# How Gender and Type of Algorithmic Group Discrimination Influence Ratings of Algorithmic Decision Making

SONJA UTZ[♦][1]
Leibniz-Institut für Wissensmedien, Germany
University of Tübingen, Germany

Algorithms frequently discriminate against certain groups, and people generally reject such unfairness. However, people sometimes display an egocentric bias when choosing between fairness rules. Two online experiments were conducted to explore whether egocentric biases influence the judgment of biased algorithms. In Experiment 1, an unbiased algorithm was compared with an algorithm favoring males and an algorithm favoring married people. Experiment 2 focused only on the first two conditions. Instead of the expected gender difference in the condition in which the algorithm favored males, a gender difference in the unbiased condition was found in both experiments. Women perceived the unfair algorithm as less fair than men did. Women also perceived the algorithm favoring married people as the least fair. Fairness ratings, however, did not directly translate into permissibility ratings. The results show that egocentric biases are subtle and that women take the social context more into account than men do.

*Keywords: algorithm acceptance, algorithmic bias, egocentric bias, fairness, permissibility*

Algorithms are increasingly involved in consequential decision making, such as who gets hired, gets a loan, or gets parole (Srivastava, Heidasta, & Krause, 2019). In many respects, algorithms are superior to humans in making these decisions: they process a larger amount of data, never get tired, and are uninfluenced by emotions. However, there are also drawbacks. When algorithms are trained with data that reflect historical biases against certain groups, they also make biased decisions. Although people tend to perceive discriminating (vs. fair) algorithms as less fair (Hong, Choi, & Williams, 2020; Starke, Baleis, Keller, & Marcinkowski, 2022), Wang, Harper, and Zhu (2020) showed that outcome favorability also colors the perceived fairness of algorithms. Humans judge algorithms that provide favorable outcomes as fairer than algorithms that provide unfavorable outcomes. The tradeoffs people make between algorithmic fairness and outcome favorability have, however, not yet been studied.

Sonja Utz: s.utz@iwm-tuebingen.de
Date submitted: 11-19-2022

[1] Sonja Utz is a member of the Machine Learning Cluster of Excellence, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC number 2064/1–Project number 390727645.

This article aims to fill this gap by exploring whether individuals whom a biased algorithm would likely advantage evaluate this algorithm as more favorable than those who would likely be disadvantaged. A second aim is to examine whether it matters whether the disadvantaged group is frequently disadvantaged in society (women) or not (singles). This comparison shows whether people are averse to unfairness in general or are especially averse to algorithms that perpetuate societal discrimination. To study these questions, two online experiments were conducted using a loan scenario and manipulating the biasedness of the algorithms. Before turning to the specific hypotheses of the article, I will situate work on algorithmic fairness in the broader context of fairness research in the social sciences.

## Fairness Definitions

People care about fairness, and social scientists have studied fairness for decades (Lind & Tyler, 1988; Messick & Sentis, 1983; Skarlicki & Folger, 1997; Van den Bos & Miedema, 2000). Various conceptualizations of fairness exist. Early research focused on distributive fairness—that is, just the allocation of resources. Three central rules for the distribution of resources are equality (everybody receives the same share), equity (individuals who contributed more/performed better receive a larger share), and need (individuals who need more receive more; Deutsch, 1975). These rules are mutually exclusive when people differ in their performance or needs, and some people are better off when another rule is applied. Therefore, later research, especially in organizational justice, suggested that procedural and interactional fairness also matter (Beugre & Baron, 2001; Greenberg, 1986; Skarlicki & Folger, 1997). Procedural fairness suggests that procedures and rules that determine the distribution of outcomes are known beforehand and applied consistently; interactional fairness refers to how individuals are treated when informed about their outcomes. If procedural and interactional fairness are high, people are more satisfied with unfavorable distribution outcomes (Brockner, 2002).

### *Algorithmic Fairness*

For algorithmic fairness, a recent review (Starke et al., 2022) found that many empirical studies on algorithmic fairness built on the concepts of distributive and procedural fairness. Algorithms are usually high in procedural fairness because they are not influenced by mood, sympathy, or bribery. However, they often disadvantage members of specific groups because they are developed using historical data (e.g., loan histories from a bank), which often contain biased human decisions or do not provide enough data for minority groups. Consequently, much attention in the machine learning community has been devoted to group discrimination, a specific form of fairness rarely considered in the social psychological literature. Group discrimination is defined as wrongfully imposing a relative disadvantage on people based on their group membership (Valera, 2021). Various technical solutions that focus on optimizing specific statistical indicators have been developed by machine learners (Binns, 2020; Harrison, Hanson, Jacinto, Ramiro, & Ur, 2020; Pessach & Shmueli, 2022; Srivastava et al., 2019; Valera, 2021). However, there is a mismatch between the work on group discrimination in the machine learning community and empirical studies on algorithmic fairness. According to the review by Starke et al. (2022), these studies either use general fairness measures or adapt procedural and interactional fairness measures instead of focusing on group discrimination.

## Research on Algorithmic Bias

Algorithmic biases, particularly group discrimination and how they can be reduced, have received much attention in the machine learning community. There are few works in the social sciences on this topic. A systematic review by Kordzadeh and Ghasemaghaei (2022) showed that most articles were conceptual, discussing the "ethical, legal, and design implications of algorithmic bias, whereas only a limited number have empirically examined them" (p. 388). The authors developed a theoretical model that should guide future empirical research. They used a stimulus-organism-response framework that proposes that algorithmic bias should negatively influence perceived fairness, which in turn should positively influence behavioral responses, such as algorithm appreciation or recommendation acceptance. Although this seems rather straightforward and some studies show relationships between bias and perceived fairness (Hong et al., 2020; Wang et al., 2020; see also review by Starke et al., 2022) or between perceived fairness and acceptance (Hong et al., 2020; Utz, Wolfers, & Göritz, 2021), they claim that these prepositions remain largely untested and require future research. They further argued that five contextual factors affect these relationships: individual characteristics, task characteristics, technology characteristics, organizational characteristics, and environmental characteristics. How these factors influence perceived fairness and algorithm acceptance is also understudied. The present research focuses on one specific individual characteristic: egocentric bias.

## Egocentric Biases

Egocentric biases are self-serving perceptions reported for the preference for fairness rules (Thompson & Loewenstein, 1992). As mentioned above, in distributive fairness, equality and equity are two central fairness rules that are mutually exclusive if people differ in their performance. Messick and Sentis (1983) demonstrated that people often prefer the fairness rule that favors them: Bad workers prefer that rewards are distributed equally, whereas good workers prefer that equity be applied. Such egocentric biases have been found repeatedly when it comes to the preference of certain fairness rules or moral judgments (Thompson & Loewenstein, 1992; Utz & Sassenberg, 2002; see also review by Bocian, Baryla, & Wojciszke, 2020).

Egocentric biases have not been systematically studied in the domain of algorithmic fairness. The difference is that an algorithm imposes fairness rules, not humans. Since the focus of egocentric biases is on one's outcome and not on who distributes it, this should not matter. Yalcin, Lim, Van Osselaer, and Puntoni (2021) found across 10 studies that favorable decisions made by humans (versus algorithms) led to somewhat more positive evaluations of the related company, while for unfavorable outcomes, there was no difference. Considering that people tend to feel more unfairly treated when they receive an unfavorable outcome, this indicates that the source of unfair treatment does not matter much. Furthermore, according to the computer-as-social-actors paradigm, people tend to treat computers, chatbots, or other agents, such as human social actors (Gambino, Fox, & Ratan, 2020; Nass, Steuer, & Tauber, 1994).

However, the fairness type might matter. For distributive fairness rules, the different rules (equality, equity, need) can be considered fair, so people can easily justify their egocentric choices because there are also arguments for alternative rules being fair. For group discrimination, it is harder to argue why a decision rule (regardless of whether it stems from a biased algorithm or a prejudiced human) that discriminates against a specific group should be accepted.

### ***Does It Matter Who Gets Discriminated?***

Work on egocentric biases in distributive fairness ratings has usually overlooked characteristics of other group members (e.g., race, sex) because the focus was on how outcomes should be distributed among group members with different performances or needs (Thompson & Loewenstein, 1992; Utz & Sassenberg, 2002). When considering group discrimination, the question arises whether people are generally negative toward group discrimination or whether this effect is stronger for group discrimination that mirrors historical discrimination. In several countries, certain groups of people are legally protected from discrimination. Sex, color, national origin, religion, age, and disability are legally protected classes (Droste, 2020).

It is, however, unclear which form of discrimination is perceived as more unfair. Different lines of argumentation are possible. Based on the robust finding that fairness is of fundamental importance and that many people are sensitive toward injustice (Baumert & Schmitt, 2009), it should not matter which groups are discriminated against. However, societal discussion on algorithmic decision making stresses the perpetuation of existing stereotypes and the legal aspects of discriminating against people from protected classes (Kordzadeh & Ghasemaghaei, 2022). Pethig and Kroenung (2023) found that women who fear being disadvantaged by prejudiced humans prefer to be evaluated by an algorithm because they (often falsely) assume that algorithms have higher relative objectivity when there is no information about algorithmic biases. From this perspective, one would expect that people from protected classes who place high hopes in algorithms will be more upset when algorithms discriminate against protected classes. However, people might also shy away from introducing new discrimination classes introduced by algorithms.

## **The Present Research**

This article aims to examine whether egocentric biases color judgments of algorithmic fairness and extends prior work by exploring whether it matters which group is discriminated against. Therefore, judgments of an unbiased algorithm were compared not only with judgments of an algorithm favoring men but also with an algorithm favoring married people. The experiment used a loan scenario. In this context, women are frequently disadvantaged (Bauer, Pfeuffer, Abdel-Karim, Hinz, & Kosfeld, 2020). In contrast, marital status forms a somewhat arbitrary basis for discrimination, especially because it could be expected that couples can better pay off depth.

First, the preposition that algorithmic bias negatively impacts perceived fairness (Kordzadeh & Ghasemaghaei, 2022) will be tested. There is some initial support for this claim. For example, Hong et al. (2020) found more positive evaluations of unbiased versus sexist algorithms in personnel selection. Similarly, Wang et al. (2020) showed that MTurkers judged an algorithm with different error rates for people of different genders, ages, and races to be less fair than an algorithm with similar error rates. Before discussing whether it matters which group is discriminated against, I propose a general hypothesis on unbiased vs. biased algorithms:

*H1: The unbiased algorithm is perceived as fairer (=less discriminatory) than the two biased algorithms.*

Next, the assumption that algorithmic bias is also related to the permissibility of algorithmic decision making (Kordzadeh & Ghasemaghaei, 2022), a measure often used in work on algorithm aversion (Bigman & Gray, 2018), will be tested. Fairness can be expected to predict perceived permissibility, but the two variables are not identical. Perceived fairness relates to the cognitive process of recognizing that an algorithm discriminates against certain groups. The permissibility of algorithmic decision making is also influenced by factors such as egocentric biases or general aversion to algorithms making consequential decisions (Bigman & Gray, 2018). Utz et al. (2021) found positive correlations between several fairness dimensions and preference for algorithmic decision making. Hong et al. (2020) found that whether an algorithm was unbiased or sexist similarly affected several dependent variables: perceived fairness, emotional reactions toward, and perceived credibility of the decision. Therefore, I expect similar patterns for permissibility as for perceived fairness:

*H2: People will find it less permissible for a biased (vs. unbiased) algorithm to make loan decisions.*

*H3: There is a positive relationship between perceived group fairness and permissibility.*

Egocentric biases might influence the evaluation of algorithmic biases. In algorithmic fairness, egocentric biases have not been systematically explored, but some works have shown that outcome favorability plays a role. People generally prefer algorithms that produce a favorable (versus unfavorable) outcome for them (Wang et al., 2020; Yalcin et al., 2021). None of these prior studies pitted algorithmic bias and outcome favorability against each other; they only showed that outcomes matter but not whether outcomes were weighted more than fairness concerns. Based on the robust finding of egocentric biases when humans make decisions (Bocian et al., 2020; Messick & Sentis, 1983; Utz & Sassenberg, 2002), it is expected that egocentric biases occur when algorithms make decisions. Therefore, the algorithm favoring men should be perceived as less fair by women, making them less willing to accept decisions from this algorithm.

*H4: There is an interaction between gender and algorithm condition; women perceive the algorithm favoring men as (a) less fair and (b) less permissible than men. No gender difference is expected in the unbiased condition.*

It is, however, unclear whether similar effects occur when a nonprotected group is disadvantaged. On one hand, one could assume that justice sensitivity transfers to all forms of injustice (Baumert & Schmitt, 2009). On the other hand, people might be averse to the fact that algorithms perpetuate existing societal discrimination, especially if such algorithms introduce new dimensions of discrimination. Because of these different lines of argumentation, an open research question is posed:

*RQ1: Is there a difference in (a) perceived (group) fairness and (b) permissibility between the two biased algorithms (favoring men vs. favoring married people)?*

That women react more strongly to injustice or unfair decisions has been demonstrated for general justice sensitivity (Schmitt, Baumert, Gollwitzer, & Maes, 2010) and for algorithmic decision making (Grgić-Hlača, Weller, & Redmiles, 2020; Wang et al., 2020). However, it is unclear whether women also react more strongly to the discrimination of a nonprotected class.

*RQ2: Is there also a gender effect in the favoring married people condition?*

It is also not known whether egocentric biases occur in favor of married people. Participants might still prefer the algorithm that favors them; more concretely, singles, being victims of this discrimination, might judge this algorithm as less fair and decision making by this algorithm as less permissible than married people (Deutsch & Steil, 1988). However, marital status might be less salient as a category because it is less stable (single people can get married, and married people can get divorced). Therefore, I pose another research question:

*RQ3: Does marital status moderate the effects of condition on perceived fairness and permissibility?*

## Experiment 1

### *Method*

#### *Participants and Design*

An online experiment was conducted with a 3 (bias: none, favoring men, favoring married people) × 2 (gender) between-subjects design. The participants' gender was a quasi-experimental variable. Hypotheses, planned sample size, and analysis were preregistered at https://aspredicted.org/1QH_SWK. I aimed for 600 participants. Participants were recruited via the provider prolific; 594 participants consented that their data could be used (299 men, 292 women, three diverse; 228 married, mean age = 34, *SD* = 10.29; cell sizes range from 85 to 110).

#### *Procedure*

The study was part of a larger survey that combined several studies (behavior during video conferences and pretests of scales for a study on [professional] social media use; see also preregistration). The part relevant to the present article was introduced as a study on algorithmic decision making. Participants were to imagine that they applied for a larger loan from their bank and that it was not self-evident that it would be granted. They were further told that the bank would use an algorithm. Participants were informed that the algorithm uses information about demographics (e.g., age, gender, marital status, race), personal situation (e.g., job, salary, housing situation), savings, loan history, loan details (amount, purpose, duration), and historical data. These parts were colored; the idea was that people with some knowledge about potentially biased training data would be made aware of a potential bias and would pay more attention to the percentages for different groups provided in the next step. After seeing the information about the algorithm, participants filled in the items on algorithm permissibility and fairness. For exploratory purposes, social dominance orientation and modern sexism were assessed. The study ended with a manipulation check and debriefing. The material, data files, and syntax for both experiments can be found at https://osf.io/gh3qr/.

### *Independent Variables*

*Bias.* In line with Wang et al. (2020), bias was manipulated by displaying error rates for different groups, specifically the percentage of people who were mistakenly granted a loan (= received a loan but could not pay it back). These percentages were either almost identical for women (8.4%) and men (8.3%) in the no bias condition, higher for men (10.8% vs. 6.2% for women) in the favoring men condition, or displayed for married people (10.8%) and singles (6.2%) in the favoring married people condition. For the age groups, in all three conditions, similar percentages were given (8.2% for people younger than 30, 8.4% for 30–65, and 8.3% for people older than 65).

*Gender and marital status*. Gender and marital status were quasi-experimental variables and were assessed in the demographics section of the study.

### *Dependent Measures*

*Algorithm permissibility***.** The permissibility of algorithmic decision making was assessed with a scale by Bigman and Gray (2018). Participants indicated their agreement to three statements such as "It is appropriate for the algorithm to make these decisions" on scales from 1 = strongly disagree to 7 = strongly agree, *α* = .83 (*M* = 3.75, *SD* = 1.29).

*Algorithmic fairness.* A new item on group discrimination was created ("This algorithm disadvantages certain groups"). Additionally, three items were adapted from Utz et al. (2021) and assessed general fairness and procedural fairness (e.g., "This algorithm evaluates according to criteria that are the same for all people"). Answers were given on 7-point scales ranging from 1 = strongly disagree to 7 = strongly agree. As stated in the preregistration, the group discrimination item provides the most direct test of the hypothesis (*M* = 5.00, *SD* = 1.37; scale for the remaining three items *M* = 4.04, *SD* = 1.13).

### *Manipulation Check*

Participants were asked whether the algorithm they just read about favored women, men, singles, or married people, or if it treated all groups equally. There was also an "I don't remember" option to reduce guessing. Most participants correctly identified whether gender or marital status was the category in which percentages differed, but many did not correctly identify which group was favored (gender condition: favored men 47.2%, favored women 28.9%; marital status condition: favored married people 33.2%, favored singles 14.8%). Interestingly, in the no-bias condition (correct: 30.3%), 29.4% believed that the algorithm favored males, and 23.9% believed that it favored females.

## Results

I did not preregister exclusions based on the manipulation checks. O'Keefe (2006) argued that no manipulation checks are needed for objective manipulations, such as the length of a text (vs. psychological constructs). The percentages were objectively more unequal in the biased conditions. Even if the participants were not aware of this in the between-subjects design, there was no reason to exclude them. The fact that they do not pay enough attention to this information or do not understand it is an interesting finding. Since it is also informative to see whether the results change when only people who correctly answered the manipulation check are included, I also report the results of the non-preregistered analysis for this subsample.

### *Complete Sample*

For the group discrimination item, a 3 (bias) × 2 (gender) ANOVA revealed a main effect of bias: $F(2,585) = 6.31$, $p = .002$, $\eta^2_p = .021$. In line with H1, the unbiased algorithm was perceived as less discriminatory ($M = 4.74$) than the algorithm favoring men ($M = 5.13$, $p = .005$) and the algorithm favoring married people ($M = 5.18$, $p = .001$). RQ1a asked whether the perceived fairness of the two biased conditions differed. The fairness ratings for the two biased algorithms did not differ significantly from each other: $p = .69$, pairwise comparisons. There was an unpredicted main effect of gender, $F(1,585) = 7.97$, $p = .005$, $\eta^2_p = .013$, showing that women perceived the algorithms as more discriminatory ($M = 5.17$) than men ($M = 4.86$). An interaction between these factors qualified these two main effects: $F(1,585) = 4.26$, $p = .015$, $\eta^2_p = .014$ (see Table 1).

In contrast to H4a, women did not perceive the algorithm favoring men as more discriminatory than men did; the means were even in the opposite direction (*Ms* = 5.05 and 5.20 for women and men, respectively: $p = .452$). Women perceived both the unbiased algorithm and the algorithm favoring married people as more discriminatory than men: *p*s = .003 and .007, respectively. Thus, the answer to RQ2A is that there is a gender difference in the algorithm favoring married people. A closer look at Table 1 also shows that H1 holds only for men. Women perceived only the algorithm favoring married people as more discriminatory than the unbiased algorithm.[2]

When using permissibility as a dependent variable, there was no significant effect of bias, $F(2,585) = 1.45$, $p = .235$, $\eta^2_p = .005$. Thus, H2 was rejected, and the answer to RQ1b is that there are no differences between the two biased algorithms. H4b, which predicted an interaction between condition and gender, was also rejected. $F(2,585) <1$, $p = .985$, $\eta^2_p < .001$.

[2] An ANOVA with the mean of the remaining three fairness items revealed only a main effect of gender, $F(1, 585) = 14.31$, $p < .001$. Women judged the algorithms in general as less fair ($M = 3.69$) than men ($M = 4.22$), indicating that the group discrimination manipulation also only influenced the group discrimination aspect of fairness, but not other fairness aspects.

***Table 1. Effects of Participants' Gender and Algorithmic Bias on Perceived Group Discrimination and Permissibility.***

| | Perceived Group Discrimination | | | Permissibility | | |
|---|---|---|---|---|---|---|
| Exp. 1, all | No bias | Favoring men | Favoring married people | No bias | Favoring men | Favoring married people |
| Men | 4.49 (1.56)$_a$ | 5.20 (1.34)$_b$ | 4.90 (1.50)$_b$ | 3.97 (1.35)$_a$ | 3.78 (1.39)$_a$ | 3.93 (1.42)$_a$ |
| Women | 5.00 (1.16)$_a$ | 5.06 (1.29)$_a$ | 5.47 (1.04)$_b$ | 3.71 (1.21) $_a$ | 3.48 (1.15) $_a$ | 3.65 (1.09)$_a$ |
| Exp. 1, correct MC | | | | | | |
| Men | 3.86 (1.79)$_a$ | 5.25 (1.28)$_b$ | 5.03 (1.40)$_b$ | 3.97 (1.45)$_a$ | 3.59 (1.56)$_a$ | 3.92 (1.48)$_a$ |
| Women | 4.66 (1.29)$_a$ | 5.17 (1.18)$_{ab}$ | 5.69 (0.93)$_b$ | 3.77 (1.25)$_{ab}$ | 3.37 (1.00) $_a$ | 4.00 (0.96)$_b$ |
| Exp. 2, all | | | | | | |
| Men | 4.20 (1.64)$_a$ | 5.24 (1.40)$_b$ | | 3.68 (1.33)$_a$ | 3.43 (1.33)$_a$ | |
| Women | 4.65 (1.52)$_a$ | 5.28 (1.28)$_b$ | | 3.50 (1.17) $_a$ | 3.31 (1.26)$_a$ | |
| Exp. 2, correct MC | | | | | | |
| Men | 3.62 (1.57)$_a$ | 5.48 (1.17)$_b$ | | 3.93 (1.23)$_a$ | 3.29 (1.27)$_b$ | |
| Women | 4.46 (1.70)$_a$ | 5.47 (1.21)$_b$ | | 3.32 (1.21)$_a$ | 3.06 (1.20)$_a$ | |

*Note.* Means within a row in a block not sharing the same subscript differ at $p$ < .05.

There was, however, a significant main effect of gender, $F$(1,585) = 6.98, $p$ = .008, $\eta^2_p$ = .012, which was relevant to RQ2b. Women generally perceived algorithmic decision making as less permissible ($M$ = 3.61) than men ($M$ = 3.89). H3 predicted a positive correlation between perceived fairness and permissibility. Permissibility war correlated at $r$(594) = .41, $p$ < .001 with the recoded group discrimination item (such that higher values correspond to higher fairness). Thus, H3 was supported.

To answer RQ3, I repeated the analyses using marital status instead of gender as the independent variable. While using the group discrimination item as a dependent variable, this yielded only the already known main effect of condition: $F$(2,541) = 6.29, $p$ = .002, $\eta^2_p$ = .023. Please note that the subsample used for this analysis was smaller because only married ($n$ = 228) and single ($n$ = 319) participants were included (cell sizes between 71 and 111). Neither the main effect of marital status, $F$(1,541) = 2.26, $p$ = .125, $\eta^2_p$ = .004, nor the interaction effect were significant, $F$(2,541) = 1.30, $p$ = .495, $\eta^2_p$ = .003. Singles especially perceived the algorithm favoring married people as more discriminatory than married people ($M$s = 5.32 vs. 4.98 for singles and married people, respectively, $p$ = .090; compared with $M$s = 4.82 vs. 4.62 for singles and married people, $p$ = .335, in the unbiased condition and $M$s = 5.17 vs. 5.17 for singles and married people, $p$ = .992, in the favoring males condition). When permissibility was used as a dependent variable, none of the effects were significant, with all $p$s > .173. Thus, the answer to RQ3 is that marital status does not moderate the effects of conditions on perceived fairness and permissibility.

### ***Subsample With Correct Manipulation Check***

To see whether the pattern is similar for participants who answered the manipulation check correctly (29 men and 32 women in the no-bias condition, 40 men and 52 women in the favoring males condition, 38 men and 26 women in the favoring married people condition), I also report the results of the non-preregistered 2 × 3 ANOVAs for group discrimination and permissibility for this subgroup. For group discrimination, the main effect of the bias condition, $F(2,211) = 12.93$, $p < .001$, $\eta^2_p = .11$, and the main effect of gender, $F(1,211) = 6.24$, $p = .013$, $\eta^2_p = .029$, were significant. Because of the smaller sample size, the interaction effect was no longer significant: $F(2,211) = 2.45$, $p = .089$, $\eta^2_p = .023$. As can be seen in the second block of Table 1, the basic pattern remained the same. Women still considered the unbiased algorithm more discriminatory than men: $p = .032$. Again, men and women did not differ in evaluating the algorithm favoring males, $p = .930$.

When permissibility was used as the dependent variable, none of the effects were significant, all *F*s < 2.14, *p*s > .121.[3] The correlation between the group discrimination item and permissibility was the same, $r(210) = .41$, $p < .001$, supporting H3.

## Discussion

This experiment examined whether biased (vs. unbiased) algorithms were perceived as less fair (= more discriminatory) and whether decision making by a biased (vs. unbiased) algorithm was perceived as less permissible. Additionally, I explored whether it mattered which group was discriminated against and whether egocentric biases colored these judgments. In line with Kordzadeh and Ghasemaghaei's (2022) proposition, men perceived the unbiased algorithm as less discriminatory than the two biased algorithms. Women, interestingly, judged the unbiased algorithm as equally discriminatory as the algorithm favoring men; they only perceived the algorithm favoring married people as more discriminatory than the unbiased algorithm. This could indicate that women assume that algorithms often discriminate against women and that they are more sensitive in noticing discrimination against people from nonprotected classes (Grgić-Hlača et al., 2020; Wang et al., 2020).

In contrast to H2, bias manipulation only affected fairness judgments but not permissibility ratings, although both were positively correlated (H3). This points to different processes: a cognitive fairness judgment based mainly on the presented percentages and a broader permissibility judgment also influenced by other (fairness) aspects. The high procedural fairness of algorithms might mitigate the effect of group discrimination (Brockner, 2002) on permissibility; the other fairness aspects were not affected by the group discrimination manipulation (see Footnote 2). Prior work showed that people are aversive toward algorithms making moral decisions (Bigman & Gray, 2018; Utz et al., 2021), so the task (loan decisions) might have contributed more to the permissibility judgments than the degree of fairness. The relatively low permissibility means are pointing to the second explanation.

[3] For exploratory reasons (see RQ3 in preregistration), modern sexism and social dominance orientation were assessed, but only sexism showed small negative correlations with perceived group discrimination. Originally, exploratory analyses of interactions with gender and marital status were also planned. However, it turned out that trying to control for gender, marital status, and their interaction effects with continuous variables in one model resulted in an overcomplicated design and underpowered analyses. (2 contrast codes needed for the 3-level bias variable, and correspondingly the double number of potential interaction-terms).

The experiment did not reveal an egocentric bias manifested in the interaction effect of bias with gender on permissibility. However, considering that men correctly perceived the biased algorithms as more discriminatory, it is surprising that they did not lower their permissibility ratings accordingly. Egocentric biases might thus manifest more subtly as a smaller preference for the unbiased algorithm based on group fairness ratings.

RQ2 asked whether there were gender effects regarding the algorithm favoring married people. Although women perceived this algorithm as more discriminatory, they did not consider decision making by this algorithm to be less permissible. In contrast, when only the subsample with the correct manipulation check was used, women strongly preferred the algorithm favoring married people over the algorithm favoring males. This might be another form of subtle egocentric bias. Women who understand discrimination by an algorithm prefer algorithms that bias another group than their ingroup (=women). Another explanation could be that women are especially averse to algorithms perpetuating existing gender discrimination (Pethig & Kroenung, 2022).

Interestingly, using marital status as a predictor did not result in interaction effects with marital status. Thus, the finding by Deutsch and Steil (1988) that members of discriminated groups are more sensitive toward injustice does not generalize to nonprotected groups. Future research should aim to conceptually replicate these findings for other groups and explore whether members of other protected classes are also more sensitive to discriminating algorithms.

A limitation of Experiment 1 is that many people failed the manipulation check. Most of those who remembered which category (gender vs. marital status) was discriminated against could not correctly indicate who was favored. Indicators such as the percentage of people who get a loan, although they cannot pay it off, might be too difficult to process for people who do not frequently deal with such statistics (Saha et al., 2020). Maybe people thought one step further and concluded that receiving a loan without being able to pay it back might lead to overindebtedness and be bad in the long run. Including only people who answered the manipulation check correctly did not substantially alter the pattern. The judgments on the discrimination item became a bit more extreme toward the manipulation, confirming that people who answered the manipulation check correctly understood the manipulation. The permissibility ratings by men were almost identical. The only difference was that women judged it more permissible that the algorithm favoring married people made loan judgments (compared with the complete sample). This indicates that understanding the type of discrimination is a prerequisite for engaging in subtle forms of egocentric bias.

The manipulation might also have been too weak. The percentages of men and women receiving a loan, although they could not pay it off, varied only slightly (10.8% vs. 6.2%). There were also almost identical percentages for the three different age groups presented, which might have diluted the effect.

To get a clearer picture, I conducted a conceptual replication of the experiment with two changes. First, I used a stronger bias manipulation (displaying only percentages for men and women and increasing the differences between the groups). Second, I used a cleaner design and skipped the favoring married people condition to explore whether there were egocentric biases in judging biased algorithms before looking for boundary conditions, such as the discriminated group.

## Experiment 2

### *Method*

*Participants and Design*

An online experiment with a 2 (bias: none, favoring men) × 2 (gender) design was conducted. Gender was a quasi-experimental variable. The hypotheses, planned sample size (400), and analysis were preregistered at https://aspredicted.org/9YZ_8JM. Participants were again recruited via prolific; 399 participants consented that their data could be used. Only participants who identified as men (195) or women (199) were included in the analysis (cell sizes between 95 and 103). The mean age of the participants was 36 years (*SD* = 11.04).

*Procedure*

The study was run at the end of an unrelated survey on knowledge-related social media use. The setup was identical to Experiment 1. Participants read the loan scenario, were presented with information about the algorithm, and judged algorithmic fairness and permissibility. The order of these two judgments was counterbalanced. An exploratory analysis showed no main or interaction effects involving order, so this variable was dropped.

*Independent Variables*

*Bias*

To strengthen the bias manipulation, only percentages for men and women, but not for different age groups, were presented, and the difference between men and women was increased: 4.2% for women and 12.8% for men. In the unbiased condition, the percentage of people who mistakenly received a loan was 8.4% for women and 8.3% for men.

*Gender*

Gender was assessed in the demographics section of the survey.

*Dependent Measures*

The dependent measures were identical to Experiment 1 (group discrimination *M* = 4.85, *SD* = 1.51; permissibility *M* = 3.48, *SD* = 1.27).

*Manipulation Check*

The manipulation check item was closer to the instructions. The answering options were "made it easier for men to get a loan, even if they cannot pay it back," "made it easier for women to get a loan, even if they cannot pay it back," "did not distinguish much between men and women" and "I don't remember." In the bias condition, 76.6% of the participants correctly answered that it was easier for men to get loans. If they answered the manipulation check incorrectly, they (14.9%) often thought that it was easier for women. Only 7% answered that the algorithm did not differ much between men and women, and 1.5% indicated not remembering. In the unbiased condition, 58.6% correctly said that the algorithm did not differentiate much between men and women. A substantial group (21.7%) indicated that it was easier for women to get a loan—which could be counted as a correct answer because the percentage was 8.4% for women and 8.3% for men. However, 18.9% believed that it was easier for men. Only 1.5% indicated that they did not remember. A larger proportion of participants answered the manipulation check correctly, with stronger manipulation.

## ***Results***

### *Complete Sample*

A 2 (bias) × 2 (gender) ANOVA with the group discrimination item as a dependent variable revealed a significant main effect of bias, $F(1,390) = 32.30$, $p < .001$, $\eta^2_p = .076$. In line with H1, the unbiased algorithm was perceived as less discriminatory ($M = 4.42$) than the algorithm favoring men ($M = 5.26$). In contrast to H4a, the interaction effect between bias and gender was not significant, $F(1,390) = 2.70$, $p = .169$, $\eta^2_p = .005$, but the pattern found in Experiment 1 was replicated (see Table 1, third block). Men and women differed in the evaluation of the unbiased algorithm. The main effect of gender was not significant this time, $F(1,390) = 1.90$, $p = .101$, $\eta^2_p = .007$.[4]

When using permissibility as a dependent variable, the effect of bias was again not significant: $F(1,390) = 2.86$, $p = .092$, $\eta^2_p = .007$, although the means were toward H2. Descriptively, the participants found it more permissible that the unbiased algorithm made the decision ($M = 3.59$) than the biased algorithm ($M = 3.37$). The main effect of gender and the interaction effect were also not significant: *F*s > 1.35, *p*s > .247. Thus, H4b was also rejected.

In line with H3, perceived group fairness and permissibility were correlated positively, $r(399) = .42$, $p < .001$.

### *Subsample With Correct Manipulation Check*

Again, I conducted the same analysis only with people who answered the manipulation check correctly (58 men and 57 women in the unbiased condition, 75 men and 77 women in the biased condition). When group discrimination was used as a dependent variable, all three effects were significant. The main effect of the condition indicated that, in general, the unbiased algorithm was perceived as less discriminatory ($M = 4.04$) than the biased algorithm ($M = 5.47$), $F(1,263) = 67.99$, $p < .001$, $\eta^2_p = .21$. Thus, H1 was again supported.

[4] This time, the bias manipulation also affected the other three items, $F(1, 390) = 8.84$, $p = .003$. The unbiased algorithm was perceived as fairer (M = 4.37) than the biased algorithm (M = 4.06). No other effect was significant, all *F*s < 1. The same pattern was found for the subsample with correct manipulation check, $F(1, 262) = 23.03$, $p < .001$, *M*s = 4.49 and 3.89 for the unbiased and biased algorithm, respectively.

Overall, women perceived the algorithms as more discriminatory (*M* = 4.96) than men (*M* = 4.55), *F*(1,263) = 5.59, *p* = .019, $\eta^2_p$ = .021. Again, the interaction effect, *F*(1,263) = 5.93, *p* = .016, $\eta^2_p$ = .022, showed that women and men differed only in evaluating the unbiased algorithm (see Table 1). H4a predicted an interaction but expected differences between men and women in the biased condition and was, therefore, rejected.

When permissibility was used as a dependent variable, both main effects were significant. In line with H2, algorithmic decision making was perceived as less permissible for the biased algorithm (*M* = 3.18) than for the unbiased algorithm (*M* = 3.62), *F*(1,263) = 8.53, *p* = .004, $\eta^2_p$ = .031. Women again found it less permissible (*M* = 3.19) than men (*M* = 3.61) that an algorithm made a loan decision, *F*(1,263) = 7.69, *p* = .006, $\eta^2_p$ = .028. In contrast to H4b, the interaction effect between bias and gender was not significant: *F*(1,263) = 1.62, *p* = .205, $\eta^2_p$ = .006. Again, the correlation between the two constructs was significant, *r*(270) = .43, *p* < .001.

### *Discussion*

Experiment 2 replicated the effect that the biased algorithm was perceived as more discriminatory than the unbiased algorithm. With the stronger manipulation of bias, a higher proportion of people answered the manipulation check correctly, and the effect size for the bias main effect was larger. When only people who answered the manipulation check correctly were included, the main effect of gender and the interaction between bias and gender became significant. The central finding from Experiment 1 was replicated: Men and women differed only in their evaluation of the unbiased algorithm. More specifically, women perceived the unbiased algorithm as more discriminatory than men did. This time, permissibility was also significantly affected when using the subsample that answered the manipulation check correctly: People found it less permissible that a biased (vs. unbiased) algorithm made loan decisions. Cognitively understanding discrimination is thus a precondition for effects on permissibility. Again, the effects on permissibility were smaller, indicating that other (fairness) aspects also determine permissibility. The expected interaction effect with gender was again not significant. Future research could examine whether egocentric biases are comparatively small effects that can only be detected with a large sample.

## General Discussion

Two experiments brought together work on algorithmic and human biases and explored how biased (vs. unbiased) algorithms and their decision making were perceived and whether egocentric biases influenced these judgments (see Table 2 for an overview of the supported/rejected hypotheses), thereby testing several assumptions of the framework by Kordzadeh and Ghasemaghaei (2022). Experiment 1 additionally tested whether it mattered that people from a protected class or a somewhat arbitrarily chosen nonprotected class were discriminated against. A significant finding was that the egocentric bias was more subtle than expected; men did not lower their permissibility ratings for the algorithm favoring males as much as expected from their fairness judgments. Interestingly, men and women differed considerably in their evaluations of the unbiased algorithm; women perceived the unbiased algorithm across both experiments to be more discriminatory than men did. Which group was discriminated against mattered only to women. They judged the algorithm favoring married people to be the most discriminatory, whereas men did not distinguish between the two types of bias.

***Table 2. Results Across Experiments.***

| | Experiment 1 | | Experiment 2 | |
|---|---|---|---|---|
| Hypotheses | Complete | Correct MC | Complete | Correct MC |
| H1: The unbiased algorithm is perceived as fairer (= less discriminatory) than the two biased algorithms. | ✓<br>(for men) | ✓ | ✓ | ✓ |
| H2: People will find it less permissible for a biased (vs. unbiased) algorithm to make loan decisions. | ✗ | ✗ | ✗ | ✓ |
| H3: There is a positive relationship between perceived group fairness and permissibility. | ✓ | ✓ | ✓ | ✓ |
| H4: There is an interaction between gender and algorithm conditions; women perceive the algorithm favoring men as (a) less fair and (b) less permissible than men. No gender difference is expected in the unbiased condition. | ✗ | ✗ | ✗ | ✗ |

*Note*. MC = manipulation check.

The experiments provide several important contributions to work on algorithmic fairness. First, they utilize an item assessing group discrimination. Although people coming from the psychological fairness literature might not consider group discrimination as a fairness aspect, recent work on algorithmic fairness considers the group level an important part of distributive fairness (Kordzadeh & Ghasemaghaei, 2022). I, therefore, followed the call by Starke et al. (2022) to use fairness measures tailored to algorithmic fairness. Overall, the results confirm Kordzadeh and Ghasemaghaei's (2022) propositions that algorithmic bias affects fairness judgments and that individual characteristics—more specifically, the gender of the participants—can moderate this. Fairness information is processed differently by men and women. Men seem to base their fairness judgments mainly on the percentages given to the groups, whereas women tend to consider the context. Women perceived the discrimination against a new and nonprotected category as more discriminatory than the discrimination against women. Prior work only reported that women often perceived algorithms as less fair than men (cf. Starke et al., 2022) but did not consider differences between discriminated groups. Since this comparison has only been included in Experiment 1, future work should try to replicate this finding and test whether it generalizes to other protected and nonprotected groups.

Second, including a measure of permissibility provided a better understanding of the cognitive and affective processes influencing algorithm acceptance. Prior work has often focused either on fairness (see Starke et al., 2022) or on permissibility (when examining algorithm aversion or acceptance; Jussupow, Benbasat, & Heinzl, 2020). Cognitive fairness judgments often did not directly translate into permissibility ratings, although both measures were positively correlated. This indicated that permissibility ratings were influenced by additional criteria, such as other fairness aspects (e.g., procedural fairness) and affective influences, such as algorithm aversion. Because the permissibility ratings were, at best, at the scale midpoint, aversion against algorithms making consequential decisions was the more likely explanation. These results align with previous work showing that people are algorithm-averse in moral decision making (Bigman & Gray, 2018; Castelo, Bos, & Lehmann, 2019; Utz et al., 2021).

Third, the findings extend previous work on algorithmic fairness by examining whether egocentric biases influence judgments. The predicted clear-cut egocentric bias did not emerge; only subtle forms occurred. This is interesting because egocentric biases have been found repeatedly in distributive fairness and justice (Messick & Sentis, 1983; Thompson & Loewenstein, 1992; Utz & Sassenberg, 2002). However, in these studies, preferences for equality versus equity have been pitted against each other. Both fairness rules can be considered societally acceptable, and the preference for the ego-serving rule can be easily justified. In the current experiments, the ego-serving algorithm was biased, and it was thus harder to justify an egocentric choice as fair. These subtle biases are also surprising when looking at previous work on outcome favorability and algorithm acceptance. In these studies, people evaluated an algorithm that afforded them a favorable (vs. unfavorable) outcome as much more positive (Wang et al., 2020; Yalcin et al., 2021). Presenting an outcome is a stronger manipulation than influencing the expected outcomes by error rates for different groups. This manipulation might also have broadened the focus from the individual outcome to the broader implications of biased algorithms for society. Future work should aim to replicate these subtle biases for other groups to test the generalizability of the findings.

A remarkable finding that had implications at the macro level was that men and women differed in their evaluations of the unbiased algorithm. Women considered the unbiased algorithm not only as more discriminatory than men, but also judged it as discriminating in absolute terms (means clearly above the scale midpoint of 4; see Table 1). An explanation for this pattern is that women might be more sensitive to information about biased algorithms because they are more affected by discrimination in their daily lives. This finding seems to contrast with that of Pethig and Kroenung (2023), who found that women, as members of a frequently discriminated group, often preferred to be judged by an algorithm because they (falsely) assumed that algorithms would treat them as fairer than biased humans. The potential group discrimination aspect might have been made salient in the present experiments by providing error rates for different groups. Even the almost equal error rates in the unbiased condition might have been too high for participants to accept algorithmic decision making (see Rebitschek, Gigerenzer, & Wagner, 2021, for similar findings). That women consider unbiased algorithms discriminatory is alarming because it shows that it might be difficult to gain the trust of protected groups, even if technical solutions that mitigate algorithmic fairness are employed. Thus, future research looking at other protected groups is needed. A practical implication is that equalizing certain statistical parameters and displaying the results might not be enough to gain women's trust. Thus, it is not only important to make algorithms fairer but also important to communicate this fairness in a way that women also trust. Another practical implication stems from the high error rates in the manipulation check items. They show that it is difficult for many people to interpret the statistical parameters used by machine learners to reduce group discrimination (Saha et al., 2020). Thus, thorough explanations of the various parameters are needed.

The relatively high error rates, especially in Experiment 1, are also a limitation of the experiments. However, the results did not fundamentally change when only people who answered the manipulation check correctly were included. The main findings that men and women mainly differ in evaluating the unbiased algorithm and that fairness judgments do not perfectly translate into permissibility judgments were stable across both experiments. Another limitation is that hypothetical scenarios were used. This might have led to socially desirable answers; egocentric biases might have been stronger when actual decisions were made. The analyses with the smaller subsamples might have been underpowered for small effects. No measure of financial literacy or prior loan experience was included as a control. A strength of the present work is that fairness and permissibility ratings were assessed, and the algorithm favoring males was compared with an algorithm favoring married people.

## Conclusion

This article aimed to bring together work on algorithmic and human biases. The results of the two experiments show that egocentric biases are more subtle than predicted; the final permissibility ratings are less influenced by group fairness judgments than could be expected. Remarkably, men and women differ considerably in evaluating the unbiased algorithm, and women, in general, consider the social context. Future work should, thus, more systematically compare members of privileged and discriminated groups.

## References

Bauer, K., Pfeuffer, N., Abdel-Karim, B., Hinz, O., & Kosfeld, M. (2020). The economic consequences of algorithmic discrimination: Theory and empirical evidence (No. 287). SAFE Working Paper. doi:10.2139/ssrn.3675313

Baumert, A., & Schmitt, M. (2009). Justice-sensitive interpretations of ambiguous situations. *Australian Journal of Psychology, 61*(1), 6–12. doi:10.1080/00049530802607597

Beugre, C. D., & Baron, R. A. (2001). Perceptions of systemic justice: The effects of distributive, procedural, and interactional justice. *Journal of Applied Social Psychology, 31*(2), 324–339. doi:10.1111/j.1559-1816.2001.tb00199.x

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*(12), 21–34. doi:10.1016/j.cognition.2018.08.003

Binns, R. (2020). On the apparent conflict between individual and group fairness. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Jan*, 514–524. doi:10.1145/3351095.3372864

Bocian, K., Baryla, W., & Wojciszke, B. (2020). Egocentrism shapes moral judgements. *Social and Personality Psychology Compass, 14*(12), 1–14. doi:10.1111/spc3.12572

Brockner, J. (2002). Making sense of procedural fairness: How high procedural fairness can reduce or heighten the influence of outcome favorability. *Academy of Management Review, 27*(1), 58–76. doi:10.5465/amr.2002.5922363

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(5), 809–825. doi:10.1177/0022243719851788

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), 137–149. doi:10.1111/j.1540-4560.1975.tb01000.x

Deutsch, M., & Steil, J. M. (1988). Awakening the sense of injustice. *Social Justice Research, 2*(1), 3–23. doi:10.1007/BF01052297

Droste, M. (2020, June 20). *What are "protected classes"?* Subscript Law. Retrieved from https://subscriptlaw.com/protected-classes/#:~:text=What%20are%20the%20protected%20classes,these%20groups%20in%20all%20circumstances

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71–86. doi:10.30658/hmc.1.5

Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology, 71*(2), 340–342. doi:10.1037/0021-9010.71.2.340

Grgić-Hlača, N., Weller, A., & Redmiles, E. M. (2020). *Dimensions of diversity in human perceptions of algorithmic fairness*. ArXiv Preprint. doi:10.48550/arXiv.2005.00808

Harrison, G., Hanson, J., Jacinto, C., Ramirez, J., & Ur, B. (2020). An empirical study on the perceived fairness of realistic, imperfect machine learning models. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Jan*, 392–402. doi:10.1145/3351095.3372831

Hong, J.-W., Choi, S., & Williams, D. (2020). Sexist AI: An experiment integrating CASA and ELM. *International Journal of Human–Computer Interaction, 36*(20), 1928–1941. doi:10.1080/10447318.2020.1801226

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). *Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion (No. 168)*. Retrieved from https://aisel.aisnet.org/ecis2020_rp/168

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. doi:10.1080/0960085X.2021.1927212

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York, NY: Springer Science & Business Media. doi:10.1007/978-1-4899-2115-4_1

Messick, D. M., & Sentis, K. (1983). Fairness, preference, and fairness biases. In D. M. Messick & K. Cook (Eds.), *Equity theory: Psychological and sociological perspectives* (pp. 61–94). New York, NY: Praeger Publishers.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94), Apr*, 72–78.

O'Keefe, D. J. (2006). Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. *Communication Theory, 13*(3), 251–274. doi:10.1111/j.1468-2885.2003.tb00292.x

Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. *ACM Computing Surveys, 55*(3), 1–44. doi:10.1145/3494672

Pethig, F., & Kroenung, J. (2023). Biased humans, (un)biased algorithms? *Journal of Business Ethics, 183*(3), 637–652. doi:10.1007/s10551-022-05071-8

Rebitschek, F. G., Gigerenzer, G., & Wagner, G. G. (2021). People underestimate the errors made by algorithms for credit scoring and recidivism prediction but accept even fewer errors. *Scientific Reports, 11*(1), 1–11. doi:10.1038/s41598-021-99802-y

Saha, D., Schumann, C., Mcelfresh, D., Dickerson, J., Mazurek, M., & Tschantz, M. (2020). Measuring non-expert comprehension of machine learning fairness metrics. *Proceedings of the 37th International Conference on Machine Learning,* 8377–8387. doi:10.5555/3524938.3525714

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data. *Social Justice Research, 23*, 211–238. doi:10.1007/s11211-010-0115-2

Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional Justice. *Journal of Applied Psychology, 82*(3), 434–443. doi:10.1037/0021-9010.82.3.434

Srivastava, M., Heidari, H., & Krause, A. (2019). Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* 2459–2468. doi:10.1145/3292500.3330664

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society, 9*(2), 1–16. doi:10.1177/20539517221115189

Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes, 51*(2), 176–197. doi:10.1016/0749-5978(92)90010-5

Utz, S., & Sassenberg, K. (2002). Distributive justice in common-bond and common-identity groups. *Group Processes & Intergroup Relations, 5*(2)*,* 151–162. doi:10.1177/1368430202005002542

Utz, S., Wolfers, L. N., & Göritz, A. S. (2021). The effects of situational and individual factors on algorithm acceptance in COVID-19-related decision-making: A preregistered online experiment. *Human-Machine Communication, 3*, 27–45. doi:10.30658/hmc.3.3

Valera, I. (2021). Discrimination in algorithmic decision making. In U. Weber (Ed.), *Fundamental questions. Gender dimensions in Max Planck research projects* (pp. 15–26). Baden-Baden, Germany: Nomos Verlagsgesellschaft mbH & Co. KG. doi:10.5771/9783748924869

Van den Bos, K., & Miedema, J. (2000). Toward understanding why fairness matters: The influence of mortality salience on reactions to procedural fairness. *Journal of Personality and Social Psychology, 79*, 355–366. doi:10.1037/0022-3514.79.3.355

Wang, R., Harper, F. M., & Zhu, H. (2020, April). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Apr*, 1–14. doi:10.1145/3313831.3376813

Yalcin, G., Lim, S., Van Osselaer, S. M. J., & Puntoni, S. (2021). Thumbs up or down: Consumer reactions to decisions by algorithms versus humans. *Journal of Marketing Research, 59*(4), 696–717. doi:10.1177/00222437211070016