

# Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation

MIN KYUNG LEE, University of Texas at Austin & Carnegie Mellon University, USA

ANURAAG JAIN, School of Computer Science, Carnegie Mellon University, USA

HAE JIN CHA, School of Design, Carnegie Mellon University, USA

SHASHANK OJHA, School of Computer Science, Carnegie Mellon University, USA

DANIEL KUSBIT, Ethics, History & Public Policy, Carnegie Mellon University, USA

As algorithms increasingly take managerial and governance roles, it is ever more important to build them to be perceived as fair and adopted by people. With this goal, we propose a procedural justice framework in algorithmic decision-making drawing from procedural justice theory, which lays out elements that promote a sense of fairness among users. As a case study, we built an interface that leveraged two key elements of the framework—transparency and outcome control—and evaluated it in the context of goods division. Our interface explained the algorithm’s allocative fairness properties (standards clarity) and outcomes through an input-output matrix (outcome explanation), then allowed people to interactively adjust the algorithmic allocations as a group (outcome control). The findings from our within-subjects laboratory study suggest that standards clarity alone did not increase perceived fairness; outcome explanation had mixed effects, increasing or decreasing perceived fairness and reducing algorithmic accountability; and outcome control universally improved perceived fairness by allowing people to realize the inherent limitations of decisions and redistribute the goods to better fit their contexts, and by bringing human elements into final decision-making.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**.

Additional Key Words and Phrases: algorithmic decision, algorithmic mediation, transparency, control, division

## ACM Reference Format:

Min Kyung Lee, Anuraag Jain, Hae Jin Cha, Shashank Ojha, and Daniel Kusbit. 2019. Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 182 (November 2019), 26 pages. <https://doi.org/10.1145/3359284>

## 1 INTRODUCTION

Computational algorithms increasingly mediate groups, making managerial and governance decisions in society. They allocate work in labor platforms, determine the locations of resources in cities, aggregate citizens’ opinions for policy and budgeting decisions, form social groups online, and distribute rewards [2, 36, 45, 46, 74, 80, 81, 87].

Algorithms enable efficient, data-driven, scalable management of social functions, and this vision is one of the driving forces behind the increasing adoption of algorithms. On the other hand,

---

Authors’ addresses: Min Kyung Lee, University of Texas at Austin & Carnegie Mellon University, USA, [minkyung.lee@austin.utexas.edu](mailto:minkyung.lee@austin.utexas.edu); Anuraag Jain, School of Computer Science, Carnegie Mellon University, USA; Hae Jin Cha, School of Design, Carnegie Mellon University, USA; Shashank Ojha, School of Computer Science, Carnegie Mellon University, USA; Daniel Kusbit, Ethics, History & Public Policy, Carnegie Mellon University, USA.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2573-0142/2019/11-ART182 \$15.00

<https://doi.org/10.1145/3359284>

emerging studies suggest that algorithms cannot yet accommodate diverse social values such as fairness, altruism, nuance, and context, all of which can be difficult to computationally model but are vital to social justice and thriving communities [31, 43, 71]. While human decisions do accommodate these factors, human decision-making is time- and resource-intensive, difficult to scale in certain contexts, less beholden to data, and more prone to biases caused by inherent power structures or interpersonal factors, all of which may create undesirable effects and inequities. How can we leverage the strengths of both algorithmic and human decision-making?

We explore this question in fair algorithmic decision-making contexts, specifically decisions involving resource allocations among people and groups. Algorithms are increasingly tasked with calculating distributions of limited resources among parties. For example, algorithms are being used in the United States to determine which food banks donated goods should be delivered to and to assign workers to different shifts, and algorithmic tools are available to the public to help determine fair divisions of household tasks or non-monetary goods based on principles of economic fairness [2, 43, 44]. In examples such as these, algorithms must adopt a philosophy of what a fair distribution is for social justice. In this paper we evaluate fairness through the lens of procedural justice, which focuses on the perceived fairness of the processes by which outcomes are reached. Previous work has shown that having both distributive and procedural justice is critical in eliciting people's trust and cooperation and adoption of social systems [9], yet procedural justice has received relatively less attention in the ongoing discussion of algorithmic fairness.

Drawing from procedural justice theory, we first present a procedural justice framework for algorithmic decision-making. We then instantiate the framework in the context of a fair division algorithm, a deterministic algorithm that formalizes allocative fairness, used for goods allocation in a cooperative group. We leveraged transparency and outcome control, two key elements of procedural justice theory. Our interface explained algorithmic assumptions and properties (standards clarity), displayed inputs and outcomes (outcome explanation), and allowed people to discuss and interactively adjust the resulting allocations as a group (outcome control). We evaluated the interface through a within-subjects laboratory study (with 23 groups (N=71)) using a goods division task on Spliddit [1], a website that applies fair division algorithms to social decisions.

Our research study makes contributions to the ongoing discussion of algorithmic fairness and transparency. Our work offers a procedural justice framework in algorithmic decision-making. We empirically show the different effects of transparency and outcome control on perceived fairness and algorithmic accountability, contributing to emerging theories on social perceptions of algorithms and design principles for human-centered algorithms.

## 2 FAIRNESS AND JUSTICE IN ALGORITHMS

Scholars define fairness in numerous and often context-specific ways [7], for example in terms of maximizing utility for groups or society [54], or respecting various rules such as individual rights and freedoms [62]. Although it is beyond the scope of this paper to address all philosophical conceptions of fairness, one influential philosophical theory of fairness comes from the 20th-century philosopher John Rawls, who famously equated fairness and justice [62], arguing broadly that fairness is “a demand for impartiality”. Since this landmark work, scholars have often used “justice” and “fairness” interchangeably [14]. While these philosophical theories present a top-down approach to fairness, in this paper we focus on distributive and procedural justice, two aspects of fairness that are rooted in human perception and are studied in social and organizational psychology [9, 14].

Fairness with a focus on distributive justice has been a central research topic in computer theory, artificial intelligence, and machine learning. Fair division algorithms have been a subject of computer theory and economics for the past fifty years [69]. More recently, fairness in machine

learning regarding the risk of discrimination has received attention [23]. However, relatively less attention has been given to procedural justice in algorithmic decision-making processes.

## 2.1 Distributive Justice in Algorithmic Fairness

Distributive justice, or the study of what constitutes a fair distribution of resources, has been studied with regard to the design and evaluation of algorithmic systems, particularly those concerned with making fair divisions of resources or removing bias from the machine learning process [23, 65].

*2.1.1 Fair Division Algorithms.* Fair division algorithms are part of a subfield of computer science, artificial intelligence, economics, and mathematics [10, 65]. These algorithms formalize philosophical or economic fairness axiomatic principles mathematically and computationally. Fair division algorithms define their outcome metrics (e.g., the utilities that each agent receives from their division outcome), take individual agents' input (e.g., preferences or needs), and allocate resources in a way that satisfies fairness properties such as Pareto-Optimality, maximin, envy-freeness, or others [64]. The benefit of fair division algorithms is that they are proven to guarantee these properties with any combination of agents and inputs. For this reason, fair division algorithms have been applied to a wide range of tasks including networked bandwidth, computer resources, organ donations, school district assignment, and physician residency matching [65].

*2.1.2 Unbiased Machine Learning Algorithms.* More recently, in response to emerging evidence that machine learning algorithms can make discriminatory decisions [71], researchers have been investigating computational techniques that make machine learning algorithms unbiased and non-discriminatory. Multiple conceptualizations of fairness and corresponding techniques have been proposed. Individual fairness techniques allow machine learning algorithms to learn to classify similar individuals similarly, depending on one's definition of similar individuals [22]. Group fairness research offers statistical techniques to treat protected groups similarly to the population as a whole [15]. Another set of validation techniques has been proposed to allow third parties to audit the process and examine whether the machine learning algorithms make decisions following individual or group fairness principles [18].

## 2.2 Emerging Work and Gaps in Procedural Justice in Algorithmic Fairness

Although it has received relatively less academic attention than distributive justice, encouraging new work has begun to investigate human perceptions of algorithmic fairness. One line of work has sought to understand people's expectations of "fair" algorithm rules or features. For example, Grgic-Hlaca et al. surveyed what features should be used for recidivism prediction [29, 30]. Other scholars conducted online studies, interviews and workshops to learn what people think "fair" algorithms are in the context of online information delivery [86], donation allocation [44] and loan decisions [66]. Another line of work has examined how people judge the fairness of algorithmic decisions. Lee et al. studied the role of decision source, showing that algorithmic hiring and evaluation [42] as well as task allocation [43] decisions are perceived less fair than human managers' decisions. Binns et al. investigated different explanation styles in algorithmic insurance decisions, finding no impact of explanation style on perceived fairness [8]. Dodge et al. further examined explanation styles and highlighted the importance of individual differences in algorithmic recidivism decisions [20]. These studies shed light on multiple factors that contribute to the perceived fairness of algorithms; yet, most of this work has been done in hypothetical settings with participants who lack domain knowledge and real experiences. For example, mTurk crowd workers judged hypothetical hiring, recidivism, and insurance decisions.

Advancing this line of research, we propose a comprehensive procedural justice framework for algorithmic fairness drawing from procedural justice theory. The framework can be used to

coherently organize findings from the emerging work and highlights areas that have received less attention, such as outcome control and decision-maker benevolence. As a case study, we applied the framework to design the interface that instantiated transparency and outcome control, and evaluated it by conducting a real division task and examining people's experiential responses to real algorithmic decisions.

### 3 PROCEDURAL JUSTICE FRAMEWORK IN ALGORITHMIC DECISION-MAKING

In this section, we draw from procedural justice literature in order to develop a framework that can guide the design of procedural fairness for algorithmic decision-making. While significant attempts have been made to develop systematic frameworks of procedural justice elements [9, 14, 47, 58], there is no single accepted comprehensive framework. The elements of procedurally-just decision-making presented here draw on multiple frameworks in the field of procedural justice and are brought together for their relevance to the design of computational decision-making systems (Figure 1). We explain how each element has value in informing the design of algorithmic decision-making processes and draw from previous work on algorithmic transparency and control.

#### 3.1 Procedural Justice Theory

Procedural justice is the perceived fairness of a decision-making process [73]. The academic study of procedural justice began when psychologists discovered that perceptions of decision outcome fairness depended not only on the outcomes themselves, but also on the process by which the outcome was determined [47, 73]. People may perceive identical outcomes as fair or unfair depending on the process used to determine the outcome, and even unfavorable outcomes can be perceived as fair given fair procedure [73]. Research in the fields of social psychology, organizational justice, and conflict resolution shows that procedural justice can lead to increased satisfaction with and adherence to negotiated agreements [53, 60, 84], as well as increased acceptance of decision-making authorities [50, 53, 77]. Procedural justice research has covered the areas of organizational management [6, 9, 25], healthcare services [75], law enforcement [21], and trust in government [78].

Our framework of procedural justice for algorithmic decision-making consists of transparency and control.

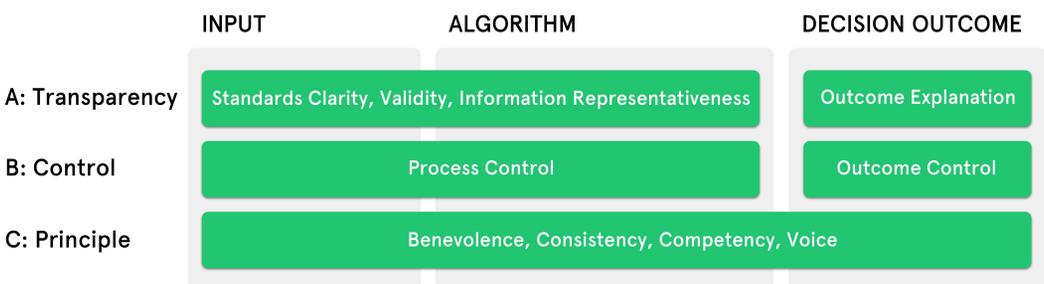


Fig. 1. Procedural justice framework in algorithmic decision-making.

#### 3.2 Transparency

Transparency enables individuals to see and evaluate how decisions are made and to draw conclusions about the decision outcomes and the decision-maker [13].<sup>1</sup> The importance of algorithmic

<sup>1</sup>Transparency does not guarantee procedural fairness. Rather, it is a method to promote procedural fairness. In our framework, transparency is used as a mechanism that enables people to judge standard clarity and validity and information representativeness. The resulting perceived fairness depends on whether people think the used standard is fair or not.

transparency has been highlighted by numerous researchers in multiple domains such as recommender systems, journalism, hiring, criminal sentencing, and online education [5, 16, 19]. More recently, European privacy law codified the “right to an explanation” for users of platforms in which data is collected by mandating that organizations give a degree of transparency to explain what data is collected and how it is used [26]. Most empirical work and proposed interaction techniques and implementations have been done in the context of individual decision-making and have focused on explanation of outcomes.

**3.2.1 Standards Clarity.** Standards clarity (Figure 1, A) communicates the rules and methods the decision-maker uses to make decisions. In organization contexts, this is often referred to as clarity of decision-maker expectations [13, 25, 58, 72]. Standards clarity enables individuals to operate under correct knowledge of how the decision-maker will interpret and respond to their input. When communication of standards is lacking, people may adopt false assumptions regarding decision-making processes and tailor their input in such a way that undesirable or unfair results are produced [13]. In the context of performance appraisal, when managers clarified their expectations and made clear the rules of evaluation during the performance review processes, perceptions of fairness increased among employees [72]. In algorithmic group mediation, the types of information that can be communicated include: what input algorithms take, what logics and models algorithms use to process the input, how fairness is operationalized to mediate different individuals’ interests and utilities, and the performance of the algorithm.

**3.2.2 Standards Validity.** Standards validity (Figure 1, A) is the extent to which the rules of the decision-maker are perceived as fair and warranted [11]. Standards are viewed as invalid if they are based on “unsound reasoning or unfounded presumptions” regarding the decision situation [58, 72]. In a study of the perceived fairness of layoff decisions in the context of corporate downsizing, Brockner et al. found that the perceived fairness of the decisions was shaped by whether individuals viewed the decision criteria (i.e., seniority, merit, job function, etc.) as appropriate [11]. In algorithmic decisions, communicating standards will allow people to judge whether they are appropriate for the decision context. Race, for example, would be considered an invalid standard for an algorithm that determines which individual receives loans after applying.

**3.2.3 Information Representativeness.** Information representativeness (Figure 1, A) is the extent to which information and decision standards reflect the “values and concerns” of the participants in the decision [47]. Information representativeness enables the decision-maker to adequately assess and respond to the opinions of individuals involved in the decision-making process and prevents them from feeling excluded. In group decision contexts, psychological studies have shown that the fairness of decision outcomes is heavily influenced by how representative the decision is of the views of group members [55]. In the context of an allocation algorithm, a high degree of information representativeness can be ensured by obtaining input from all members of the decision.

**3.2.4 Explanation of Decision Outcomes.** Once the decision has been made, whether and how the results are explained to individuals influences the extent to which the decision is viewed as fair [28, 67] (Figure 1, A). Adequate explanation may be especially important in the wake of poor decision outcomes [27, 28], as decision explanations can illustrate the degree to which the decisions were necessary given the circumstances [6, 11]. In Colquitt and Rodell’s model this is called “informational justice” and is characterized by providing truthful and thorough explanations for how decisions were reached [14]. Taking the time to explain results also improves fairness perceptions on an interpersonal level by being interpreted as a signal of respect for individuals, reinforcing to them that they are of value to the group [53]. Studies in organizational contexts

involving layoff decisions have found that explaining how layoff decisions were made significantly increased the perceived fairness of those decisions [6, 28].

Prior empirical work has focused on decision explanations in algorithmic systems, including types and effects of explanations for algorithmic decisions in recommender, context-aware, and personal analytics systems [3]. This line of work suggests multiple methods to explain mediated decision outcomes. For example, explanations can focus on explaining input and output following the white box model [33] or explaining “why,” with rationales for decisions in text form without exposing input and output data [85]. Design dimensions of explanations have also been explored, including the soundness and completeness of explanations [41] and types of information such as what and when [48].

### 3.3 Control

Thibaut and Walker developed the control model of procedural justice that conceptualizes fairness as a function of the degree of control over the decision that individuals receive [73]. When individuals have more control over the processes that lead to decision outcomes (hereafter “process control”) and decision outcomes (“outcome control”), they perceive the results to be fairer [51, 73].

**3.3.1 Process Control.** Process control is the ability to influence what evidence or data is considered by the decision-maker, how that evidence is presented, and the rules by which the evidence is interpreted [35, 73] (Figure 1, B). Instrumentally, process control increases fairness perceptions by enabling individuals to shape the decision outcome in their favor with their input [9, 35, 73]. Socially, it increases fairness perceptions by symbolizing status recognition and benevolence on the part of the decision-maker [9, 52, 76]. People tend to prefer legal processes in which they have greater degree of control regarding the selection and presentation of evidence [35, 51, 79].

In algorithmic decision-making contexts, process control may include allowing individuals to determine input data or giving individuals the ability to influence the rules and logics of the algorithm itself. Much of previous work on algorithmic control has focused on process control in the context of decisions for individuals. Examples include allowing end-users to control the input and rules of recommendation engines [32, 37, 57] and enabling individuals to train algorithms by providing examples and demonstrations [4]. The group mediation context brings up the new research challenge of determining how to distribute levels of control across individuals, as the process influences multiple people rather than one individual.

**3.3.2 Outcome Control.** Outcome control is based on the term “decision control” coined by Houlden et al. and refers to the ability to appeal or modify the outcome of a decision once it has been made [35] (Figure 1, B). Outcome control enables correctability and possible recourse against decisions that are wrong or improper [47]. This is also known as the right to appeal [17]. Outcome control is a recognition that, as Leventhal states, “even the most well-intentioned and competent decision-makers commit errors or oversights” [47]. In the United States, the right to appeal legal decisions to a higher court is embedded in law. Procedural justice research has documented the ability of outcome control to increase perceptions of fairness in layoff decision-making [6] and employee evaluation decisions [56, 58].

In algorithmic decision-making contexts, outcome control will enable individuals or groups to reject algorithmic outcomes through appeal or by finding alternative outcomes. While there is recent recognition that this is an important area [34], this dimension of control has been less explored, probably because in many recommender systems, users are not bound to take the recommendations and thus have full control over whether to use the outcomes. Further research is needed to study outcome control in algorithmic decision-making that considers both individual and social levels of decision impact. For example, individuals’ decisions to accept or reject an outcome can influence

other members of the group, and decisions can be renegotiated among individuals with the aid of the algorithm as an ongoing decision-making tool.

### 3.4 Overarching Principles

Four procedural justice elements apply to all phases of algorithmic decision-making in our model (Figure 1, C): Demonstrations of consistency, competency, benevolence, and voice.

**3.4.1 Consistency.** Consistency is adopted from Leventhal's landmark work and is, in his words, the ability to make decisions in like fashion, "across persons and over time" [47]. Consistency includes equal, consistent application of rules across differing parties and situations. Consistency has been found to be an important factor in making determinations of fairness in job application contexts [70]. Consistency is an important factor in data collection, decision-making (the treatment of data), and in the communication of results. Failure to behave in a consistent manner may violate egalitarian norms of equal treatment, thereby resulting in discriminatory behavior [68]. Algorithms have the potential to increase consistency in decision-making, compared to human decision-makers.

**3.4.2 Competency.** Competency is the perception that the decision-maker is qualified to make a decision [58, 72]. To design fair procedures, the competency of the decision-maker (algorithm) should be communicated during all phases of the decision-making process. Demonstrating competency is intended to bolster trust in the service and its validity and acceptance of the results.

**3.4.3 Benevolence.** Benevolence relates to beliefs regarding the decision-making authority's motivations, including their "willingness to consider one's needs and try to make fair decisions" [9, 53]. Colquitt and Rodell's model of justice contains a similar concept of "interpersonal justice" that is based upon the enactment of procedures that are respectful and sincere [14]. Communication of the benevolent intent behind decision-making algorithms can increase perceptions of fairness. Those that are communicated in ways that stress their impartiality and desire to provide the best outcome for all users should achieve this aim.

**3.4.4 Voice.** Voice is the opportunity to provide input or feedback in the decision-making process. [49]. When individuals are given the opportunity to express their opinions, even after a decision has been finalized, perceptions of outcome fairness increase [24, 49]. For algorithmic decision-making, having a channel through which people can express concerns or opinions at each stage of decision-making will be critical.

### 3.5 Applying the Framework to Contexts

This framework has identified procedural elements that can make algorithmic decision-making more fair. Design decisions as to how to implement these elements are context-dependent. For example, the allocation of rides in for-profit organizations such as Uber or Lyft ridesharing systems versus the allocation of food donations for non-profits in a city will allow for different levels of transparency, control, and attributed benevolence because of their differences in power and group structure (e.g., hierarchical groups in which managers make decision for workers versus non-hierarchical groups in which non-profits make decisions for themselves), cultures (e.g., cooperative versus competitive), and concerns about proprietary knowledge. It is not feasible to explore all design dimensions and contexts in one study or paper. As a first step in exploring this space, we focus on algorithms that allocate goods to a cooperative group where individuals make decisions for themselves using an algorithmic tool. In this context, we explore how we can leverage transparency and outcome control, and empirically examine their effects on perceived fairness in algorithmic decision-making.

## 4 APPLYING PROCEDURAL JUSTICE TO INTERFACE DESIGN FOR FAIR DIVISION

We explore our research question—how to make algorithmic decision-making procedurally fair—in the context of goods division in which groups make allocation decisions for themselves using an algorithm. Our focus was motivated by a real-world use case in a smart city. A city office in the US asked our research group to develop an algorithmic tool that non-profit organizations can use to fairly allocate and share resources such as food donations among themselves. With this goal, we extended Spliddit [1], a social division application that employs fair division algorithms, as the algorithmic tool. In this section, we explain Spliddit and its allocation algorithm, and our approach to designing transparency and outcome control for its interface.

### 4.1 Spliddit: Fair Division Application

Spliddit [1] is a non-profit, public website that employs fair division algorithms to find solutions to everyday division problems such as division of rent, credit, goods, and tasks. Ariel Procaccia, a computer scientist, and his students at Carnegie Mellon University developed this website. On Spliddit, a user can determine the variables that the group needs to consider to make the division. All users are then able to input their preferences for each option as individuals. Once everyone finishes inputting their preferences, they are presented with the results for the entire group. In our study, we focused on division of goods.

*4.1.1 Goods Division.* Spliddit's goods feature aims to fairly divide any kind of goods among two or more individuals [12, 40, 59]. In divisions between two people, the equitability property is met as long as both participants believe that their sets of goods have the same value. Envy-freeness is guaranteed as long as an individual's set of goods is at least as valuable to them as the other set and neither participant is willing to swap goods [40]. That is to say, envy-freeness ensures that individuals value their outcomes as much or more than the outcomes given to other individuals. There are no possible trades in such a scenario that could increase overall utility. The efficiency property is guaranteed in that the algorithm assigns goods in a way that would make it impossible for one participant to be assigned another more beneficial set of goods without making another participant worse off. In divisions between three or more people, the envy-freeness property is guaranteed along with the maximin share fairness property. A person's maximin share is the amount that they could have been given if they were allowed to divide the goods into sets, but the other participants were allowed to choose their sets before them. The algorithm guarantees that each person will receive at least  $\frac{1}{3}$  of his/her maximin share, with a greater likelihood that each individual will receive their total maximin share.

### 4.2 Designing Interfaces for Procedurally-Fair Algorithmic Decision-Making

In our study, we designed interfaces to add transparency and outcome control to the goods division algorithm on Spliddit.

*4.2.1 Transparency of Fair Division Algorithms.* We implemented two elements of transparency, standards clarity and decision outcome explanation. Throughout the design process, we did informal, iterative testing to find a design that people can best understand.

**Standards Clarity.** For fair division algorithms, the information about decision-making standards included fairness properties, types of input, and performance. Our goal was to avoid mathematical formulas or economics jargon, and show how their input was processed and how fairness

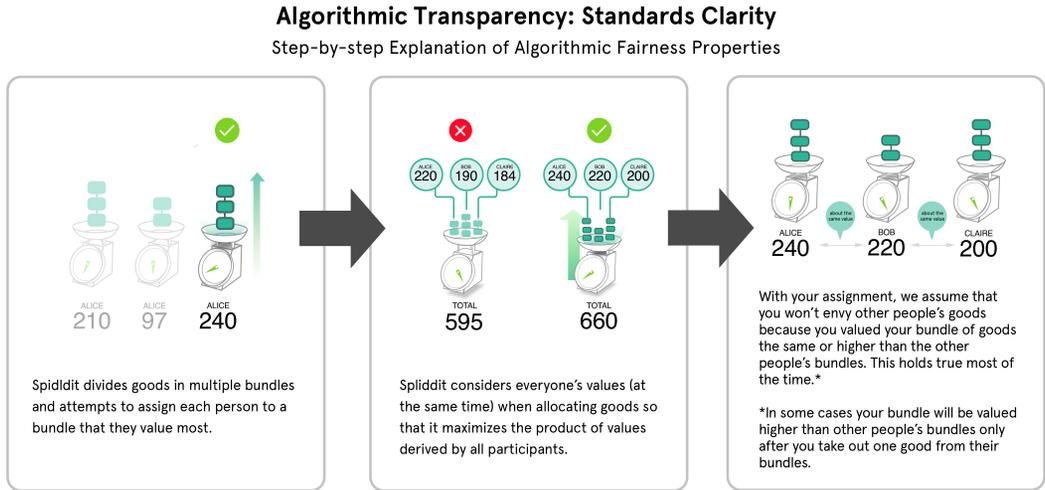


Fig. 2. Standards clarity is implemented as a step-by-step explanation of the fairness properties of the fair division algorithm in our interface.

is operationalized in the context<sup>2</sup>. To do so, our interface presented a step-by-step process that described how the algorithm processed inputs and what guarantees it provided in illustrations (Figures 2 and 3a).

**Outcome Explanation.** We chose to implement full transparency of both individual- and social-level information using a “white-box” style explanation showing input and output in a matrix table (Figure 3b). Columns represented individuals and rows represented goods. Each cell of the table showed an individual’s preference input for the corresponding goods, and highlighted cells indicated the decision outcomes, or in other words, the goods assigned to the individuals. Full transparency was suitable for this context to give people complete understanding of their own and others’ preferences so that they could use the knowledge to adjust outcomes collectively. For the input and output values, we considered whether to show the exact values or abstract symbols that indicated only relative magnitude. In this context, we chose to use exact values because people wanted to know the exact values when they saw the abstract symbols, and there was no privacy concern.

**4.2.2 Outcome Control of Fair Division Algorithms.** We implemented outcome control by allowing people to click different table cells to adjust the algorithmic allocations (Figure 3c). Our goal was to support people’s process of finding alternative outcomes if they disagreed with the algorithmic allocation. The interface showed how alternative allocations would influence the algorithm’s objective function metrics (i.e., each individual’s gained utilities from allocations), in case people wanted to see the allocation from the algorithm’s perspective. We created a two-step process of outcome control. Individual outcome control allowed each individual to create an allocation that they found suitable. The individual control process was essential so that each individual had a

<sup>2</sup>When we crafted the explanation, we avoided using the term “envy-freeness” which is economic jargon (this term may be understood differently between lay people vs. economists). We chose to keep “maximizing,” because our pretest indicated people understood the term and it conveyed the idea in a succinct manner. The interviews from our actual study also suggest that people understood the term (they used it frequently in the interview). Still, we agree that this verb may not be easily understood by members of all societies and cultures with varying levels of education.

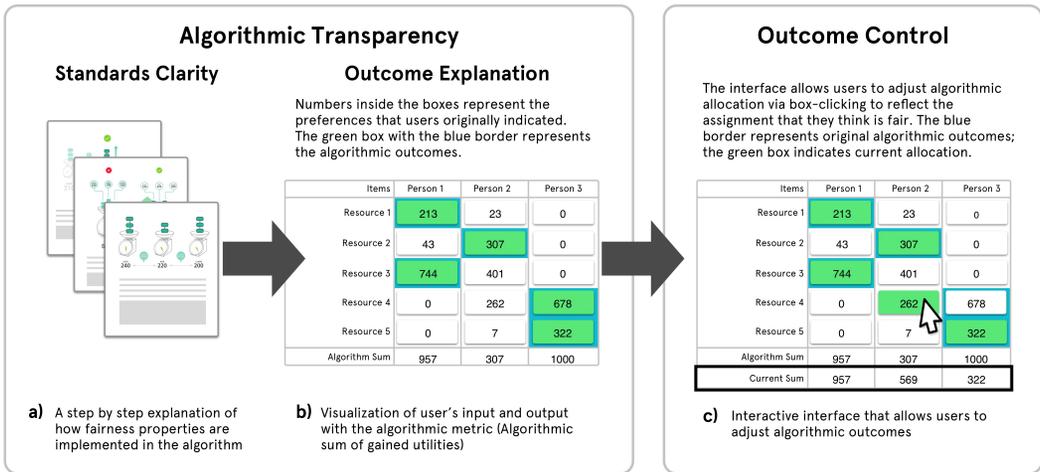


Fig. 3. Our interface first explained how the algorithm implemented the fairness properties (a. Standards Clarity), and showed the visualization of input and output with the algorithmic metric (b. Outcome Explanation). People could then adjust the algorithmic allocation (c. Outcome Control).

chance to understand the results before group discussion. Group outcome control allowed people to discuss the results and find alternative allocations that they agreed on.<sup>3</sup>

### 4.3 Interface Implementation

We used Spliddit's original interface and algorithm to collect and process user input. Our interface was connected to the Spliddit website, and visualized the algorithmic division outcomes. We used ReactJS and CSS as the primary tools to build a dynamic interface to allow users to interact with the algorithm's output. The backend system comprised of a webserver written in flask and postgres, and was hosted on Amazon Web Services.

## 5 METHOD

To evaluate the role of algorithmic transparency and outcome control in perceived fairness of decision-making, we conducted a within-subjects laboratory study in which groups of two to seven people divided food items using Spliddit and our interface.

### 5.1 Participants

Participants were recruited through a recruitment website managed by the university and through flyers posted on campus. We ran 71 participants in 23 sessions. Participants had diverse ethnicities: there were 39 Asians (or Pacific Islanders), 20 Caucasians, 7 African Americans, 2 Asian and Caucasians, and 1 other (and 2 preferred not to answer). Participants recorded an average education

<sup>3</sup>In our current study, discussion was done in person; for online discussion, different individual's alternate allocations, along with their description, could be shared in a group, and individuals could vote to determine alternative allocations.

level of 4.2 (“associate degree”=4, “bachelor degree”=5). Participants reported having basic concepts of programming on average ( $M=2.4$  ( $SD=1.1$ )) and basic concepts of algorithms ( $M=2.0$  ( $SD=1.1$ )).<sup>4</sup>

## 5.2 Task: Dividing Goods

Participants used the goods division algorithm to divide food items that they kept after the study. Participants were told before the study began that they would keep all items that were distributed as part of the study. We chose food items as goods for two reasons: people have varying levels of preferences for different food, and people can easily determine their preferences. Food items have been used in many other studies that explored people’s decision-making patterns in behavioral economics [38, 63]. In our study, participants were shown the food items that they were going to split among the group before inputting their preferences into the interface on Spliddit. To ensure that some items would be more desirable than others, we chose items with prices ranging from \$1 to \$5. Examples of food distributed in the study include 12 oz. cans of San Pellegrino sparkling fruit beverage (retail \$2 USD), Lay’s potato chips (retail \$1.50 USD) and Ghirardelli chocolate bars (retail \$5 USD). For groups of two to three individuals, three to five food items were offered. For groups of four to seven individuals, seven to twelve items were offered. This quantity variation ensured that there was always 1.5-2 times more items than the number of people in each group. Per our IRB’s recommendation, we asked incoming participants about food allergies before beginning the study. We encountered two participants who had gluten allergies and we removed all gluten-containing foods for these rounds of the study.

## 5.3 Procedures

A study session took between 40 and 70 minutes, depending on the size of the group, and each participant was compensated \$10. Participants worked in groups ranging from two to seven people. Each session could accommodate up to seven participants, and the group size was determined based on the total number of participants who signed up and showed up to the session. All participants signed the consent form prior to participating in the study and used laptops individually to use the interfaces and fill out the surveys. In order to measure the impact of transparency and control on perceived fairness, we conducted three surveys with identical measures throughout the study. We considered measuring perceived fairness after each of the transparency sub-methods, standards clarity and outcome explanation, but we chose to measure the combined main effect of transparency because the sub-methods played complementary roles; only improving standards clarity by explaining fair division properties was not effective according to [43], but it was still essential for people to understand how their input was used to devise outcomes.

**5.3.1 Baseline Phase.** No discussion was held prior to using Spliddit. Participants were given a handout with an explanation of how to use the Spliddit interface. After reviewing the handout, participants then added their inputs to Spliddit individually, without consulting other group members. Once all inputs were finalized, each participant received the results of everyone’s assignment. After receiving algorithmic outcomes, participants were instructed to fill out Survey 1.

**5.3.2 Transparency Phase.** Once everyone was done with Survey 1, the researcher asked participants to access our interface to see the algorithmic standards clarity diagram (Figure 2), followed by the input-output outcome explanation (Figure 3). The researcher also displayed the interface

<sup>4</sup>To measure programming knowledge, we used the following 4-point scale: “No knowledge at all,” “A little knowledge-I know basic concepts in programming,” “Some knowledge-I have coded a few programs before,” “A lot of knowledge-I code programs frequently.” To measure knowledge in computational algorithms, we used the following 4-point scale: “No knowledge,” “A little knowledge-I know basic concepts in algorithms,” “Some knowledge-I have used algorithms before,” “A lot of knowledge-I apply algorithms frequently to my work or I create algorithms.”

for all participants to see on a TV screen. After seeing the fair division standards and the outcome explanations, participants were instructed to fill out Survey 2.

**5.3.3 Outcome Control Phase.** Once everyone was done with Survey 2, participants were given the option to individually alter the algorithm's results to what they considered to be a fair result. Once all participants' results were finalized, the researcher took photos of each participant's changes for records. A group discussion was held immediately after. The group referred to the interface displayed on the TV screen to finalize and change the algorithm's results if they chose to do so. After the group discussion, participants were instructed to fill out Survey 3.

**5.3.4 Interview and Demographic Survey.** After participants completed Survey 3, the researcher conducted an interview with each individual in a separate room. After the interview, participants filled out the demographic survey.

## 5.4 Measures

We recorded the input of each participant, the outcomes from the Spliddit algorithm, the changes that each participant made to the algorithmic outcomes, the group discussion, and the group's final result. To measure the perceived fairness of the participants' individual and group division outcomes, we used a survey that asked participants to indicate how much they agreed or disagreed, on a 7-Likert scale, with the statement, "My assignment is fair," and, referring to each other group member, the statement, "This participant's assignment is fair." We measured perceptions of the overall fairness of the results for the whole group by asking them to indicate how much they agreed with the statement, "The overall result for the group was fair."

## 5.5 Interview

We conducted 10-15 minute long semi-structured interviews with each participant individually. We chose individual interviews, rather than a focus group, so that participants could express their opinions and report experiences without social pressures. For example, people who thought the outcomes were unfair may not have wanted to say so in front of the whole group. We had 2-3 interviewers on site depending on the group size, so 2-3 interviews were conducted simultaneously in different rooms. All interviewers used the same interview protocol and were trained through pilot studies to develop consistent interview styles. Interviews started with questions about participants' initial perceptions of their and other group members' results, before they saw our interface. We then asked about their thoughts after seeing the fair division property and outcome explanation, their experiences with the control phase, and whether and how their perception of the division outcome was influenced by them. We showed the interface on a laptop during the interview, so that the participants could interactively show us how they felt and what they did.

## 5.6 Analysis

All interviews were recorded and transcribed. Two researchers qualitatively analyzed the interview transcripts. Both coders coded all transcripts in a bottom-up analysis. We first open-coded the transcripts at the sentence or paragraph level to note factors that influenced participants' perceptions of the fairness of decision outcomes and impacts of transparency and control, which resulted in 22 concepts. We also used the survey answers to group the concepts and looked for any patterns in those whose fairness perception changed. We then synthesized the concepts into themes, which resulted in 11 high-level categories.<sup>5</sup> We used the survey answers in order to understand participants' initial

<sup>5</sup>In reporting the results, we note the number of participants who expressed corresponding themes. However, this is a small-sample study and the numbers should not be interpreted as indicators of prevalence or importance.

ratings of their outcomes, and whether participants' experiences as reported in the interviews varied depending on their initial perceptions. Overall, both individual and group outcomes were perceived somewhat fair on average (MeanIndividual=5.38 (SD=1.75), MeanGroup=4.79 (SD=1.62)). We used the fairness rating 5 ("somewhat fair") to divide the participants into two groups based on their initial perceptions: a Fair group of those who perceived the outcomes as "fair" or "very fair" and a Less Fair group of those who perceived the outcomes from "somewhat fair" to "very unfair." (Own outcome: Fair Group N=44, Less Fair Group N=27; Group outcome: Fair Group N=28, Less Fair Group N=43). We then analyzed how their fairness perception changed after each of the transparency and outcome control phases by triangulating interview themes and fairness ratings. (See Table 1 in Appendix A for fairness ratings after each phase.)

## 6 RESULTS

In this section, we describe our study results to explain the effects of transparency and control on how people perceived the fairness of algorithmic outcomes.

### 6.1 Initial Reactions Before the Interface

Initial reactions to algorithmic outcomes before using our interface were diverse, showing similar patterns documented in previous work done by Lee and Baykal [43]. Some participants were satisfied with what they received, but unsure how to judge other participants' results without knowing their preferences. Some participants were unsatisfied with their results because they had received fewer items than other participants. This decreased the fairness perception of the group as well, because they had expected a more even distribution of items. Still others initially thought their outcome was fair because they had received items that they rated highly and assumed that the group outcome was fair based on their own results.

### 6.2 Effects of Transparency

In our interface, standards clarity was designed as a description of what algorithm was used and how it worked. Many understood how the algorithm worked and how fairness in division was operationalized based on this step-by-step description, and they later used the knowledge to interpret the input and output matrix table. At the same time, participants told us that the standards clarity alone did not make them trust the algorithm or see the results as being automatically fair. On the other hand, outcome explanation, the input and output visualization, which made everyone's preferences and outcomes transparent, had a large impact on how people judged the fairness of their outcomes.

In the next sections, we first describe how transparency, particularly outcome explanation, increased perceived fairness: it allowed people to understand equalities in utility distribution and the role of individual input, and made them attribute less accountability to algorithms in distributive outcomes. We then describe how transparency decreased perceived fairness: outcome explanation made participants recognize uneven distributions and revealed differences in strategies across participants. Additionally, some participants concluded that the algorithm embodied fairness principles that they disagreed with, drawing from the algorithm process description (standard clarity) and outcome explanation.

*6.2.1 Understanding Equalities in Utility Distribution.* Even distribution of the utility in assigned items had a positive impact on overall fairness perceptions. Equality in utility distribution helped people understand the reasons for the distribution of items within the group, clarifying false assumptions about people's choices of items and dispelling other theories of how the items were allocated (n=16). Some participants initially thought their own outcomes were less fair because

they did not receive their top choice, but they increased their own fairness ratings when they saw that the distributed items were of equal value for each participant. For instance, Participant 14 (Group 5) said *“So my initial thought was more like speculation, but once I saw it I was like, ‘Okay, that makes sense’”* and *“I am satisfied just because [of] the numerical value [...], the fact that it was like evenly split and that the numbers were split up as closely as possible.”*

In several instances, seeing other participants’ preferences resulted in an increase in group fairness ratings by dispelling false assumptions of what other participants desired. For example, Participant 22 (Group 9) said: *“Before you look at other people’s preferences you kind of look at the final allocation based on your own preferences. You might form an opinion on that because you might make false assumptions about what other people want.”*

Other times, participants initially felt bad for other participants who received items that they themselves would not enjoy, but increased their perception of group fairness once group members’ preferences were made clear. For instance, Participant 27 (G8) said: *“I felt bad for [another participant] at first, because I didn’t really want [Item 1] or [Item 2], and so I thought maybe he was getting the short end of the stick on that one, but then once I saw how much he wanted each of them, I felt more comfortable with that.”*

Visualization of the relationship between their own results and the group’s input in our interface also increased group fairness perceptions. Participant 25 (Group 10) stated *“Even though I knew the assignment was fair for me, I didn’t know if it was fair for him or not, because I didn’t know if he actually valued [Item] the most. But after seeing the chart and all the values, I realized it was a pretty fair assignment for both of us.”*

Without explanations, even people who got more items thought the outcome was unfair, as they created their own theories to explain their results for themselves. For instance, Participant 15 (Group 6) originally thought she had received too many items, and wondered if the other person *“gave it all to me because I was a woman?”* But then she saw that the other participant actually only wanted specific kinds of items. Transparency increased understanding of what each participant wanted, thus increasing perceptions of group fairness.

**6.2.2 Understanding the Role of Individual Input.** Showing how input preferences translated into results also made people realize their roles in the division outcomes. For many, seeing this connection was enough for people to feel accountable for the outcome, which increased their perception of the fairness of their own outcome (n=42). For instance, Participant 57 (Group 20) initially thought her result was less fair because she got her first choice and one that she liked least (Item 1) instead of her second choice; but after seeing the interface, she said: *“It looked really fair and as soon as I saw the screen [...] I looked at [other participant’s] number for (Item 1), and he put zero. [...] I was fine with it, because 50 [her preference input] means that I would be willing to take it. If I didn’t really want it, I would put zero, like [other participant] did.”*

**6.2.3 Algorithms Becoming Less Accountable.** The increased accountability of the participants themselves decreased the accountability of the algorithm. For example, some participants concluded that unsatisfactory results were not the algorithm’s fault, but rather the fault of participants during the input phase (n = 12). Participant 20 (Group 8) said, *“Because there was no real way I could move anything around without reducing the score, so that means that if there was any perceived unfairness it would be not because of the algorithm, it would be because one of us or the other rated the objects incorrectly.”*

Participant 24 (Group 10) considered the algorithm fair in computing numbers, but said that *“if you feel it’s unfair, that’s your fault,”* referring to the input phase. *“Once you are the one who’s picking the value, the algorithm’s just trying to maximizing it for you. If something goes wrong or if you feel it’s unfair, that’s your fault.”* Participant 7 (Group 3) also noted that the group’s input had

more influence on the allocation than the algorithm itself: *“I think everyone theoretically should have been pretty satisfied with their choice. I think if they weren’t satisfied, it had more to do with how they were making their [own] allocation decisions rather than how the algorithm worked out.”*

The increased role of their own preferences was sometimes even seen as an absence of a mediator or decision-maker. Participants felt there was no one to blame, so they considered the outcomes acceptable. Participant 69 (Group 23) said: *“Had we had this discussion right from the scratch, and had a person allocated this, people would not have been that satisfied. But just because maybe they didn’t have a person to, like, blame or something, they were like, ‘Okay, fine.’”*

**6.2.4 Understanding uneven distributions.** Transparency also decreased perceived fairness in some cases. Generally, uneven distributions of utilities for the group were perceived to be unfair, as participants valued equality across the board (n=30). Comparisons of utility sums across the board exposed the fact that some people received more or fewer points than others for items received. Participant 23 (Group 9) stated: *“I think that it’s fair if you just compare [other participant] and me. However, when looking at [another participant], her score is one third of our score or about one third. So just looking at the numbers, it doesn’t look fair at all, at least for [another participant].”*

Even participants who initially found their assignment fair became less satisfied when the uneven distribution was revealed. For instance, Participant 29 (Group 12) was initially somewhat satisfied with her items, but upon realizing her sum of utilities was much lower than others, decreased her own fairness perceptions. Fairness perceptions also decreased in the case of participants who received their first-choice items (n=15). Without the interface, they tended to judge both their own and the group outcome as fair because they assumed other group members would have received their first choice items as well. When they saw the outcome explanation and realized that they had gotten more utility points than the others, their perceived fairness decreased. For example, when Participant 62 (Group 22) received the most points for his assignment, while someone else got much lower than he did, he said: *“The algorithm favored my choices over the others.”*

**6.2.5 Revealed Differences in Input Strategies.** The interface also revealed that people in the group might have used different input strategies. While some chose to spread out their points across the items, others allocated all their points to one or two items, leaving “0” points for some items. Skewed inputs could have led to skewed results that did not accurately represent each individual’s preferences. This was revealed when the outcome was explained and lowered perceived fairness outcomes for some participants. For example, Participant 9 (Group 3) noted that “I was surprised that everyone else had a zero except for me” when the inputs were revealed. Participant 61 (Group 21) also noted that part of why his group fairness rating dropped when seeing the interface was the fact that he saw that each person had a different strategy in inputting his or her values. He stated: *“So [other participant] and I both got two [items] for each just because we devoted [points] to several ones but [another participant] just got one.... We had different strategies but I don’t think the strategy should make such a different result.”*

**6.2.6 Different Concepts of Fairness.** Understanding of algorithms sometimes decreased fairness perceptions for people whose fairness concepts were different from the algorithm’s assumption (n=8). As described in the algorithm process explanation (standards clarity), the fair division algorithm sought to maximize the product of all participants’ gained utilities, but Participant 65 (Group 23) valued equality in distribution more: *“I think what [other participant] was trying to do, made a little more sense than this algorithm, which was kind of reduce the range of like the maximum/minimum points. But this one ... was trying to make everyone’s points similar, but then a second factor they also put in was trying to maximize the product of everyone’s values? I guess mathematically that is the same thing as trying to make the average the same, but just like if there*

was a way to just primarily just decrease the overall range, even if the values themselves are not as high as they could be.” This is an interesting finding as this participant did not think the algorithm’s fairness standards itself were unfair initially. Only after seeing the algorithm’s outcomes and that the algorithm, following the standard, could make uneven material distributions across participants did he realized that the algorithm was not in line with his beliefs.

Participants sometimes disagreed or felt uncomfortable with the algorithm’s assumptions of mathematical fairness. Participant 32 (Group 13) was originally somewhat satisfied, but was unsatisfied when he realized that the assignment was based purely on technical fairness. He noted that his way of allocating values was different from how another participant had allocated his values to one item in order to get what he really wanted: *“The whole fairness thing, there’s two parts to it that I can understand. There’s a technical fairness. I read the About page, there’s this technical fairness that you can calculate, but there’s also a sense of fairness, like how you feel that it’s fair. I definitely agree that it’s technically fair and I can’t argue about that because it’s literally like an equation or formula but it never really felt fair because I never got to tell them what I want. There’s no human interaction.”* Participant 54 (Group 19) also said understanding how the algorithm worked made him uncomfortable with the idea of basing decisions purely on the basis of numbers: *“I did kind of understand the logic behind the allocation but an assumption that two items on one side equals one item on the other may not be the best way to divide things up— purely on the basis of numbers [it] seems misleading ... the want should be the final decider.”* Additionally, Participant 23 (Group 9) said, *“If anything they made me feel very negative in the beginning about how somebody was losing out just because a computer told me that these are the values.”*

### 6.3 Effects of Control

Outcome control, or having a chance to explore the algorithmic allocation by changing one’s own or another group member’s results, caused participants to perceive algorithmic outcomes as fair. 22% of the groups actually changed their final outcomes; but all groups, regardless of the actual changes, thought that algorithmic allocation was fair after having a chance to adjust their outcomes.

**6.3.1 Realization of Inherent Limitations.** Control of the algorithm’s allocation surfaced the inherent limitations of possible division outcomes given the resources and the people who were in the group. Many participants noted that through exploration, they realized that the algorithmic result might be the best allocation, which increased their overall fairness perceptions ( $n=17$ ). For example, Participant 29 (Group 12) said: *“I was just thinking about what the totals of the bottom would be, if I changed them and it was just basically like I couldn’t, like the total of all of them would still be lower if I changed anything. So it was like, ‘Okay, I guess there’s not a better way to arrange it.’”* For Participant 57 (Group 20), control over allocation made her realize that there was no other way to improve the results: *“I think by trying to change [the original allocation], it makes you even more solid about admitting the result.”*

**6.3.2 Redistribution Through Discussion.** Individuals used the interface as a basis for group discussion to redistribute the sum of utilities ( $n=22$ ). Participant 23 (Group 9) indicated that the interface provided a baseline, while the group discussion allowed human interaction to generate the final decision: *“I think that changing the algorithm is very fair because the algorithm doesn’t have feelings. It doesn’t know what we’re thinking. So it can’t really tell us how we value our choices. But it can give us something to work from.”*

Across groups, participants sought to redistribute to satisfy the participants who received lower utility sums. Many participants attempted to be “spokespeople” for people who received the lowest sum of utilities and explained how they could changed the results. The interface helped groups to explore other ways to allocate resources according to people’s preferences: *“If we’d just seen the*

results, the two, two, and one distribution,... it would have been more difficult to find possible choices to satisfy people, yes. I think having this grid, this, uh, spreadsheet in front of us was a help" (Participant 63 (Group 22)).

**6.3.3 The Human Element.** Discussion itself had a positive effect on fairness perceptions because of the human element of acknowledgement (n=26) and understanding of others' preferences on a deeper level (n=45). For instance, Participant 32 (Group 13) said: "Once they were talking about it, that's when things started to feel more fair because there's people behind the words." Participants reported that discussion clarified the group's preferences on a deeper level. Participant 31 (Group 12) said: "When the group agrees (verbally) it feels more fair than when the group is just told. So it kind of gave a deeper sense of okay, we both like this thing. Maybe we should find an answer outside of the algorithm that would work better." Participant 6 (Group 2) also mentioned that the reasons behind preferences were disclosed through the discussion phase: "So after the discussion, whenever for example [other participant] said that she wanted the [item] for her flight, I was like well she should definitely have that then because she's got a real reason to want [item]... So the discussion really helped because ... you know why like those values came from."

Discussion helped participants to understand the altruistic intentions of other participants even when they did not materialize in the results (n=8). Although Participant 12 (Group 2) did not receive her most-wanted item, she found it kind that another group member offered her two other items to compensate her for her low utility sum: "[S]he was willing to offer me the other two items. Yeah, so I guess that kind of helped even though I didn't want them."

## 6.4 Binding Effects of Outcome Explanations

While algorithmic outcome metrics allowed people to understand algorithmic decisions, they also had undesirable binding effects on discussion. Some individuals found it difficult to go against the algorithmic outcome metrics (n=11). In Group 19, a person who got one item wanted to give it to someone who valued it more. Even though both the giver and the receiver agreed on it, another person commented that this would significantly lower the giver's utility sum. The comment made the receiver uncomfortable, even though the giver was fine with that outcome, and in the end, the group didn't change the allocation. On the other hand, some participants found it difficult to express their dissatisfaction during their group discussion because their algorithmic sum or points allocation stated otherwise (n=7). Participant 35 (Group 9) did not receive what she wanted, but she found it challenging to articulate her dissatisfaction because her values stated that she had gotten her second choice, which was still considered highly ranked: "I felt okay with [item] because everyone could see that it was my second choice in the product list and what second choice numerically that I picked."

## 7 DISCUSSION

In this section, we discuss our findings and their implications for emerging research on algorithmic transparency, explanation, and fairness. We lay out questions and opportunities for future research.

### 7.1 Summary of Findings

We implemented a procedurally fair process for fair algorithmic goods division as a case study for our procedural justice framework for algorithmic decision-making. For transparency, our interface communicated algorithmic rules using a step-by-step diagram (standards clarity), then explained algorithmic outcomes by displaying inputs and outputs (outcome explanation). For control, individuals and groups used the interface to explore, discuss, and ultimately choose alternative outcomes (outcome control).

The results from a within-subjects study suggest that standards clarity and outcome explanation allowed people to judge whether the fairness properties of the algorithm were in line with their fairness concepts. While a fairness property, such as envy-freeness, is difficult to refute, the step-by-step explanation of algorithmic calculations allowed some people to exactly pinpoint where their beliefs differed from the algorithmic assumptions (i.e., instead of maximizing the sum of the utility and finding the solution that minimizes within-differences in division, strive for a more equal distribution of the utilities). Outcome explanations helped people understand why the algorithm made certain allocations and whether the allocations were correct from their or their group's perspectives; however, this had mixed effects, increasing or decreasing perceived fairness and reducing algorithmic accountability.

With outcome control, participants thought their outcomes were fair regardless of whether they actually changed the outcomes or not. People discovered the inherent constraints of the problem as they tried to find alternate outcomes themselves, thus coming to understand the algorithm's unequal utility distributions. Discussion gave people the opportunity to voice their concerns and understand other people's benevolent intentions and reasons for their preferences, and allowed them to change the algorithm's outcomes to make them fairer for their groups (22% of the groups).

## 7.2 Differentiating Standards Transparency from Outcome Transparency

Our findings suggest that more nuanced definitions of transparency, specifically those that clarify whether it is transparency of algorithmic standards or of outcomes, can further unpack the role of algorithmic transparency. In previous work, common ways to implement algorithmic transparency are “how” explanations, describing systems' inputs and outputs, and “why” explanations, providing justification for outcomes [39, 61], and they have shown mixed effects. For example, in Rader's study, “how” explanations did not improve judgments of the “correctness” of news feed algorithms [61]; yet in our study “how” explanations helped people judge the correctness of allocation outcomes. One potential explanation for this is that “how” explanations in [61] described the algorithmic rules and the standard used in news feed curation, while the correctness measure was oriented toward newsfeed outcomes. Our findings suggest that the “how” explanation of the allocation algorithm (standards clarity) helped people understand and make judgments as to whether they agreed or disagreed with the standard itself, but were less helpful in judging their specific outcomes. On the other hand, “how” outcome explanations of specific inputs and outputs helped participants judge the fairness of the outcomes. Previous work done by Kizilcec [39] made a distinction between levels of transparency, where medium transparency was implemented as transparency of standards (i.e., how an algorithm fairly synthesized peer-reviewed grades), and high transparency was implemented as transparency of outcomes (i.e., showing the original peer-reviewed grades and their algorithmic adjustment). In his study, standards transparency improved trust while the outcome explanation lowered trust among those who did not get the grades they expected. Distinction between standard and outcome explanation may help us synthesize these different findings and design future research that distinguishes the mixed impacts of transparency.

## 7.3 Unexpected Effects of Transparency

Our study offer considerations for introducing transparency into algorithmic systems.

*7.3.1 Misalignment between Human versus Algorithmic Evaluation.* Expectation violation, or the difference between predicted versus actual outcomes, has been highlighted as an important factor in understanding people's responses to algorithmic outcomes [39]. Our study adds to this work by showing that misalignment in human versus algorithmic evaluation is another point where expectation violation can occur. Showing algorithmic evaluation metrics, the utility sums, decreased

the fairness judgments of some participants who initially evaluated their outcomes to be satisfactory but received lower utility sums than others, and of those who initially assumed that everyone got their first choice because they themselves got theirs, and then realized this was not the case. This highlights the importance of and complexities in understanding people's dynamic responses to algorithmic evaluation and managing their expectations.

*7.3.2 Outcome Explanation without Control.* Outcome explanation played a critical role in helping people understand whether the algorithm worked in the way it was intended, which is one of the key potential values of transparency [5]; however, seeing outcome explanations alone did not lead participants who initially thought that their outcomes were unfair to feel that their outcomes were fairer. This suggests that making algorithmic decision-making systems transparent without giving users a sense of control or voice may backfire against the intention of transparency, as “transparency is disconnected from power” [5].

## 7.4 Materialization of Transparency and Control in Design

Our findings offer empirical evidence that the materialization of transparency and control in design influences perceptions of algorithmic decisions. Depending on the design choices, transparency and control can help the algorithmic service achieve its ideal potential, or create harmful or deceitful effects [5]. We reflect on four design choices that promote transparency and control that seem to have influenced people's accountability attributions, fairness judgments, and willingness to adhere to algorithmic decisions. We discuss questions and opportunities future research should address to produce systematic knowledge on different transparency and control designs that can produce positive effects across different contexts.

*7.4.1 Communicating Difficulties in Algorithmic Decision-Making.* In our study, interactive exploration of algorithmic outcomes, particularly alternative allocation outcomes, made people realize the inherent constraints in some decisions, thereby making people perceive the same algorithmic outcomes as fairer compared to their initial assessment. This finding suggests a new kind of transparency to consider in order to increase fairness perceptions. For example, communicating the difficulty of the algorithm's decision task (e.g., the number of data points that it needed to go through, alternative solutions it computed, or sizes of solution spaces in relation to particular inputs) or giving people the experience of making the decisions themselves can improve their adoption of the decisions. However, depending on the algorithms, interactive exploration designs may become too complex, making problems seemingly more complex or difficult than they are and creating the illusion that an algorithm is fair [5]. In accordance with Abdul et al.'s review and proposal of an HCI research agenda, future research should explore intuitive and interactive ways of exploring algorithmic decisions [3].

*7.4.2 Transparency Design that Influences Algorithmic Accountability.* Our study visualized the inputs and outputs of algorithms; this visualization helped people understand the exact role of their input and acknowledge their own responsibility for the outcomes. This made them perceive the outcomes as fairer. This realization shifted most of the responsibility for the allocation from the algorithm to the participants. Before this shift, participants viewed the algorithm as solely responsible for division outcomes and thus blamed it for unfair divisions [43]. This shift in accountability can be more or less desirable depending on the algorithm's actual performance. In retrospect, we believe our design choice of using a matrix format that shows inputs and outputs may have magnified this pattern. While the format allowed people to explore alternative outcomes, it did not visually show the algorithms, making their presence less salient. Future studies need to look at this dynamic between implementation of the “how” (input-output) explanation and accountability shift,

so that we can create an ideal balance of input data versus algorithm accountability depending on the context.

*7.4.3 Numeric Representation of Algorithmic Judgments.* Our interface showed actual numeric values of inputs and algorithmically driven utilities. While this allowed for exact comparisons of people's different utility sums, it also seemed to have binding effects, making it difficult for some people to explore other outcomes or override the algorithmic utilities. Our choice of numbers may have led people to pay attention to sometimes insignificant numeric differences (e.g., a 50-point difference in a total of 1,000 points). Future research should explore other ways of representing algorithmic evaluation (i.e., gained utilities or scores) in order to create balance between people's ability to understand algorithmic evaluations and their willingness to override algorithmic judgments. For example, more abstract representations (such as sizes of circles) may convey the magnitude and direction of changes, yet does not bind people in minor numerical differences.

*7.4.4 Levels of Social Transparency & Algorithmic Fairness Judgments.* We implemented full outcome transparency by showing people the entire group allocation results as well as the algorithmic evaluation. In our context, this full transparency helped participants judge fairness based on actual overall distribution rather than one's initial, often self-centered assumptions alone, so that actual injustices could be identified and addressed. We acknowledge that full social transparency (e.g., exact display of numeric inputs and algorithmic outcome utilities) is neither feasible nor appropriate in all settings, particularly in commercial sectors. On the other hand, our work suggests that social transparency is an important element of fairness perception in algorithmic decisions. Previous work shows that social comparison of outcomes is a key element in whether and how people judge fairness and justice [82]. In a recent investigation of different experimentation styles, Binns et al. [8] found that a lack of social comparison might be one of the factors that contributed to the similar impacts of explanation styles that all focused on individuals' own outcome explanations. Future research can investigate how social transparency can be adopted into outcome transparency in a way that balances the accuracy of fairness judgments and the privacy of other involved parties. This decision should be made depending on the importance of instilling a sense of fairness of outcomes in the decision context, the perceived benevolence of the algorithmic systems (or institutions that use the algorithmic systems), and the performance and stability of the algorithmic systems—For example, in an early stage, a transparent system is more desirable to catch edge cases; when it is tested thoroughly, the system can be closed and opaque.

## 7.5 Social Research to Complement Computational Fairness Properties

Finally, our findings suggest that research that develops computational methods for “fair” decisions should be complemented with research that investigates social and contextual factors that are critical to making these fair algorithmic systems work in the real world. Veale et al.'s recent work describes social practices in public sectors that seek to make algorithms perform fairly [83]. Our work echoes the importance of their findings. In their work, many developers of the tool were also the users or the decision-makers, or there was close interaction between the two communities, which eases the process of making and testing algorithmic systems that are appropriate for different contexts and use cases. However, as algorithmic decision tools become more prevalent, there will be contexts and institutions where developers of the tools are not the main users and decision-makers using the systems (one prevalent example is COMPAS, used by US courts to predict defendant recidivism). In this case, clearly describing how “fairness” is implemented in the algorithm will be important so that people can decide whether that particular operationalization of fairness fits their context. Additionally, our work suggests the importance of preserving a human element in the process by providing people opportunities to voice their concerns and appeal or change their

outcomes. Creating a channel for these concerns demonstrating the benevolence of the decision makers will be critical to enable decisions that are both socially and computationally fair.

## 8 LIMITATIONS

Like any study, our study has several limitations that readers should keep in mind. As a first step in exploring the space of procedurally fair decision-making processes for algorithmic decision-making, we focused on one fair division algorithm and one interface design for a goods division task in a context where collocated groups made allocation decisions for themselves. The present study was also done in a laboratory with a within-subjects design for a one-time decision with a low-stakes task. Future studies should examine whether our findings hold true in different contexts—for example, decisions over time, or with different resources or high-stakes tasks—and with other study designs, such as between-subjects, or settings that involve explicit decision-makers.

Given the difference between white-box and black-box algorithms, our findings are unlikely to contribute to black-box algorithms such as neural networks. Still, we believe that our findings can be applicable to interpretable machine learning algorithms that computationally guarantee fairness properties such as individual or meritocratic fairness because these principles can be explained to users. In this context of machine learning algorithms, further research would need to be done to account for people's understanding of predictive algorithmic outcomes and associated uncertainty.

We used groups of varying sizes because previous work that explored the same algorithmic system did not find an impact of group size on participants' perceptions of algorithmic decisions [43] and because all groups were still relatively small. However, having a fixed number of individuals in each group session would provide even more control. Additionally, although our interviews and demographic surveys do not indicate that any of our participants were experts in social choice theory, economics, or fair division algorithms, we acknowledge as a limitation that we did not explicitly measure participants' knowledge of fair division algorithms. Further delineating the role of understanding of the general algorithm process versus social transparency is important future work. Finally, representation of the Latinx community is limited both in our study and in the geographic region in which the study was conducted. Future work in this area should seek to address this limitation as pioneered in [86].

## 9 FUTURE WORK

Future work should explore how our algorithmic procedural justice framework can be extended in different domains and for different kinds of algorithms. For example, the framework can be tested in a setting in which people's shifts and tasks are algorithmically divided based on people's preferences—for example, schedule and task assignment for healthcare, retail, and restaurant workers. These contexts are similar to the context of our study, and will allow an examination of the effects in a naturalistic setting. We will also examine what effects the framework and design strategies have on the perceived fairness of algorithms that moderate social media by flagging content, banning accounts, or hiding someone's content when inappropriate. This line of work can explore the design of interfaces that vary standard and outcome transparency and control to further delineate the effects of the strategies. It can also investigate the role of other components of our procedural justice framework such as the "perceived benevolence of organizations and decision-makers" (Figure 1c) and "process control [45] (Figure 1b)."

In exploring other domains, one would need to examine the context in which the algorithmic outcome is a prediction (e.g., predicted risk), rather than utility distribution. In prediction settings, other kinds of information, such as confidence level, may influence perceived fairness. We believe

that cognitive overload is another important consideration for any future work. Usable and accessible information is a prerequisite for transparency to be effective. In domains where this may not be possible, promoting procedural justice could be difficult.

It will also be important to examine the boundary conditions for our framework. For example, there could be an interaction effect of transparency and control, meaning that having one without the other versus having both could have different effects. Additionally, there could be an effect of repeat participation; for example, post-transparency participants, equipped with knowledge of how the algorithm works, may attempt to manipulate the results using these rules, or interpret the outcomes for fellow users differently than they did in the first round.

Finally, it is necessary to explore the tension between the positive and negative consequences of framework elements when the framework is used in practice. For example, once algorithm creators understand that transparency can hurt perceptions of fairness, particularly among those who receive less utility, this knowledge may incentivise them to either improve treatment of those with lower utility outcomes or adaptively reduce transparency for them. Future work should explore ethical and policy issues that need to complement psychological understanding of perceived fairness.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation CNS-1651566 grant, the Uptake & CMU Machine Learning for Social Good grant, and the CMU Block Center for Technology & Society grant. We thank anonymous reviewers for providing helpful comments and our research participants for sharing valuable insights with us.

## REFERENCES

- [1] [n.d.]. Provably Fair Solutions. <http://www.spliddit.org/>
- [2] [n.d.]. Stanford Participatory Budgeting Platform. <https://pbstanford.org/>
- [3] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 582.
- [4] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. 2014. Power to the people: The role of humans in interactive machine learning. *AI Magazine* 35, 4 (2014), 105–120.
- [5] Mike Ananny and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20, 3 (2018), 973–989. <https://doi.org/10.1177/1461444816676645>
- [6] Robert J Bies, Christopher L Martin, and Joel Brockner. 1993. Just laid off, but still a “good citizen?” Only if the process is fair. *Employee Responsibilities and Rights Journal* 6, 3 (1993), 227–238. <https://doi.org/10.1007/BF01419446>
- [7] Reuben Binns. 2018. Fairness in machine learning: Lessons from political philosophy. *Proceedings of Machine Learning Research* 81 (2018), 1–11.
- [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. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 377.
- [9] Steven L Blader and Tom R Tyler. 2003. A four-component model of procedural justice: Defining the meaning of a “fair” process. *Personality and Social Psychology Bulletin* 29, 6 (2003), 747–758. <https://doi.org/10.1177/0146167203029006007>
- [10] Steven J Brams, Michael A Jones, Christian Klamler, et al. 2006. Better ways to cut a cake. *Notices of the AMS* 53, 11 (2006), 1314–1321.
- [11] Joel Brockner, Tom R Tyler, and Rochelle Cooper-Schneider. 1992. The influence of prior commitment to an institution on reactions to perceived unfairness: The higher they are, the harder they fall. *Administrative Science Quarterly* (1992), 241–261. <https://doi.org/10.2307/2393223>
- [12] Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. 2016. The unreasonable fairness of maximum Nash welfare. In *Proceedings of the 2016 ACM Conference on Economics and Computation*. ACM, 305–322. <https://doi.org/10.1145/2940716.2940726>

- [13] W Chan Kim and Renée Mauborgne. 1998. Procedural justice, strategic decision making, and the knowledge economy. *Strategic management journal* 19, 4 (1998), 323–338. [https://doi.org/10.1002/\(SICI\)1097-0266\(199804\)19:4<323::AID-SMJ976>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-0266(199804)19:4<323::AID-SMJ976>3.0.CO;2-F)
- [14] Jason A Colquitt and Jessica B Rodell. 2015. Measuring justice and fairness. *Oxford handbook of justice in the workplace* 187 (2015), 202.
- [15] Sam Corbett-Davies and Sharad Goel. 2018. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023* (2018).
- [16] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wierlinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455. <https://doi.org/10.1007/s11257-008-9051-3>
- [17] Harlon Leigh Dalton. 1985. Taking the right to appeal (more or less) seriously. *The Yale Law Journal* 95, 1 (1985), 62–107.
- [18] 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>
- [19] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic transparency in the news media. *Digital Journalism* 5, 7 (2017), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
- [20] Jonathan Dodge, Q Vera Liao, Yunfeng Zhang, Rachel KE Bellamy, and Casey Dugan. 2019. Explaining models: an empirical study of how explanations impact fairness judgment. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, 275–285.
- [21] Christopher Donner, Jon Maskaly, Lorie Fridell, and Wesley G Jennings. 2015. Policing and procedural justice: a state-of-the-art review. *Policing: an international journal of police strategies & management* 38, 1 (2015), 153–172. <https://doi.org/10.1108/PIJPSM-12-2014-0129>
- [22] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS)*. ACM, 214–226.
- [23] FAT\*. [n.d.]. ACM Conference on Fairness, Accountability, and Transparency.
- [24] Robert Folger. 1977. Distributive and procedural justice: Combined impact of voice and improvement on experienced inequity. *Journal of personality and social psychology* 35, 2 (1977), 108. <https://doi.org/10.1037//0022-3514.35.2.108>
- [25] Robert Folger and Mary A Konovsky. 1989. Effects of procedural and distributive justice on reactions to pay raise decisions. *Academy of Management journal* 32, 1 (1989), 115–130. <https://doi.org/10.2307/256422>
- [26] Bryce Goodman and Seth Flaxman. 2016. EU regulations on algorithmic decision-making and a right to explanation. In *ICML workshop on human interpretability in machine learning (WHI 2016)*, New York, NY. <http://arxiv.org/abs/1606.08813> v1.
- [27] Jerald Greenberg. 1990. Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of applied psychology* 75, 5 (1990), 561. <https://doi.org/10.1037/0021-9010.75.5.561>
- [28] Jerald Greenberg and R Cropanzano. 1993. The social side of fairness: Interpersonal and informational classes of organizational justice. *Justice in the workplace: Approaching fairness in human resource management*. Hillsdale, NJ: Lawrence Erlbaum Associates (1993).
- [29] Nina Grgić-Hlača, Elissa M Redmiles, Krishna P Gummadi, and Adrian Weller. 2018. Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. *arXiv preprint arXiv:1802.09548* (2018).
- [30] Nina Grgić-Hlača, 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*, Vol. 1. 2.
- [31] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*. ACM, 305–318. <https://doi.org/10.1145/2663716.2663744>
- [32] F Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting users in control of their recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 3–10. <https://doi.org/10.1145/2792838.2800179>
- [33] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.
- [34] Tad Hirsch, Kritzia Merced, Shrikanth Narayanan, Zac E Imel, and David C Atkins. 2017. Designing contestability: Interaction design, machine learning, and mental health. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. ACM, 95–99. <https://doi.org/10.1145/3064663.3064703>
- [35] Pauline Houlden, Stephen LaTour, Laurens Walker, and John Thibaut. 1978. Preference for modes of dispute resolution as a function of process and decision control. *Journal of Experimental Social Psychology* 14, 1 (1978), 13–30. [https://doi.org/10.1016/0022-1031\(78\)90057-4](https://doi.org/10.1016/0022-1031(78)90057-4)

- [36] Farnaz Jahanbakhsh, Wai-Tat Fu, Karrie Karahalios, Darko Marinov, and Brian Bailey. 2017. You Want Me to Work with Who?: Stakeholder Perceptions of Automated Team Formation in Project-based Courses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3201–3212. <https://doi.org/10.1145/3025453.3026011>
- [37] Judy Kay and Bob Kummerfeld. 2012. Creating personalized systems that people can scrutinize and control: Drivers, principles and experience. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 4 (2012), 24. <https://doi.org/10.1145/2395123.2395129>
- [38] Uzma Khan and Ravi Dhar. 2007. Where there is a way, is there a will? The effect of future choices on self-control. *Journal of Experimental Psychology: General* 136, 2 (2007), 277. <https://doi.org/10.1037/0096-3445.136.2.277>
- [39] René F Kizilcec. 2016. How much information?: Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2390–2395. <https://doi.org/10.1145/2858036.2858402>
- [40] Flip Klijn. 2000. An algorithm for envy-free allocations in an economy with indivisible objects and money. *Social Choice and Welfare* 17, 2 (2000), 201–215. <https://doi.org/10.1007/s003550050015>
- [41] Todd Kulesza, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. 2013. Too much, too little, or just right? Ways explanations impact end users' mental models. In *2013 IEEE Symposium on Visual Languages and Human Centric Computing*. IEEE, 3–10. <https://doi.org/10.1109/VLHCC.2013.6645235>
- [42] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 1–16. <https://doi.org/10.1177/2053951718756684>
- [43] Min Kyung Lee and Su Baykal. 2017. Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 1035–1048.
- [44] Min Kyung Lee, Ji Tae Kim, and Leah Lizarondo. 2017. A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3365–3376.
- [45] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, Alexandros Psomas, and Ariel Procaccia. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM : Human-Computer Interaction* 3, CSCW (2019), Article 181, 35 pages.
- [46] Min Kyung Lee, Daniel Kusbit, Evan Metsky, and Laura Dabbish. 2015. Working with machines: The impact of algorithmic and data-driven management on human workers. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*. ACM, 1603–1612. <https://doi.org/10.1145/2702123.2702548>
- [47] Gerald S Leventhal. 1980. What should be done with equity theory? In *Social exchange*. Springer, 27–55. [https://doi.org/10.1007/978-1-4613-3087-5\\_2](https://doi.org/10.1007/978-1-4613-3087-5_2)
- [48] Brian Y Lim, Anind K Dey, and Daniel Avrahami. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2119–2128.
- [49] E Allan Lind, Ruth Kanfer, and P Christopher Earley. 1990. Voice, control, and procedural justice: Instrumental and noninstrumental concerns in fairness judgments. *Journal of Personality and Social psychology* 59, 5 (1990), 952. <https://doi.org/10.1037//0022-3514.59.5.952>
- [50] E Allan Lind, Carol T Kulik, Maureen Ambrose, and Maria V de Vera Park. 1993. Individual and corporate dispute resolution: Using procedural fairness as a decision heuristic. *Administrative Science Quarterly* (1993), 224–251. <https://doi.org/10.2307/2393412>
- [51] E Allan Lind, Robin I Lissak, and Donald E Conlon. 1983. Decision Control and Process Control Effects on Procedural Fairness Judgments 1. *Journal of Applied Social Psychology* 13, 4 (1983), 338–350. <https://doi.org/10.1111/j.1559-1816.1983.tb01744.x>
- [52] E Allan Lind and Tom R Tyler. 1988. *The social psychology of procedural justice*. Springer Science & Business Media.
- [53] E Allan Lind, Tom R Tyler, and Yuen J Huo. 1997. Procedural context and culture: Variation in the antecedents of procedural justice judgments. *Journal of Personality and Social Psychology* 73, 4 (1997), 767. <https://doi.org/10.1037//0022-3514.73.4.767>
- [54] John Stuart Mill. 2016. Utilitarianism. In *Seven masterpieces of philosophy*. Routledge, 337–383.
- [55] Charles E Miller, Patricia Jackson, Jonathan Mueller, and Cynthia Scherschling. 1987. Some social psychological effects of group decision rules. *Journal of Personality and Social Psychology* 52, 2 (1987), 325. <https://doi.org/10.1037//0022-3514.52.2.325>
- [56] Brian P Niehoff and Robert H Moorman. 1993. Justice as a mediator of the relationship between methods of monitoring and organizational citizenship behavior. *Academy of Management journal* 36, 3 (1993), 527–556. <https://doi.org/10.2307/256591>
- [57] Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67. <https://doi.org/10.1016/j.ijhcs.2015.01.007>

- [58] Richard Arthur Posthuma. 1999. The effect of context on the multiple dimensions of procedural justice. (1999).
- [59] Ariel D Procaccia and Junxing Wang. 2014. Fair enough: Guaranteeing approximate maximin shares. In *Proceedings of the fifteenth ACM conference on Economics and computation*. ACM, 675–692. <https://doi.org/10.1145/2600057.2602835>
- [60] Dean G Pruitt, Robert S Peirce, Neil B McGillicuddy, Gary L Welton, and Lynn M Castrianno. 1993. Long-term success in mediation. *Law and Human Behavior* 17, 3 (1993), 313–330. <https://doi.org/10.1007/BF01044511>
- [61] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 103.
- [62] John Rawls. 2009. *A Theory of Justice*. Harvard University Press.
- [63] Daniel Read and George Loewenstein. 1995. Diversification bias: Explaining the discrepancy in variety seeking between combined and separated choices. *Journal of Experimental Psychology: Applied* 1, 1 (1995), 34. <https://doi.org/10.1037/1076-898X.1.1.34>
- [64] John E Roemer. 1998. *Theories of distributive justice*. Harvard University Press.
- [65] Maurice Salles. 2017. Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel Procaccia (eds), Handbook of Computational Social Choice. *(Economia. History, Methodology, Philosophy* 7-4 (2017), 609–618. <https://doi.org/10.1017/CBO9781107446984>
- [66] Nripsuta Ani Saxena, Karen Huang, Evan DeFilippis, Goran Radanovic, David C Parkes, and Yang Liu. 2019. How Do Fairness Definitions Fare?: Examining Public Attitudes Towards Algorithmic Definitions of Fairness. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, 99–106.
- [67] John Schaubroeck, Douglas R May, and F William Brown. 1994. Procedural justice explanations and employee reactions to economic hardship: A field experiment. *Journal of applied psychology* 79, 3 (1994), 455. <https://doi.org/10.1037//0021-9010.79.3.455>
- [68] Shlomi Segall. 2012. What’s so Bad about Discrimination? *Utilitas* 24, 1 (2012), 82–100. <https://doi.org/10.1017/S0953820811000379>
- [69] Amartya Sen. 2017. *Collective Choice and Social Welfare: Expanded edition*. Penguin UK.
- [70] Ming Singer. 1990. Determinants of perceived fairness in selection practices: An organizational justice perspective. *Genetic, Social, and General Psychology Monographs* (1990).
- [71] Latanya Sweeney. 2013. Discrimination in online ad delivery. *arXiv preprint arXiv:1301.6822* (2013). <https://doi.org/10.1145/2460276.2460278>
- [72] M Susan Taylor, Kay B Tracy, Monika K Renard, J Kline Harrison, and Stephen J Carroll. 1995. Due process in performance appraisal: A quasi-experiment in procedural justice. *Administrative science quarterly* (1995), 495–523. <https://doi.org/10.2307/2393795>
- [73] John W Thibaut and Laurens Walker. 1975. *Procedural justice: A psychological analysis*. L. Erlbaum Associates.
- [74] Huey-Ru Debbie Tsai, Yasser Shoukry, Min Kyung Lee, and Vasumathi Raman. 2017. Towards a socially responsible smart city: dynamic resource allocation for smarter community service. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, Article 13. <https://doi.org/10.1145/3137133.3137163>
- [75] Aki Tsuchiya, Luis Silva Míguel, Richard Edlin, Allan Wailoo, and Paul Dolan. 2005. Procedural justice in public healthcare resource allocation. *Applied Health Economics and Health Policy* 4, 2 (2005), 119–127. <https://doi.org/10.2165/00148365-200504020-00006>
- [76] Tom Tyler, Peter DeGoey, and Heather Smith. 1996. Understanding why the justice of group procedures matters: A test of the psychological dynamics of the group-value model. *Journal of personality and social psychology* 70, 5 (1996), 913. <https://doi.org/10.1037//0022-3514.70.5.913>
- [77] Tom R Tyler. 1989. The psychology of procedural justice: a test of the group-value model. *Journal of personality and social psychology* 57, 5 (1989), 830. <https://doi.org/10.1037//0022-3514.57.5.830>
- [78] Tom R Tyler. 1998. Trust and democratic governance. *Trust and governance* 1 (1998), 269.
- [79] Tom R Tyler and Andrew Caine. 1981. The influence of outcomes and procedures on satisfaction with formal leaders. *Journal of Personality and Social Psychology* 41, 4 (1981), 642. <https://doi.org/10.1037//0022-3514.41.4.642>
- [80] Umair Ul Hassan, Sean O’Á’Riain, and Edward Curry. 2013. Effects of expertise assessment on the quality of task routing in human computation. In *Proceedings of the 2nd International Workshop on Social Media for Crowdsourcing and Human Computation., Paris, France*. <https://doi.org/10.14236/ewic/sohuman2013.1>
- [81] Rajan Vaish, Snehal Kumar Neil S Gaikwad, Geza Kovacs, Andreas Veit, Ranjay Krishna, Imanol Arrieta Ibarra, Camelia Simoiu, Michael Wilber, Serge Belongie, Sharad Goel, et al. 2017. Crowd research: Open and scalable university laboratories. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. ACM, 829–843. <https://doi.org/10.1145/3126594.3126648>
- [82] Kees Van den Bos, E Allan Lind, Riël Vermunt, and Henk AM Wilke. 1997. How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect. *Journal of personality and social psychology* 72, 5 (1997), 1034. <https://doi.org/10.1037//0022-3514.72.5.1034>

- [83] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 440.
- [84] Laurens Walker, Stephen LaTour, E Allan Lind, and John Thibaut. 1974. Reactions of Participants and Observers to Modes of Adjudication 1. *Journal of Applied Social Psychology* 4, 4 (1974), 295–310. <https://doi.org/10.1111/j.1559-1816.1974.tb02601.x>
- [85] Weiqian Wang and Izak Benbasat. 2007. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* 23, 4 (2007), 217–246. <https://doi.org/10.2753/MIS0742-1222230410>
- [86] Allison Woodruff, Sarah E Fox, Steven Rousso-Schindler, and Jeffrey Warshaw. 2018. A qualitative exploration of perceptions of algorithmic fairness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 656.
- [87] Zhilin Zheng, Tim Vogelsang, and Niels Pinkwart. 2015. The impact of small learning group composition on student engagement and success in a MOOC. In *Proceedings of the 8th International Conference of Educational Data Mining*. 500–503.

## A PERCEIVED FAIRNESS OF ALGORITHMIC OUTCOMES

	Baseline		Transparency		Outcome Control	
	Own	Group	Own	Group	Own	Group
Initially Fair	6.50 (SE=.14)	6.36 (SE=.22)	5.89 (SE=.14)	6.14 (SE=.22)	6.48 (SE=.18)	6.36 (SE=.22)
Initially Less Fair	3.56 (SE=.24)	3.77 (SE=.18)	4.04 (SE=.24)	4.32 (SE=.18)	5.22 (SE=.24)	5.63 (SE=.18)

Table 1. Perceived fairness of algorithmic outcomes measured at the end of the baseline, transparency (standards clarity and outcome explanation), and outcome control phases. Participants rated the fairness of their own outcome and their group’s outcomes using a 7-Likert Scale (1=Strongly disagree that the outcome is fair, 7=Strongly agree that the outcome is fair).

Received April 2019; revised June 2019; accepted August 2019