Disentangling Perceptions of Informational and Procedural Fairness for Algorithmic Decision-Making Processes

1 INTRODUCTION

In an effort to develop responsible algorithmic decision-making systems, organizations both in the public [13, 18, 31, 55] and private [22, 29, 43] sphere have published ethical guidelines with values[1] that these systems should preserve. However, a mismatch between formal value-driven system requirements and how people perceive them can significantly influence their adoption [36, 48, 57, 59]. That is why a growing body of work now focuses on capturing end users' ethical preferences through empirical studies [28, 30, 45, 60].

One of the main values in the context of algorithmic decision-making is *fairness*. Previous work on human decision-making defined fairness as a four-factor construct [11, 14] that enables the equitable and impartial treatment of decision subjects [20].[2] This multidimensional characterization not only considers the equitable allocation of outcomes (i.e., *distributive fairness*) [14], but also deals with the nature of the process that leads to those decisions (i.e., *procedural fairness*) [54] as well as the information (i.e., *informational fairness*) [11, 14] and the treatment (i.e., *interpersonal fairness*) [6] received by decision subjects.

Despite recent efforts towards fair algorithmic decision-making systems (e.g., adding *explanations* [8, 17, 47, 52], *human oversight* [16, 39, 40, 46, 59], and *contestability* [42, 56] to algorithmic decisions), comparatively little attention has been paid to the *evaluation* of fairness. Earlier work has predominantly followed a one-factor approach, measuring overall fairness as opposed to multiple dimensions of fairness and thus overlooking a long-standing practice in domains such as organizational justice. This has resulted in a lack of knowledge regarding the effect that the evaluated algorithmic configurations have on each of the dimensions of fairness. Unpacking and disentangling these effects is key to developing algorithmic decision-making systems that enhance feelings of justice irrespective of outcome favorability [38].

Drawing from research on organizational justice for human decision-making [3, 5–7, 23, 25] and studies on perceptions of fairness in algorithmic systems [4, 26, 36, 48, 50, 52, 56, 57, 59], we aim at disentangling perceptions of two of the four fairness dimensions: *informational* and *procedural* fairness. To this end, we systematically study configurations of algorithmic decision-making procedures with varying levels of *explainability*, *human oversight*, and *contestability* (i.e., referring to the presence of a decision appeal process [42]), and evaluate the perceived informational and procedural fairness of each configuration for high- and low-stakes decisions. We then examine the relationships between informational fairness, procedural fairness, and overall fairness. Our work is guided by the following research questions:

- **RQ1:** What effect do explanations, human oversight, and contestability have on perceived informational and procedural fairness in algorithmic decision-making processes?
- **RQ2:** Does the stakes involved in the decision have an effect on perceived informational and procedural fairness?
- **RQ3:** Does users' perceived informational and procedural fairness predict overall perceived fairness?

2 STUDY SETUP AND HYPOTHESES

Study type. Randomized controlled trial; between-subjects design with ($2 \times 2 \times 3 \times 2 = 24$) groups.

Context. Loan approval process.

Hypotheses.

- **Hypotheses related to RQ1**

Hypothesis 1a (H_{1a}). **End users perceive algorithmic decision-making processes as more informationally fair when they are accompanied with explanations.**
Rationale. Schoeffer and Kuehl [51] found that, as the provided amount of information about an algorithmic decision grows, users' informational fairness perception towards the system that made the decision increases. This study was focused on measuring informational fairness in a home loan approval process. We extend this study to evaluate the effect of explanations on informational fairness in both high-stakes (home loan) and low-stakes (holiday loan) decisions. We expect to see Schoeffer et al. [52]'s findings replicated in our own experimental setting.

Hypothesis 1b (H_{1b}). **End users perceive algorithmic decision-making processes as more procedurally fair when these processes are supplemented by human oversight rather than fully automated.**
Rationale. Previous studies have found that users consider human decisions to be fairer than fully automated, algorithmic decisions; especially for practices that are highly complex and are perceived to require human skills [36, 46]. Although recent research has not found any evidence that users generally perceive *hybrid* decision-making as fairer than entirely algorithmic decision-making [59], we do expect that human oversight will lead to increased *procedural fairness* perceptions among users in sensitive contexts (e.g., loan approval processes).

Hypothesis 1c (H_{1c}). **End users' procedural fairness perceptions differ based on the contestation procedure of an algorithmic decision-making process**.

[1]We will adopt the definition of *values* used in philosophy of science, following Birhane et al. [9]. Values of an entity are, thus, defined as properties that are desirable for that kind of entity.
[2]Colquitt [14] uses the terms *fairness* and *justice* interchangeably.

Rationale. Contestability has been defined as one of the core principles for designing ethical AI [20] and a key element for ensuring procedurally fair decision-making processes [54]. However, little is known about what contestability entails in relation to *algorithmic* decision-making [41]. We hypothesize that, as it is the case in human decision-making [54], contestation procedures in algorithmic decision-making processes affects perceived procedural fairness.

Hypothesis 1d (H_{1d}). **The effect of contestability on end users' procedural fairness perceptions is moderated by the presence of explanations.**
Rationale. Schoeffer et al. [52] found that, although including more information in explanations led to an increased perception of informational fairness, end users would not agree with the way in which different factors were being used for decision-making. This finding was in line with other studies that highlighted the need for making explanations actionable [8, 17]. We thus hypothesize that, aside from a general effect of contestability on users' procedural fairness perception (see H_{1c}), the presence of explanations and contestability on the algorithmic decision *interact* in affecting users' perceived procedural fairness.

Hypothesis 1e (H_{1e}). **The effect of contestability on end users' procedural fairness perceptions is moderated by the presence of human oversight.**
Rationale. Various studies have demonstrated end users' concern for fully automated, highly complex decision-making processes [36, 46]. That is why we expect that configurations where end users can contest what factors influence an algorithmic decision lead to varying degrees of procedural fairness perceptions in users depending on whether the final decision was made by a fully automated or hybrid system. In the study performed by Lyons et al. [42], the fairest contestation strategy was the one where information could be reconsidered. However, in that case the original decision was always made by a fully automated system. We hypothesize that, based on the decision maker of the original scenario, the urge to have an appeal process where the decision maker is reconsidered or one where factors are reconsidered will be different.

- **Hypothesis related to RQ2**

Hypothesis 2a (H_{2a}). **The effect of explanations on end users' informational fairness perceptions is moderated by the stakes of the task.**
Rationale. Binns et al. [8] found an interaction between the effect of explanation types on fairness perceptions and the nature of the presented scenario. In line with these findings, we hypothesize that, based on the nature of the task at stake (i.e., involving high or low stakes), end users will be satisfied differently with the amount of information they received.

Hypothesis 2b (H_{2b}). **The effect of human oversight on end users' procedural fairness perceptions is moderated by the stakes of the task.**
Rationale. Lee [36] demonstrated that fairness perceptions regarding the decision maker (i.e., a fully automated system or a human) were moderated by task characteristics. Nagtegaal [46] also found that the effect of involving humans on perceptions of procedural justice varied based on the complexity of the task. Despite the context being different (both these studies focused on managerial decisions) and our study considering fully automated vs hybrid decision making, we hypothesize that the stakes of the task (i.e., involving high or low stakes) will similarly moderate the effect of human oversight on procedural fairness perceptions in our study.

Hypothesis 2c (H_{2c}). **The effect of contestability on end users' procedural fairness perceptions is moderated by the stakes of the task.**
Rationale. Previous work has suggested that perceptions of fairness regarding the decision maker generally depend on the nature of the task [36]. We thus hypothesize that the stakes of the task (i.e., involving high or low stakes) also moderate the effect of contestability on users' procedural fairness perceptions.

- **Hypothesis related to RQ3**

Hypothesis 3a (H_{3a}). **End users' informational fairness perceptions are positively associated with their overall fairness perceptions.**
Rationale. This hypothesis is in line with findings in human decision-making, where informational fairness was found to influence perceptions of overall fairness [14, 24]. We hypothesize that the trend will be the same for algorithmic decision-making.

Hypothesis 3b (H_{3b}). **End users' procedural fairness perceptions are positively associated with their overall fairness perceptions.**
Rationale. Studies dealing with procedural fairness in human decision-making processes [24, 54] demonstrated that participants with a strong influence over the decision-making process were more likely to perceive a negative outcome as fair [32]. We hypothesize that for algorithmic decision-making processes, there will also be a positive relation between perceptions of procedural fairness and overall fairness.

Hypothesis 3c (H_{3c}). **End users' perceived informational and procedural fairness interact in predicting overall fairness** *Rationale.* Research in human decision-making has demonstrated that explanations provide the "information needed to evaluate structural aspects of decision-making" [25]. In line with these findings, we hypothesize that perceptions of overall fairness are not just dependent on both informational and procedural fairness, but that these two factors *interact* in predicting overall fairness perceptions.

3 METHOD

3.1 Variables

3.1.1 Independent variables. Tables 1 and 2 show how each each independent variable is displayed in practice.

- **Explanations** (*categorical, between-subjects*). We will assign each participant to one of two configurations (Table 2):
  (1) No explanation: participants will see what information the fictional loan requester has been asked to provide but not how this information is used.
  (2) Explanations are given:[3] participants will have an insight into what weight each piece of information has in the final decision (*input influence-based explanation*) and the hypothetical scenarios where Kim would have been able to have the loan approved (*counterfactuals*) These explanations are textual to limit presentation complexity [12, 57].
- **Human oversight** (*categorical, between-subjects*). We will randomly assign each participant to one of two configurations:
  (1) No human oversight: participants are told that the algorithmic decision-making process is fully automated.
  (2) With human oversight: participants are told that the loan approval process combines the usage of an algorithmic system with human expertise. The approval process will involve two steps: a first step where the algorithmic system receives the online loan request and evaluates each case; and a second step where the human expert [49] (bank employee) oversees the decision if the confidence obtained by the algorithmic decision-making system is low.
- **Contestability**[4] (*categorical, between-subjects*). Each participant will be randomly assigned to one of three configurations:
  (1) No contestability: participants are told that, due to the desire of the bank to handle the loan application process in a timely manner, in case of a rejection, there is no option for the fictional loan requester to contest the decision.
  (2) Option to contest the initial decision and provide additional information: participants are told that, in case of a rejection, the fictional loan requester has the option to make objections about the initial decision and provide any information to support the application.
  (3) Contest decision maker: participants are told that, in case of a rejection, the fictional loan requester has the opportunity to ask a human (different from the one who oversaw the process if there was already a human involved in the original decision) to review the process. This human reviewer will make a completely new decision with the information that Kim already provided for the initial decision.
- **Task stakes** (*categorical, between-subjects*). Each participant will be randomly assigned to one of two configurations:
  (1) High-stakes decision: the purpose of the loan application is to buy a house.
  (2) Low-stakes decision: the purpose of the loan application is to be able to go on a holiday trip.

3.1.2 Dependent variables.

- **Informational fairness perception** (*continuous*). Measured by the average score on the four items defined by Colquitt [14], (adapted) [5], based on Bies and Moag [6] and Shapiro et al. [53]; see Appendix A.
- **Procedural fairness perception** (*continuous*). Measured by the average score on the seven items defined by Colquitt [14] (adapted), based on Thibaut and Walker [54] and Leventhal [37]; see Appendix A.
- **Overall fairness perception** (*continuous*). Measured by a single-item rated on a 7-point Likert scale, as previously used [35, 36] (adapted); see Appendix A.

3.1.3 Descriptive and exploratory measurements. We use these variables to describe our sample and for exploratory analyses, but we do not conduct any conclusive hypothesis tests on them.

- **Age group** (*categorical*). Participants will select their age group from multiple choices; see Appendix B.
- **Level of education** (*categorical*). Participants will select the highest level of education they have completed from multiple choices; see Appendix B.
- **AI literacy** (*continuous*). Average score of the four items defined by Schoeffer et al. [52] (adapted); see Appendix B.
- **Affinity to technology** (*continuous*). Average score of the four items defined by Franke et al. [21] [6](adapted); see Appendix B.
- **Personal experience** (*continuous*). Average score of the two items defined by Kramer et al. [34] (adapted); see Appendix B.
- **Task stakes perception** (*continuous*). In this study we have considered a home loan evaluation process to be a high-stakes decision and a holiday loan evaluation process to be a low-stakes decision. Since the stakes involved in a decision are subjective and personal [33], we will capture participants' task stakes perceptions. This will be measured through an adapted version of the item defined by Lyons et al. [42]; see Appendix B.

3.2 Planned sample

We will test approximately 261 participants. We computed required sample size using the software *G*Power* [19] for an ANOVA with main effects and interactions; specifying the default effect size of 0.25, a significance threshold of $\alpha = \frac{0.05}{11} = 0.0045$ (i.e., due to testing multiple hypotheses' see Section 3.4), a desired power of 0.8, 24 groups, and the respective degrees of freedom for the different hypotheses we aim to test.

[3]In a preliminary exploratory study we compared five types of explanations and tested their *understandability* [8, 52], *actionability* [8, 52] and *contestability enactment* [2, 27, 41]. We found that participants rated the combination of input influence-based explanations [15] and counterfactuals [58] highest in every evaluation criterion.

[4]In the same preliminary exploratory study we provided participants with Kim's scenario and, through open-ended questions, asked what Kim should contest. Through a thematic analysis [10] we found that end users would like to contest two main aspect of the decision-making process: (1) the factors (and their weights) used by the bank and (2) the use of an algorithmic system for decision-making. Based on these insights we designed two different contestation strategies. These strategies resonate with the *new information condition* and *new decision condition* (with a human reviewer) defined by Lyons et al. [42].

[5]We pilot tested the wording and layout of the presented scenarios. Based on the insights we got from the pilot test, we rephrased some of the items to make them more understandable for participants.

[6]In these four items, the term "technical system" is used to refer to systems that include some technology that converts inputs to outputs using a transformation process [44].

A bank has implemented a new loan application system where potential customers apply for a loan online and then the company assesses the eligibility of the customer for the loan. <Configuration [*No human oversight*] or [*With human oversight*]> Kim, a potential customer, is looking for funding opportunities to <task> and has thus decided to apply for a <task> loan through the bank's online platform. As part of the <task> loan application process, the bank has requested the following information: – Applicant annual income – Co-applicant (if any) annual income – Credit score – Date of birth – Employment status – Education – Loan amount requested – Loan amount term (months) – Loan purpose – Number of dependents A few hours after sending the requested information, Kim has received an email with the final decision: the loan has been rejected. <Configuration [*No explanation*] or [*With explanations*]> <Configuration [*No contestability*] or [*Contest initial decision*] or [*Contest decision maker*]>

Table 1: Overview of the scenario.

We will recruit participants from *Prolific* (https://prolific.co). Each participant must be at least 18 years old, have a high proficiency in English, and can participate in our study only once. Participants will be excluded from data analysis if they do not pass at least one of the attention checks in the experiment. The study itself will be conducted on *Qualtrics* (https://www.qualtrics.com/).

3.3 Procedure

Our study will be conducted on our own server, where participants will authenticate with a registration token received on *Prolific*.

Step 1. Participants state their age group and level of education. Furthermore, their degrees of AI literacy, affinity to technology, personal experience and task stakes perception are measured.

Step 2. Participants are presented with a fictional loan approval scenario involving a person called Kim (Table 1).[7] As part of this process, Kim has applied for a loan online and is waiting for the bank to assess their eligibility. Depending on the stakes of the task that the participant has been assigned to, the purpose of this loan approval will be either to buy a house (high stakes) or to go on a holiday trip (low stakes). Participants will be informed about the information Kim has provided to the bank to evaluate the loan request. As part of the scenario, every participant will then be informed that Kim's loan request has been rejected and they will get to know the process through which the loan request made by Kim has been evaluated. Based on which of the ($2 \times 2 \times 3 \times 2 =$)24 between-subject scenarios a participant was randomly placed in, participants will know whether there was a human expert overseeing the process, they will (or will not) receive explanations and whether and how Kim can contest the decision. Participants will then respond to an attention check, where they will be asked about the purpose of the loan request.

Step 3. Participants evaluate the informational fairness, procedural fairness, and overall fairness of the decision-making process. Additionally, this step will include a second attention check that asks participants to select a specific option from a Likert scale.

Step 4. Participants are asked two open-ended questions (see Appendix B), where they have the option to describe what kind of information they would have liked to receive (if any) and what element would make the decision-making process fairer (if any).

3.4 Analysis plan

We will analyze the hypotheses we specified in Section 2 in three separate statistical analyses. First, to test H_{1a} and H_{2a}, we will conduct a multi-way ANOVA with *explanations*, *human oversight*, *contestability*, and *task stakes* as between-subjects factors and *informational fairness perception* as dependent variable.[8] Second, to test $H_{1b\text{-}e}$ and $H_{2b\text{-}c}$, we conduct another multi-way ANOVA with the same between-subjects factors but with *procedural fairness perception* as the dependent variable. Third, to test $H_{3a\text{-}c}$, we conduct a multiple linear regression analysis with *informational fairness perception* and *procedural fairness perception* as independent and *overall fairness perception* as dependent variables.

[7]Gender, age, and other demographics of Kim are not disclosed.

[8]Although we do not specifically hypothesize about effects of human oversight and contestability on informational fairness perception, we include these variables here for exploratory analyses.

<table>
<tr><th>Parameters</th><th>Conditions</th><th>Descriptions</th></tr>
<tr><td rowspan="2">Explainability</td><td>No explanation</td><td>The artificial intelligence system uses some of this information for making the loan decision.</td></tr>
<tr><td>With explanations</td><td>In the email received by Kim, an explanation of how the decision-making system has reached the conclusion is included. The email includes the importance that each piece of information provided by Kim had in the final decision. Factors are listed from the most important to the least important factor based on the bank's criteria. The magnitude of the contribution of each piece of information (negative (−) means that it contributed to the rejection decision) is added between brackets:

Credit Score (−0.15) > Loan amount requested (−0.12)> Total annual income (−0.09)> Loan purpose (+0.02)> Employment status (+0.02)> Loan amount term (months) (−0.03)> Date of birth (+0.03)> Co-applicant (if any) income (+0.01)> Number of dependents (−0.07)> Education (+0.02)

The email also includes information about scenarios where the individual would have been granted the loan. Kim would have been granted a loan if one of the following scenarios had been true:

– The loan amount requested had been 5% lower
– The total annual income of the individual had been 10% higher
– The credit score of the individual had been "Very Good"</td></tr>
<tr><td rowspan="2">Human oversight</td><td>No human oversight</td><td>Given the latest technological advances and in an effort to make loan decisions in a timely manner, the loan application process is now fully automated. An artificial intelligence system receives the online requests and evaluates each case. An email is sent to the applicants with the final verdict.</td></tr>
<tr><td>With human oversight</td><td>Given the latest technological advances and in an effort to make loan decisions in a timely manner, the loan application process is now hybrid: it combines artificial intelligence with human expertise. This involves a two-step approval process. In the first step, an artificial intelligence system receives the online requests and evaluates each case. If the artificial intelligence system reaches a decision (approve or reject) with a high confidence, an email is sent to the applicant with the final verdict. If the artificial intelligence system has a low confidence over the decision, there is a second step where a human oversees the decision and makes the final verdict and an email is sent to the applicant.</td></tr>
<tr><td rowspan="3">Contestability</td><td>No contestability</td><td>Since the reason for introducing an artificial intelligence system is to handle home loan applications in a timely manner, Kim has no option to request a review of the decision.</td></tr>
<tr><td>Contest initial decision</td><td>Kim has decided to appeal the decision and has asked for a review of the process. As part of the review procedure, Kim has the opportunity to make objections about the initial decision and provide any information to support the application. The same artificial intelligence system will then reevaluate the home loan application.</td></tr>
<tr><td>Contest decision maker</td><td>Kim has decided to appeal the decision and has asked for a review of the process. As part of the review procedure, Kim has the opportunity to ask for a human to review the process. This human reviewer will make a completely new decision with the information that Kim already provided for the initial decision.</td></tr>
<tr><td rowspan="2">Task stakes</td><td>High stakes</td><td>Buy a house / home loan</td></tr>
<tr><td>Low stakes</td><td>Go on holiday / holiday loan</td></tr>
</table>

Table 2: Summary of the experimental design.

Because we are testing 11 hypotheses as part of this study, we apply a Bonferroni correction to our significance threshold, reducing it to $\frac{0.05}{11} = 0.0045$. This means that p-values that result from the analyses described above will only be regarded as significant if they are below this reduced threshold.

In addition to the analyses described above, we may conduct posthoc tests (i.e., to analyze pairwise differences), Bayesian hypothesis tests (i.e., to quantify evidence in favor of null hypotheses), and exploratory analyses (i.e., to note any unforeseen trends in the data) to better understand our results.

4 ADDITIONAL COMMENTS

As of submitting this preregistration, data collection has not yet begun.

A MEASUREMENT OF DEPENDENT VARIABLES

A. Items to measure *informational fairness.* Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

(1) The bank thoroughly explains how the information provided by Kim is used for making a decision.
(2) The explanations regarding the <task> decision-making are reasonable.
(3) The explanations are tailored to Kim's specific needs.
(4) I understand the way the bank uses the information to make decisions.

B. Items to measure *procedural fairness.* Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

(1) Kim is able to express their views and feelings during the <task> decision-making process.
(2) Kim has influence over the decision arrived at by this procedure.
(3) The <task> decision-making is applied consistently
(4) The <task> decision-making is free of bias
(5) The <task> decision-making is based on accurate factors.
(6) Kim is able to appeal the decision arrived at by this process.
(7) The <task> decision-making process upholds ethical and moral standards.

C. Item to measure *overall fairness.* Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

(1) The <task> decision-making process is fair.

B MEASUREMENT OF DEMOGRAPHICS AND DESCRIPTIVE VARIABLES

A. Questionnaire for determining *age range.*

What is your age range?

- A1: 0-18
- A2: 19-25
- A3: 26-35
- A4: 36-50
- A5: 50-80
- A6: 80+

B. Questionnaire for determining *level of education.*

What is the highest level of school that you have completed or the highest degree you have received?

- A1: High school incomplete or less.
- A2: High school graduate or GED (includes technical / vocational training that does not award college credit)
- A3: Some college (some community college, associate's degree).
- A4: Four year college degree / bachelor's degree
- A5: Some postgraduate or professional schooling, no postgraduate degree
- A6: Postgraduate or professional degree, including master's, doctorate, medical or law degree

C. Items to measure *AI literacy.* Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

(1) I have a good knowledge in the field of *artificial intelligence.*
(2) My current employment includes working with *artificial intelligence.*
(3) I am confident interacting with *artificial intelligence.*
(4) I understand what the term *artificial intelligence* means.

D. Items to measure *Affinity to technology.* Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

(1) I like to occupy myself in greater detail with technical systems.
(2) I like testing functions of new technical systems.
(3) It is enough for me that a technical system works; I don't care about how or why. (r) [9]
(4) It is enough for me to know the basic functions of a technical system. (r)

E. Items to measure *personal experiences.* Assessed on a seven-point Likert scale (1 = completely disagree, 7 = completely agree).

(1) I have heard or had experience with a human making loan decisions for <task>.
(2) I have heard about or had experience with an *artificial intelligence system* making loan decisions for <task>.

F. Item to measure *task stakes perception.* Assessed on a seven-point Likert scale (1 = very low stakes, 7 = very high stakes).

(1) What are the stakes involved in a <task> loan decision-making process based on the impact that this decision has on end users' lives?

G. Open-ended questions.

(1) Do you think the bank offers appropriate [10] information about the decision-making process? Why? If not, what information do you think the bank should offer Kim?

[9] Reverse-coded item

[10] Schoeffer et al. [52] tested the effect of the amount of information on end users' perceptions of informational fairness. Through this open-ended question we would like to explore end users' perceptions towards the *quality* of the information they received [1].

(2) Do you think that the procedure that the bank has put in place for making <task> loan decisions in a timely manner is fair? Why? If not, what would make the procedure fairer?

REFERENCES

[1] Kars Alfrink, Ianus Keller, Neelke Doorn, and Gerd Kortuem. 2022. Tensions in transparent urban AI: designing a smart electric vehicle charge point. *AI & SOCIETY* (3 2022). https://doi.org/10.1007/s00146-022-01436-9

[2] Kars Alfrink, T. Turel, A. I. Keller, N. Doorn, and G. W. Kortuem. 2020. Contestable City Algorithms. International Conference on Machine Learning Workshop.

[3] Karl Aquino. 1995. Relationships among pay inequity, perceptions of procedural justice, and organizational citizenship. *Employee Responsibilities and Rights Journal* 8, 1 (3 1995), 21–33. https://doi.org/10.1007/BF02621253

[4] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H. de Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY* 35, 3 (9 2020), 611–623. https://doi.org/10.1007/s00146-019-00931-w

[5] Julian Barling and Michelle Phillips. 1993. Interactional, Formal, and Distributive Justice in the Workplace: An Exploratory Study. *The Journal of Psychology* 127, 6 (11 1993), 649–656. https://doi.org/10.1080/00223980.1993.9914904

[6] R.J. Bies and J. F. Moag. 1986. Interactional Justice: Communication Criteria of Fairness. . *Research on Negotiations in Organizations* 1 (1986), 43–55.

[7] Robert J. Bies and Debra L. Shapiro. 1987. Interactional fairness judgments: The influence of causal accounts. *Social Justice Research* 1, 2 (6 1987), 199–218. https://doi.org/10.1007/BF01048016

[8] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions. (1 2018). https://doi.org/10.1145/3173574.3173951

[9] Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2021. The Values Encoded in Machine Learning Research. (6 2021).

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (1 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[11] David Chan. 2011. Perceptions of fairness. *Research Collection School of Social Sciences* (2011).

[12] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F Maxwell Harper, and Haiyi Zhu. 2019. Explaining Decision-Making Algorithms through UI: Strategies to Help Non-Expert Stakeholders. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300789

[13] China Electronics Standardization Institute. 2018. Original CSET Translation of "Artificial Intelligence Standardization White Paper". https://cset.georgetown.edu/research/artificial-intelligence-standardization-white-paper/

[14] Jason A. Colquitt. 2001. On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology* 86, 3 (6 2001), 386–400. https://doi.org/10.1037/0021-9010.86.3.386

[15] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. In *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 598–617. https://doi.org/10.1109/SP.2016.42

[16] Berkeley J. Dietvorst, Joseph P. Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114–126. https://doi.org/10.1037/xge0000033

[17] Jonathan Dodge, Q. Vera Liao, Yunfeng Zhang, Rachel K. E. Bellamy, and Casey Dugan. 2019. Explaining Models: An Empirical Study of How Explanations Impact Fairness Judgment. (1 2019). https://doi.org/10.1145/3301275.3302310

[18] European Commission. 2019. Ethics guidelines for trustworthy AI. https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

[19] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (5 2007), 175–91. https://doi.org/10.3758/bf03193146

[20] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal* (2020). https://doi.org/10.2139/ssrn.3518482

[21] Thomas Franke, Christiane Attig, and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (4 2019), 456–467. https://doi.org/10.1080/10447318.2018.1456150

[22] Google. 2018. AI at Google: Our Principles. https://www.blog.google/technology/ai/ai-principles/

[23] Jerald Greenberg. 1987. A Taxonomy of Organizational Justice Theories. *The Academy of Management Review* 12, 1 (1 1987), 9. https://doi.org/10.2307/257990

[24] Jerald Greenberg. 1990. Organizational Justice: Yesterday, Today, and Tomorrow. *Journal of Management* 16, 2 (6 1990), 399–432. https://doi.org/10.1177/014920639001600208

[25] J. Greenberg. 1993. The social side of fairness: Interpersonal and informational classes of organizational justice. *Justice in the workplace: Approaching fairness in human resource management.* (1993), 79–103.

[26] Nina Grgic-Hlaca, Muhammad Bilal Zafar, Krishna P. Gummadi, and Adrian Weller. 2016. The Case for Process Fairness in Learning: Feature Selection for Fair Decision Making. In *NIPS SYMPOSIUM ON MACHINE LEARNING AND THE LAW 8*.

[27] Clément Henin and Daniel Le Métayer. 2021. Beyond explainability: justifiability and contestability of algorithmic decision systems. *AI & SOCIETY* (7 2021). https://doi.org/10.1007/s00146-021-01251-8

[28] César Hidalgo, Diana Orghian, Jordi Albo-Canals, Filipa de Almeida, and Natalia Martin. 2021. *How Humans Judge Machines.* MIT Press. https://hal.archives-ouvertes.fr/hal-03058652

[29] IBM. 2019. IBM Everyday Ethics for AI. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf

[30] Maurice Jakesch, Zana Buçinca, Saleema Amershi, and Alexandra Olteanu. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. (5 2022). https://doi.org/10.1145/3531146.3533097

[31] Technology Japanese Cabinet Office, Council for Science and Innovation. 2019. Social Principles of Human-Centric Artificial Intelligence. https://www8.cao.go.jp/cstp/english/humancentricai.pdf

[32] Denise Jepsen and John Rodwell. 2009. A New Dimension of Organizational Justice: Procedural Voice. *Psychological Reports* 105, 2 (10 2009), 411–426. https://doi.org/10.2466/PR0.105.2.411-426

[33] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. 2022. "Because AI is 100% right and safe": User Attitudes and Sources of AI Authority in India. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–18. https://doi.org/10.1145/3491102.3517533

[34] Max F. Kramer, Jana Schaich Borg, Vincent Conitzer, and Walter Sinnott-Armstrong. 2018. When Do People Want AI to Make Decisions?. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 204–209. https://doi.org/10.1145/3278721.3278752

[35] Markus Langer, Tim Hunsicker, Tina Feldkamp, Cornelius J. König, and Nina Grgić-Hlača. 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 *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–28. https://doi.org/10.1145/3491102.3517527

[36] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (1 2018). https://doi.org/10.1177/2053951718756684

[37] Gerald S. Leventhal. 1980. What Should Be Done with Equity Theory? In *Social Exchange*. Springer US, Boston, MA, 27–55. https://doi.org/10.1007/978-1-4613-3087-5{_}2

[38] E. Allan Lind and Tom R. Tyler. 1988. *The Social Psychology of Procedural Justice*. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4899-2115-4

[39] Jennifer M. Logg, Julia A. Minson, and Don A. Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (3 2019), 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

[40] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to Medical Artificial Intelligence. *Journal of Consumer Research* 46, 4 (12 2019), 629–650. https://doi.org/10.1093/jcr/ucz013

[41] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. (2 2021). https://doi.org/10.1145/3449180

[42] Henrietta Lyons, Senuri Wijenayake, Tim Miller, and Eduardo Velloso. 2022. What's the Appeal? Perceptions of Review Processes for Algorithmic Decisions. In *CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3491102.3517606

[43] Microsoft. 2018. AI Principles. https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6

[44] Enid N.H. Montague, Brian M. Kleiner, and Woodrow W. Winchester. 2009. Empirically understanding trust in medical technology. *International Journal of Industrial Ergonomics* 39, 4 (7 2009), 628–634. https://doi.org/10.1016/j.ergon.2009.01.004

[45] Albert Musschenga. 2005. Empirical Ethics, Context-Sensitivity, and Contextualism. *The Journal of Medicine and Philosophy* 30, 5 (10 2005), 467–490. https://doi.org/10.1080/03605310500253030

[46] Rosanna Nagtegaal. 2021. The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly* 38, 1 (1 2021), 101536. https://doi.org/10.1016/j.giq.2020.101536

[47] Yuri Nakao, Simone Stumpf, Subeida Ahmed, Aisha Naseer, and Lorenzo Strappelli. 2022. Towards Involving End-users in Interactive Human-in-the-loop AI Fairness. (4 2022).

[48] Gideon Ogunniye, Benedicte Legastelois, Michael Rovatsos, Liz Dowthwaite, Virginia Portillo, Elvira Perez Vallejos, Jun Zhao, and Marina Jirotka. 2021. Understanding User Perceptions of Trustworthiness in E-Recruitment Systems. *IEEE Internet Computing* 25, 6 (11 2021), 23–32. https://doi.org/10.1109/MIC.2021.3115670

[49] Christina A. Pan, Sahil Yakhmi, Tara P. Iyer, Evan Strasnick, Amy X. Zhang, and Michael S. Bernstein. 2022. Comparing the Perceived Legitimacy of Content Moderation Processes: Contractors, Algorithms, Expert Panels, and Digital Juries. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (3 2022), 1–31. https://doi.org/10.1145/3512929

[50] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. 2019. How Do Fairness Definitions Fare?. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, New York, NY, USA, 99–106. https://doi.org/10.1145/3306618.3314248

[51] Jakob Schoeffer and Niklas Kuehl. 2021. Appropriate Fairness Perceptions? On the Effectiveness of Explanations in Enabling People to Assess the Fairness of Automated Decision Systems. (8 2021). https://doi.org/10.1145/3462204.3481742

[52] Jakob Schoeffer, Niklas Kuehl, and Yvette Machowski. 2022. "There Is Not Enough Information": On the Effects of Explanations on Perceptions of Informational Fairness and Trustworthiness in Automated Decision-Making. (5 2022). https://doi.org/10.1145/3531146.3533218

[53] Debra L. Shapiro, E.Holly Buttner, and Bruce Barry. 1994. Explanations: What Factors Enhance Their Perceived Adequacy? *Organizational Behavior and Human Decision Processes* 58, 3 (6 1994), 346–368. https://doi.org/10.1006/obhd.1994.1041

[54] J. W. Thibaut and L. Walker. 1975. Procedural Justice: A Psychological Analysis. *L. Erlbaum Associates, Hillsdale.* (1975).

[55] National Science United States Executive Office of the President and Technology Council Committee on Technology. 2016. Preparing for the Future of Artificial Intelligence. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

[56] Kristen Vaccaro, Christian Sandvig, and Karrie Karahalios. 2020. "At the End of the Day Facebook Does What ItWants". *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (10 2020), 1–22. https://doi.org/10.1145/3415238

[57] Niels van Berkel, Jorge Goncalves, Daniel Russo, Simo Hosio, and Mikael B. Skov. 2021. Effect of Information Presentation on Fairness Perceptions of Machine Learning Predictors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445365

[58] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. (11 2017).

[59] Zezhong Wang, Jacob Ritchie, Jingtao Zhou, Fanny Chevalier, and Benjamin Bach. 2021. Data Comics for Reporting Controlled User Studies in Human-Computer Interaction. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2 2021), 967–977. https://doi.org/10.1109/TVCG.2020.3030433

[60] Takuya Yokota and Yuri Nakao. 2022. Toward a decision process of the best machine learning model for multi-stakeholders: a crowdsourcing survey method. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, NY, USA, 245–254. https://doi.org/10.1145/3511047.3538033