.

Berezina, K., Ciftci, O. and Cobanoglu, C. (2019) 'Robots, artificial intelligence, and service automation in restaurants', in *Robots, artificial intelligence, and service automation in travel, tourism and hospitality*. Emerald Publishing Limited.

Berry, L.L. and Seiders, K. (2008) 'Serving unfair customers', *Business Horizons*, 51(1), pp. 29–37. Available at: https://doi.org/10.1016/j.bushor.2007.09.002.

Blake, P.R. *et al.* (2015) 'The ontogeny of fairness in seven societies', *Nature*, 528(7581), pp. 258–261. Available at: https://doi.org/10.1038/nature15703.

Brandao, M. *et al.* (2020) 'Fair navigation planning: A resource for characterizing and designing fairness in mobile robots', *Artificial Intelligence*, 282, p. 103259.

Brosnan, S.F. (2013) 'Justice-and fairness-related behaviors in nonhuman primates', *Proceedings of the National Academy of Sciences*, 110(supplement_2), pp. 10416–10423.

Brosnan, S.F. and de Waal, F.B. (2014) 'Evolution of responses to (un) fairness', *Science*, 346(6207), p. 1251776.

Byrd, K. *et al.* (2021) 'Robot vs human: expectations, performances and gaps in off-premise restaurant service modes', *International Journal of Contemporary Hospitality Management* [Preprint].

Campbell, W.K. and Sedikides, C. (1999) 'Self-Threat Magnifies the Self-Serving Bias: A Meta-Analytic Integration', *Review of General Psychology*, 3(1), pp. 23–43. Available at: https://doi.org/10.1037/1089-2680.3.1.23.

Celis, L.E. *et al.* (2019) 'Classification with Fairness Constraints: A Meta-Algorithm with Provable Guarantees', in *Proceedings of the Conference on Fairness, Accountability, and Transparency*. New York, NY, USA: Association for Computing Machinery (FAT* '19), pp. 319–328. Available at: https://doi.org/10.1145/3287560.3287586.

Chang, M.L. *et al.* (2021) 'Unfair! Perceptions of Fairness in Human-Robot Teams', in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Vancouver, BC, Canada: IEEE, pp. 905–912. Available at: https://doi.org/10.1109/RO-MAN50785.2021.9515428.

Chang, M.L. and Thomaz, A. (2021) 'Valuable Robotic Teammates: Algorithms That Reason About the Multiple Dimensions of Human-Robot Teamwork', in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. *HRI '21: ACM/IEEE International Conference on Human-Robot Interaction*, Boulder CO USA: ACM, pp. 580–582. Available at: https://doi.org/10.1145/3434074.3446355.

Claure, H. *et al.* (2020) 'Multi-Armed Bandits with Fairness Constraints for Distributing Resources to Human Teammates', in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. *Conference on Human Robot Interaction*, 2020: ACM. Available at: http://arxiv.org/abs/1907.00313 (Accessed: 9 February 2022).

Colquitt, J.A. *et al.* (2018) 'Bringing "The Beholder" center stage: On the propensity to perceive overall fairness', *Organizational Behavior and Human Decision Processes*, 148, pp. 159–177.

Colquitt, J.A. and Rodell, J.B. (2015) 'Measuring justice and fairness.'

Colquitt, J.A., Zapata-Phelan, C.P. and Roberson, Q.M. (2005) 'Justice in Teams: A Review of Fairness Effects in Collective Contexts', in J. J. Martocchio (ed.) *Research in Personnel and Human Resources Management*. Emerald Group Publishing Limited (Research in Personnel and Human Resources Management), pp. 53–94. Available at: https://doi.org/10.1016/S0742-7301(05)24002-1.

Curhan, J.R., Elfenbein, H.A. and Xu, H. (2006) 'What do people value when they negotiate? Mapping the domain of subjective value in negotiation.', *Journal of Personality and Social Psychology*, 91(3), pp. 493–512. Available at: https://doi.org/10.1037/0022-3514.91.3.493.

Deutsch, M. (1975) 'Equity, Equality, and Need: What Determines Which Value Will Be Used as the Basis of Distributive Justice?', *Journal of Social Issues*, 31(3), pp. 137–149. Available at: https://doi.org/10.1111/j.1540-4560.1975.tb01000.x.

Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) 'Algorithm aversion: people erroneously avoid algorithms after seeing them err.', *Journal of Experimental Psychology: General*, 144(1), p. 114.

Fehr, E. and Schmidt, K.M. (1999) 'A theory of fairness, competition, and cooperation', *The quarterly journal of economics*, 114(3), pp. 817–868.

Følstad, A. and Taylor, C. (2020) 'Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives', in A. Følstad et al. (eds) *Chatbot Research and Design*. Cham: Springer International Publishing, pp. 201–214. Available at: https://doi.org/10.1007/978-3-030-39540-7_14.

Greenberg, J. (1993) 'The Social Side of Fairness: Interpersonal and Informational Classes of Organisational Justice', in Cropanzano (ed.) *justice in the workplace: Approaching fiarness in human resource management.* Hillsdale, pp. 79–103.

Gulati, S., Sousa, S. and Lamas, D. (2019) 'Design, development and evaluation of a human-computer trust scale', *Behaviour & Information Technology*, 38(10), pp. 1004–1015. Available at: https://doi.org/10.1080/0144929X.2019.1656779.

Hartzog, W. (2014) 'Unfair and deceptive robots', *Md. L. Rev.*, 74, p. 785.

Holstein, K. *et al.* (2019) 'Improving fairness in machine learning systems: What do industry practitioners need?', in *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–16.

Hurtado, J.V., Londoño, L. and Valada, A. (2021) 'From Learning to Relearning: A Framework for Diminishing Bias in Social Robot Navigation', *Frontiers in Robotics and AI*, 8. Available at: https://www.frontiersin.org/articles/10.3389/frobt.2021.650325 (Accessed: 22 August 2022).

Jung, M.F. *et al.* (2020) 'Robot-Assisted Tower Construction—A Method to Study the Impact of a Robot's Allocation Behavior on Interpersonal Dynamics and Collaboration in Groups', *ACM Transactions on Human-Robot Interaction*, 10(1), pp. 1–23. Available at: https://doi.org/10.1145/3394287.

Keyes, O. (2018) 'The misgendering machines: Trans/HCI implications of automatic gender recognition', *Proceedings of the ACM on human-computer interaction*, 2(CSCW), pp. 1–22.

Langer, M. *et al.* (2022) '"Look! It's a Computer Program! It's an Algorithm! It's AI!": Does

Terminology Affect Human Perceptions and Evaluations of Algorithmic Decision-Making Systems?', in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '22), pp. 1–28. Available at: https://doi.org/10.1145/3491102.3517527.

Lee, M.K. *et al.* (2012) 'Ripple effects of an embedded social agent: a field study of a social robot in the workplace', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '12), pp. 695–704. Available at: https://doi.org/10.1145/2207676.2207776.

Lee, M.K. (2018) 'Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management', *Big Data & Society*, 5(1), p. 2053951718756684.

Lee, M.K. and Baykal, S. (2017) 'Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division', in *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pp. 1035–1048.

Lee, N., Madotto, A. and Fung, P. (2019) 'Exploring Social Bias in Chatbots using Stereotype Knowledge.', in *WNLP@ ACL*, pp. 177–180.

Lee, W.-H., Lin, C.-W. and Shih, K.-H. (2018) 'A technology acceptance model for the perception of restaurant service robots for trust, interactivity, and output quality', *International Journal of Mobile Communications*, 16(4), pp. 361–376.

Leventhal, G.S. (1980) 'What Should Be Done with Equity Theory?', in K.J. Gergen, M.S. Greenberg, and R.H. Willis (eds) *Social Exchange: Advances in Theory and Research*. Boston, MA: Springer US, pp. 27–55. Available at: https://doi.org/10.1007/978-1-4613-3087-5_2.

Marcinkowski, F. and Starke, C. (2019) 'Wann ist Künstliche Intelligenz (un-) fair?', in J. Hofmann et al. (eds) *Politik in der digitalen Gesellschaft| Band*, p. 269.

McFarlin, D.B. and Sweeney, P.D. (1992) 'Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes', *Academy of management Journal*, 35(3), pp. 626–637.

Nagataki, S. *et al.* (2019) 'Can Morality Be Ascribed to Robot?', in *Proceedings of the XX International Conference on Human Computer Interaction*, pp. 1–4.

Nass, C., Steuer, J. and Tauber, E.R. (1994) 'Computers are social actors', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 72–78.

Ok, C., Back, K.-J. and Shanklin, C.W. (2006) 'Service Recovery Paradox: Implications from an Experimental Study in a Restaurant Setting', *Journal of Hospitality & Leisure Marketing*, 14(3), pp. 17–33. Available at: https://doi.org/10.1300/J150v14n03_03.

Ötting, S.K. *et al.* (2017) 'Why Criteria of Decision Fairness Should be Considered in Robot Design', *CSCW 2017: The 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, p. 7.

Pastaltzidis, I. *et al.* (2022) 'Data augmentation for fairness-aware machine learning: Preventing algorithmic bias in law enforcement systems', in *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 2302–2314.

Radhakrishnan, J. and Gupta, S. (2020) 'Artificial Intelligence in Practice – Real-World Examples and Emerging Business Models', in S.K. Sharma et al. (eds) *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation*. Cham: Springer International Publishing, pp. 77–88. Available at: https://doi.org/10.1007/978-3-030-64849-7_8.

Riedl, M.O. (2019a) 'Human-centered artificial intelligence and machine learning', *Human Behavior and Emerging Technologies*, 1(1), pp. 33–36.

Riedl, M.O. (2019b) 'Human-centered artificial intelligence and machine learning', *Human Behavior and Emerging Technologies*, 1(1), pp. 33–36.

Ruf, B. and Detyniecki, M. (2022) 'A Tool Bundle for AI Fairness in Practice', in *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1–3.

Salem, M. *et al.* (2015) 'Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust', in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15. the Tenth Annual ACM/IEEE International Conference*, Portland, Oregon, USA: ACM Press, pp. 141–148. Available at: https://doi.org/10.1145/2696454.2696497.

Schmidt, A. and Herrmann, T. (2017) 'Intervention user interfaces: a new interaction paradigm for automated systems', *interactions*, 24(5), pp. 40–45.

Seiders, K. and Berry, L.L. (1998) 'Service Fairness: What It Is and Why It Matters', *The Academy of Management Executive (1993-2005)*, 12(2), pp. 8–20.

Seo, K.H. and Lee, J.H. (2021) 'The Emergence of Service Robots at Restaurants: Integrating Trust, Perceived Risk, and Satisfaction', *Sustainability*, 13(8), p. 4431. Available at: https://doi.org/10.3390/su13084431.

Shneiderman, B. (2022) *Human-Centered AI*. Oxford University Press. Available at: https://doi.org/10.1093/oso/9780192845290.001.0001.

Short, E. *et al.* (2010) 'No fair!! An interaction with a cheating robot', in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 219–226. Available at: https://doi.org/10.1109/HRI.2010.5453193.

Siu, N.Y.-M., Zhang, T.J.-F. and Yau, C.-Y.J. (2013) 'The Roles of Justice and Customer Satisfaction in Customer Retention: A Lesson from Service Recovery', *Journal of Business Ethics*, 114(4), pp. 675–686. Available at: https://doi.org/10.1007/s10551-013-1713-3.

Starke, C. *et al.* (2021) 'Fairness Perceptions of Algorithmic Decision-Making: A Systematic Review of the Empirical Literature'. arXiv. Available at: http://arxiv.org/abs/2103.12016 (Accessed: 23 August 2022).

Torta, E. *et al.* (2013) 'The ultimatum game as measurement tool for anthropomorphism in human–robot interaction', in *International Conference on Social Robotics*. Springer, pp. 209–217.

Tyler, T.R. and Bies, R.J. (1990) 'Beyond Formal Procedures: The Interpersonal Context of Procedural Justice', in *Applied Social Psychology and Organizational Settings*. Psychology Press.

Wang, R., Harper, F.M. and Zhu, H. (2020) 'Factors Influencing Perceived Fairness in Algorithmic Decision-Making Algorithm Outcomes, Development Procedures, and Individual Differences', in. *Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM.

Winograd, T. (1996) *Bringing design to software*. ACM.

Yean, T.F. and Yusof, A.A. (2016) 'Organizational Justice: A Conceptual Discussion', *Procedia - Social and Behavioral Sciences*, 219, pp. 798–803. Available at: https://doi.org/10.1016/j.sbspro.2016.05.082.

Zhou, J. *et al.* (2021) 'Understanding Relations Between Perception of Fairness and Trust in Algorithmic Decision Making'. arXiv. Available at: http://arxiv.org/abs/2109.14345 (Accessed: 23 August 2022).