algorithms

MDPI

Article

Fairness in Algorithmic Decision-Making: Applications in Multi-Winner Voting, Machine Learning, and Recommender Systems

Yash Raj Shrestha [1] and Yongjie Yang [2,*]

[1] Chair of Strategic Management and Innovation, Eidgenössische Technische Hochschule Zürich (ETH Zürich), 8092 Zürich, Switzerland; yshrestha@ethz.ch
[2] Chair of Economic Theory, Saarland University, 66123 Saarbrücken, Germany
* Correspondence: yyongjiecs@gmail.com

Received: 31 July 2019; Accepted: 16 September 2019; Published: 18 September 2019

Abstract: Algorithmic decision-making has become ubiquitous in our societal and economic lives. With more and more decisions being delegated to algorithms, we have also encountered increasing evidence of ethical issues with respect to biases and lack of fairness pertaining to algorithmic decision-making outcomes. Such outcomes may lead to detrimental consequences to minority groups in terms of gender, ethnicity, and race. As a response, recent research has shifted from design of algorithms that merely pursue purely optimal outcomes with respect to a fixed objective function into ones that also ensure additional fairness properties. In this study, we aim to provide a broad and accessible overview of the recent research endeavor aimed at introducing fairness into algorithms used in automated decision-making in three principle domains, namely, multi-winner voting, machine learning, and recommender systems. Even though these domains have developed separately from each other, they share commonality with respect to decision-making as an application, which requires evaluation of a given set of alternatives that needs to be ranked with respect to a clearly defined objective function. More specifically, these relate to tasks such as (1) collectively selecting a fixed number of winner (or potentially high valued) alternatives from a given initial set of alternatives; (2) clustering a given set of alternatives into disjoint groups based on various similarity measures; or (3) finding a consensus ranking of entire or a subset of given alternatives. To this end, we illustrate a multitude of fairness properties studied in these three streams of literature, discuss their commonalities and interrelationships, synthesize what we know so far, and provide a useful perspective for future research.

Keywords: algorithmic fairness; bias; machine learning; recommender system; algorithmic decision-making; multi-winner-voting; proportional representation; survey

1. Introduction

Decision-making by algorithms is becoming a ubiquitous part of our societal and economic lives. Algorithmic decisions increasingly appear in a plethora of domains such as healthcare, legal, education, banking, e-commerce, etc. In healthcare, for example, algorithms are being used to routinely monitor biochemical signals in patients, and immediately alert clinicians when anomalies arise [1]. Deep learning algorithms are able to process anonymized electronic health records and flag potential emergencies, to which clinicians are then promptly able to respond. Similarly, in US courts, an algorithmic system known as COMPASS is used to estimate the risk of recidivism. Human Resource departments in various companies are increasingly resorting to algorithms that are able to filter from the initial set of potential applications to reduce human time and effort in the evaluation of

applications [2]. Similarly, universities and colleges have begun using algorithmic predictions on big data to estimate which students will do well before accepting their admission applications [3]. With banks moving towards mobile payments to offer a seamless and fast customer experience, payment services based on machine learning algorithms verify and identify credit fraud in real-time. Similarly, insurance companies use automated data credibility assessment methods to quickly perform complex rounds of approval, verification, and evaluation so as to flag duplicate or otherwise unusual activities. Online retailers such as Amazon and Alibaba routinely deploy recommender systems algorithms in order to filter the set of product items that are displayed on the users dashboard. It is becoming evident that (with or without our desire) algorithmic decisions leave their footprints in our day to day activities from the way we do grocery shopping to the way we do banking. This increasing application and deployment of algorithmic decision-making in economy and society are driven by their high accuracy, effectiveness, low cost, and efficiency. Acceleration in the adoption of algorithmic decision-making is further supported by the access to mass volumes of data that is being currently collected in the digital economy as well as advancement in development of hardware such as General Processing Units (GPUs) and Tensor Precessing Units (TPUs).

In addition to the notable benefits and growing prevalence of algorithmic decisions, we are also witnessing growing concerns and skepticism in academia and popular media with respect to algorithmic unfairness and the evidences that they may inadvertently discriminate against certain minority groups. Evidence has shown that algorithmic decisions not only counteract and expose biases but also afford new mechanisms for introducing biases with unintended and detrimental effects [4]. Specifically, algorithmic decisions have been shown to amplify biases and unfairness embedded in data in terms of sensitive features such as gender, culture, race, etc. For example, in their recent study, Caliskan, Bryson, and Narayanan [5] found that natural language processing algorithms do capture historic discrimination against gender, such as by more closely associating words like "doctor" with males and "nurse" with females. As such algorithms are trained on historical data, past discrimination and stereotypes prevalant in the society are reflected in their predictions. These concerns become particularly alarming when algorithmic decisions are interacting and influencing almost every aspect of economic and social life of groups and individuals. As an example, consider the work of Angwin et al. [6], who found that COMPASS is biased against African-American defendants. As the tools' error rates were asymmetric, African-American defendants were more vulnerable to be incorrectly labeled as higher-risk than they actually were when compared to their white defendants. In another example, recommender algorithms deployed for personalization have been shown to propagate or even create biases that may influence decisions and opinions of the user at the receiving end [7,8]. Such phenomena has been observed in social media platforms such as Facebook and Twitter, resulting in an inflation in the polarization of society by over 20 percent in the last eight years [9]. Algorithmic decisions have also been shown to amplify biases with respect to gender embedded in data. For example, algorithms trained on data which feature under representation of women in science, technology, engineering and mathematics (STEM) topics output decisions more biased towards men [10].

Nevertheless, it is encouraging to observe that as a response to the above-mentioned scrutiny and following debates in popular media, computer science scholars have been swift in beginning to collaborate with lawyers, policy-makers, economists, social scientists, and others in designing fair, transparent, and reliable algorithms. This has also led to the organization of the relatively new yet much influential ACM conference on Fairness Accountability and Transparency in Machine Learning (FATML), which is particularly targeted at bringing together researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems. Even though these outlets mainly focus on algorithmic fairness pertaining to machine learning algorithms, this represents an important step in achieving fairness in algorithmic decision-making in general.

Taking a further step in this direction, in this paper, we review the proliferation of research on fairness in algorithms, synthesize our present understanding, and conclude with identification of

major challenges (if any) and pressing open questions and future research directions. We particularly look into three domains of research on decision-making, namely, multi-winner voting algorithms (Section 2), machine Learning algorithms (Section 3), and recommendation systems algorithms (Section 4). A major portion of recent research on fairness in algorithmic decision-making falls into one of these three domains. We recall many concrete concepts of fairness in these areas and discuss their importance, interrelationships, as well as other related problems. Notably, understanding and keeping in touch with latest research in fairness in algorithms is of great importance to policy makers and practitioners interested in introducing algorithmic decision-making into their organizations and businesses. This article also aims to provide a concise overview to readers looking for an outlook on diverse dimensions of algorithmic fairness research in one place. This is important because, as is evident in our review, algorithmic fairness research is advancing in diverging directions with a variety of definitions, designs of fair algorithms and mechanisms being developed with uncorrelated and non intersecting idiosyncratic assumptions. It is therefore imperative to have some sort of convergence in the further development of the field in order to facilitate more beneficial societal impact.

Note that research has revealed that biases in algorithmic decision can come from multitude of sources, such as human decisions on how the data was collected, noisy preferences provided by decision makers, features selected, steps taken in cleaning and preprocessing the data, and even the choice of algorithms itself. These elements are largely dependent on courses of action taken by the user of the algorithm. However, in this article, our main focus is to stick with fairness in algorithmic decision-making precluding the biases introduced by the courses of action by humans. Accordingly, given as input a set of alternatives $A = \{a_1, a_2, \ldots, a_n\}$, the decision-making task is required to evaluate this set based on a clearly defined objective function. Based on this evaluation, the decision-making algorithm groups the alternatives and ranks the groups into a particular order. This function symbolizes applications such as (1) collectively selecting a fixed number of winners from a set of alternatives; (2) clustering the set of alternatives into disjoint groups; or (3) finding a consensus ranking of a smaller subset of or all alternatives. We believe these specific tasks are covered by the applications of algorithms in multi-winner voting, machine learning, and recommender systems.

2. Multi-Winner Voting

Collective decision-making is a significant branch of social choice theory and has wide applications in both economics and computer systems. Examples of applications include political elections, committee selections (e.g., journal editorial board selection), selecting items to display in online shops, recommending multiple items to users in recommender systems, company or institute employee recruitment, heuristic algorithms selection in meta-heuristics, selecting data to load into caches in cloud computing systems, etc. Concretely, collective decision-making is mainly concerned with deriving consensus outcomes based on preferences of a number of decision-making participants over possible outcomes. Without a doubt, voting is one of the most popular approaches for collective decision-making. In this setting, we have a set of candidates C (possible outcomes), a set of voters V (decision-making participants) each of whom has a preference over candidates in C, and then we either aim to select a subset of exactly k candidates as winners for some integer, k, or find a ranking of candidates from the best to the worst for the community. It should be pointed out that voters need not necessarily to be human beings, they can also be certain criteria, robots, functions, or even algorithms. A large number of algorithms or multi-winner voting rules have been proposed for the purpose of the former. However, as fairness properties were not comprehensively taken into account when these rules were coined, many of them may result in unfair outcomes. For instance, assume that we have 100 voters who are divided into two groups: the majority and the minority. In particular, the majority consists of 90 voters, all of whom approve their spoiled candidates $c_1, c_2, \ldots, c_{10}$. The remaining 10 voters, who are a minority, approve only the last candidate denoted by c', probably because only this candidate has positive utility to them. If we aim to select 10 winners and apply the prevalent approval voting, then

$\{c_1, c_2, \ldots, c_{10}\}$ will be selected, as they are approved by the maximum number of voters. This result is clearly biased against the minority since their opinion is completely ignored.

In this section, we shall survey recent progress on the study of fairness properties in multi-winner decision-making. Regarding fairness, an important concern is fair for whom and fair at which level. These two questions are important guidance for us to define different fairness properties. In voting, we have two types of entities, namely, the candidates and the voters, both of whom may need to be fairly considered. For single-winner voting rules ($k = 1$), which aim to select exactly one winner, the *neutrality* property ensures that candidates are treated equally, whereas the *anonymity* property ensures that voters are treated equally [11]. Recall that neutrality says that the winners' identities remain the same after candidates are renamed, and anonymity says that all voters have the equal power and the order of them have no impact on the results (see the work by the authors of [12] for the formal definitions). These two properties have long been studied in the literature. Neutrality and anonymity are of course also desired for multi-winner voting rules, where a fixed number of winners are selected [13,14]. However, these two properties only provide individual-level fairness by regarding each voter and each candidate as an independent individual, but do not say anything about group-level fairness, which is of particular importance in some real-world applications. Consider the above example and consider what should be a fair result for both the majority and the minority. As the minority accounts to 10% of all voters, should 10% of the winners also come from their approved candidates? If this is the case, then a fair result would be that selecting c' and nine of $\{c_1, c_2, \ldots, c_{10}\}$ as the winners. To fill the gap, proportional fairness properties of multi-winner decision-making have been proposed and received a considerable amount of study in the literature in recent years. Generally speaking, these properties stipulate that certain groups of voters should be proportionally represented in a committee according to the strengths of their numbers.

This section is devoted to numerous important proportionality properties studied in the recent literature. We discuss mainly two preference models: the dichotomous preference model and the linear preference model.

> **Dichotomous preference.** Each voter classifies candidates into two classes, namely, the *approved candidates* and the *disapproved candidates*. In particular, all approved candidates are preferred to all disapproved candidates, and candidates inside each class are equally preferred.
> **Linear preference.** Each voter ranks all candidates in a linear order $\succ$, from the best to the worst. For two candidates, a and b, $a \succ b$ means that the corresponding voter strictly prefers a to b.

Multi-winner voting rules with dichotomous preferences are often referred to as approval-based multi-winner voting rules, and with linear preferences are referred to as ranking-based rules.

We divide our discussions into four subsections. In Sections 2.1 and 2.2, we survey fairness properties for ranking-based and approval-based voting, respectively. These properties are aimed at certain groups of voters. We shall give the definitions of these properties, discuss the relations among them, point out the complexity of two important problems related to these concepts, and offer an overview of the most important voting rules studied in the literature and whether they fulfill these properties. In Section 2.3, we discuss recent research on the setting where candidates have sensitive attributes or are labeled, and fairness are provided for groups of candidates. Section 2.4 is aimed at discussing stability concept-based fairness properties.

2.1. Voter Fairness in Ranking-Based Voting

We consider first ranking-based voting where voters are asked to report linear order preferences over candidates. A voter v's preference is denoted by $\succ_v$, so that $a \succ_v b$ represents that this voter prefers the candidate a to the candidate b. A crucial notion in this setting is *solid coalition* which was first mentioned in the work by the authors of [15]. Particularly, for a subset $C' \subseteq C$ of candidates, a solid coalition is a subset of voters $U \subseteq V$ such that all voters in U rank all candidates in C' above all

the other candidates, i.e., for all voters $v \in U$, it holds that $a \succ_v b$ for all $a \in C'$ and $b \in C \setminus C'$. In this case, we say that U supports C' and call U a C'-solid coalition.

The following proportional property provides fairness for solid coalitions: it states that for a solid coalition of a certain scale, a guaranteed number of candidates supported by this coalition should be selected as winners.

q-Proportionality for solid coalition (q-PSC) [16]. For a rational number, q, a k-committee $w \subseteq C$ satisfies q-PSC if for every positive integer ℓ and for every solid coalition $U \subseteq V$ supporting some $C' \subseteq C$ such that $|U| \geq \ell q$, it holds that $|w \cap C'| \geq \min\{\ell, |C'|\}$.

Normally, we are only interested in the case where $\frac{n}{k+1} < q \leq \frac{n}{k}$, where n denotes the number of voters. One of the reason is that when $k = 1$, i.e., we select only one winner, a q-PSC committee is a singleton consisting of a candidate who is most preferred by at least a majority of the voters, whenever such a candidate exists (note that such a candidate must be a Condorcet winner). In addition, q-PSC is not guaranteed to exist if $q \leq n/(k+1)$. Moreover, if $q > n/k$, any q-PSC committee must provide some counter-intuitive properties (see the work by the authors of [16] for the details).

Specifically, if q is equal to the so-called Hare quota n/k, the property is referred to as Hare-PSC (q_{H}-PSC). Besides, if q is equal to the Droop quota $\left\lfloor \frac{n}{k+1} \right\rfloor + 1$, the property is referred to as Droop-PSC (q_{D}-PSC).

Proportionality for solid coalition seems to be first considered by Dummett [15]. Many of its variants have been studied very recently [16,17]. For example, the weak q-PSC puts constraints only on solid coalitions supporting a limited sized committee, and asks a committee w to contain all candidates who are supported by these solid coalitions.

Weak q-PSC. [16]. A committee $w \subseteq C$ satisfies weak q-PSC if the following holds, for every positive integer ℓ, every $C' \subseteq C$ such that $|C'| \leq \ell$, and every C'-solid coalition U of size at least ℓq, it holds that $C' \subseteq w$.

Similar to q-PSC, we are particularly interested in the case where $\frac{n}{k+1} < q \leq \frac{n}{k}$. Weak q_{H}-PSC and weak q_{D}-PSC are referred to as weak q-PSC, where q takes the Hare quota and the Droop quota, respectively. Both PSC and weak PSC are designed to guarantee fairness for voters at group levels, but they differ at the degree of fairness they could provide. In fact, due to the definitions, we know that q-PSC implies weak q-PSC, but not necessarily the other way around.

Given a concept of fairness, a significant question is whether we can compute a committee providing the fairness efficiently. Let τ be a fairness property.

τ-Computing	
Input:	An election (C, V) and a positive integer $k \leq \|C\|$.
Question:	Is there a k-committee $w \subseteq C$ which provides the τ property at (C, V)?

Prior to the proposal of many fairness properties, a large body of voting rules have been extensively and widely studied in the literature. Analyzing whether the outcomes of these voting rules provide some specific fairness property is of also particular importance. This motivation brings the following decision problem into the line of research.

τ-Testing	
Input:	An election (C, V) and a committee $w \subseteq C$.
Question:	Does w satisfy τ at (C, V)?

Concerning the first decision problem, Aziz and Lee [16] proved that both computing and testing q-PSC and weak q-PSC are polynomial-time solvable for all possible values of q. See Table 1.

Table 1. Complexity of computing a committee satisfying a proportional property, or testing whether a given committee satisfies a proportional property. In the table, "P" stands for "polynomial-time solvable". All results are from the work by the authors of [16].

	Complexity of Computing	Complexity of Testing
q-PSC	P	P
weak q-PSC	P	P

For the second decision problem, we survey the results for many voting rules studied in the literature. Let (C, V) be an election. For a vote $\succ \in V$, the position of a candidate c in $\succ$, denoted by $\mathsf{pos}_{\succ}(c)$, is the number of candidates ranked before c plus one, i.e.,

$$\mathsf{pos}_{\succ}(c) = |\{c' \in C \setminus \{c\} : c' \succ c\}| + 1.$$

Committee scoring rules. Under a committee scoring rule, each voter provides a score to each committee based on the positions of the committee-members in the preference of this voter, and winning committees are those with the maximum total score. Committee scoring rules were first studied by Elkind et al. [17] as a general framework to encapsulate many concrete multi-winner voting rules, including, e.g., Bloc, k-Borda, Chamberlin–Courant, etc.

- **k-Borda.** Each voter gives $m - i$ points to each candidate ranked in the i-th position, where m denotes the number of candidates. The score of a committee from a voter is the sum of the scores of all its members from the voter.
- **Bloc.** Every voter gives 1 point to all of their top k ranked candidates. The score of a committee from a voter is the sum of the scores of all its members from the voter.
- **Single nontransferable vote (SNTV).** Every voter gives 1 point to her top ranked candidate. The score of a committee from a voter is the sum of the scores of all its members from the voter.
- **Chamberlin–Courant (CC).** Different from the above three rules where all members of the winning committee are counted to accumulate the satisfaction of a voter, in CC, for each voter, only the best candidate in the winning committee contributes to the satisfaction of this voter. In other words, each voter is assumed to be only represented by her best candidate in the winning committee. Precisely, each voter has a nonincreasing mapping $\alpha : \mathbb{N} \to \mathbb{N}$, such that $\alpha(i)$ is a voter's satisfaction of a candidate ranked in the i-th position. For a voter v with preference $\succ_v$ and a nonempty committee $w \subseteq C$, let $\mathsf{top}^w(\succ_v)$ be the top-ranked candidate of v among w, i.e., $\mathsf{top}^w(\succ_v)$ is the candidate $c \in w$, such that $c \succ_v c'$ for all $c' \in w \setminus \{c\}$. The CC score of a committee $w \subseteq C$ from a voter with mapping α is then $\alpha(\mathsf{pos}_{\succ_v}(\mathsf{top}^w(\succ_v)))$. In this section, we consider only the Borda satisfaction function $\alpha : \mathbb{N} \to \mathbb{N}$, which, for m candidates, holds that $\alpha(i) = m - i$.

Monroe's rule. This rule is similar to the CC rule but with a further restriction that every candidate can represent at most $\lceil \frac{n}{k} \rceil$ voters. Let $g : V \to C$ be an assignment function and $g^-(c)$, $c \in C$ be the set of voters, $\succ \in V$, such that $g(\succ) = c$. Moreover, let $\mathcal{G}$ be the set of all assignment functions from V to C. The Monroe score of a k-committee $w \subseteq C$ is then defined as

$$\max_{\substack{g \in \mathcal{G} \text{ s.t.} \\ |g^-(c)| \leq n/k \text{ for all } c \in C}} \left\{ \sum_{\succ \in V} \alpha_{\succ}(g(v)) \right\},$$

where $\alpha_{\succ} : \mathbb{N} \to \mathbb{N}$ is a mapping as in CC. Monroe's rule selects k-committees with the maximum score as winning committees.

Single-transferable voting (STV). STV rules are a large class of voting rules each of which is featured by a rational number q and some vote-reweighting approach. A common principle of these rules is to guarantee certain groups of voters are proportionally represented. Fixing a

rational quota q and a vote-reweighting approach, the STV rule selects winning committees iteratively as shown below. For a candidate c, let $V^{\mathsf{top}}(c)$ be the set of voters ranking c in the top.

1. Initially, we associate to each voter $v \in V$ a weight denoted by $\mathsf{weight}(v)$. (Usually, all voters have weight 1 initially, but this is not necessarily the case.)
2. If there is a candidate, $c \in C$, that is ranked in the top by at least q voters, that candidate is added to the winning committee. Then, we apply the vote-reweighting approach so that the total weight of all votes ranking c in the top are reduced by $\min\{q, p\}$, where $p = \sum_{v \in V^{\mathsf{top}}(c)} \mathsf{weight}(v)$ is the sum of the weights of all voters ranking c in the top before the reweighting. Moreover, the candidate c is deleted from C and from all votes.
3. If there is no such a candidate c as discussed above, then a candidate that is ranked in the top by the least number of voters is eliminated.
4. The procedure terminates until k candidates are selected.

Many of concrete STV rules have been considered in the literature (see the works by the authors of [18,19] for a history and a summary of many important STV rules). However, for simplicity, in this survey, we discuss only STV rules where initially all voters have weight 1, and the *uniform reweighting approach* is used in Step 2. Particularly, according to this reweighting approach, in Step 2, the weight of a voter v which ranks c in the top is reduced to $\min\{0, \mathsf{weight}(v) \cdot (1 - \frac{q}{p})\}$. Two important STV rules are those when q is equal to the Hare quota or the Droop quota, i.e., $q = \frac{n}{k}$ and $q = \left\lfloor \frac{n}{k+1} \right\rfloor + 1$. We denote these two special STV rules as D-STV and H-STV, respectively.

Much research has been done to investigate whether the above defined multi-winner voting rules provide the PSC fairness property, see Table 2 for a summary of the current known results. According to this table, by using STV rules, we are able to obtain, in polynomial time, winning committees that provide both q_{H}-PSC and q_{D}-PSC fairness simultaneously. Nevertheless, it is important to point out that STV rules fail many monotonic properties [16]. For a nice remedy, Aziz and Lee recently proposed a new rule, which they named "Expanding Approvals Rule" (EAR). In particular, they showed that EAR has the following advantages compared with any other concrete rules studied in the literature to date. First, an EAR winning committee can be always computed in polynomial-time. Second, EAR committees provide both q_{H}-PSC and q_{D}-PSC fairness. Third, EAR committees satisfy many monotonic properties. Finally, EAR works not only for strict preference elections but also for the case where voters hold weak order preferences over candidates. See the work by the authors of [16] for the definition of EAR and the detailed discussions.

Table 2. A summary of the PSC properties satisfied by several important multi-winner voting rules and the complexity of computing a winning committee with respect to these rules. In the table, "N" means that the rule in the corresponding row does not satisfy the property in the corresponding column, and "Y" means that the rule satisfies the property. Observing that weak q-PSC is a too strong property for many rules to satisfy, Elkind et al. [17] studied three weak versions, namely, solid coalitions, consensus committee, and unanimity. They showed that each of SNTV, Bloc, k-Borda, CC, and Monroe fails at least one of these weak versions, and these results imply the ones for these rules in the table.

	q_{H}-PSC	q_{D}-PSC	Weak q_{H}-PSC	Weak q_{D}-PSC	Complexity
k-Borda	N [17]	N [17]	N [17]	N [17]	P (trivial)
Bloc	N [17]	N [17]	N [17]	N [17]	P (trivial)
SNTV	N [17]	N [17]	N [17]	N [17]	P (trivial)
CC	N [17]	N [17]	N [17]	N [17]	NP-complete [20]
Monroe	N [17]	N [17]	N [17]	N [17]	NP-complete [21]
H-STV	Y [16]	Y [16]	Y [16]	Y [16]	P (trivial)
D-STV	Y [16]	Y [16]	Y [16]	Y [16]	P (trivial)

2.2. Voter Fairness in Approval-Based Voting

This section is devoted to fairness properties of approved-based multi-winner voting rules, where each vote $v \in V$ consists of a subset of candidates, the candidates approved by the corresponding voter. Several important proportional properties have been put forward in the literature. In general, these properties aim at providing fairness for certain group of voters who approve some candidates in common. In particular, they ensure that for such a group of enough large size, at least a certain number of candidates approved by all (or some) members of this group should be selected.

Justified representation (JR). A k-committee, $w \subseteq C$, provides JR, if, for every subset $U \subseteq V$ of at least $\frac{n}{k}$ votes such that $\bigcap_{u \in U} u \neq \emptyset$, at least one of the candidates approved by some vote in U is included in w, i.e.,

$$w \cap \left(\bigcup_{u \in U} u \right) \neq \emptyset.$$

This property was proposed by Aziz et al. [22,23].

Proportional justified representation (PJR). A k-committee ,$w \subseteq C$, provides PJR if for every positive integer $\ell \leq k$, and for every subset $U \subseteq V$ of at least $\ell \cdot \frac{n}{k}$ votes such that $|\bigcap_{u \in U} u| \geq \ell$, the committee w contains at least ℓ candidates from $\bigcup_{u \in U} u$, i.e., $|w \cap (\bigcup_{u \in U} u)| \geq \ell$. This property was proposed in the work by the authors of [24].

Extended justified representation (EJR). A k-committee $w \subseteq C$ provides EJR if for every positive integer $\ell \leq k$ and for every subset $U \subseteq V$ of at least $t \geq \ell \cdot \frac{n}{k}$ votes such that $|\bigcap_{u \in U} u| \geq \ell$, the committee w contains at least ℓ candidates from every vote $u \in U$, i.e., $|w \cap u| \geq \ell$ for all $u \in U$. This property was proposed by Aziz et al. [23].

Perfect representation (PR). PR is defined for special elections. Particularly, let (C, V) be an election such that $|V| = t \cdot k$ for some integer t. A k-committee $w = \{c_1, c_2, \ldots, c_k\}$ provides PR if there is a partition $(V_1, V_2, \ldots, V_k)$ of V such that $|V_i| = t$ for all $1 \leq i \leq k$ and c_i is approved by all votes in V_i. This property was studied in the work by the authors of [24].

From the definitions, it is easy to see that EJR implies PJR, and PJR implies JR [22,24].

As discussed in the previous section, two important questions are (1) whether we can always calculate a committee providing a certain property efficiently and (2) whether we can determine whether a given committee provides a certain fairness property efficiently. For JR, we have positive answers to both questions. However, for the two more refined concepts EJR and PJR, we have only the positive answer to the first question. See Table 3 for a summary of the concrete results.

Table 3. Complexity of computing a committee satisfying a proportional property or testing whether a given committee satisfies a proportional property.

	Complexity of Computing	Complexity of Testing
justified representation	P [23]	P [23]
extended Justified representation	P [25]	co-NP-complete [23]
proportional Justified representation	P [24]	co-NP-complete [25]
perfect representation	NP-complete [24]	P [24]

The next important question is, therefore, whether there are committees providing several fairness properties simultaneously, and whether we calculate such a committee in polynomial time if it exists. We check the answer by surveying several well-studied and natural multi-winner voting rules.

Approval voting (AV). The AV score of a candidate is the number of votes approving this candidate, and a winning k-committee consists of k candidates with the highest AV scores.

Satisfaction approval voting (SAV). The SAV score of a candidate c is defined as

$$\sum_{c \in v \in V} \frac{1}{|v|} - \sum_{\substack{v \in V, \\ c \notin v}} \frac{1}{m - |v|}$$

where m denotes that number of candidates. A winning k-committee consists of k candidates with the highest SAV scores.

Minimax approval voting (MAV). This rule aims to find a committee that is most close to every voter's opinion. More precisely, the Hamming distance between a committee w and a vote v is $d_{\mathrm{H}}(v, w) = |w \setminus v| + |v \setminus w|$, and this rule selects a k-committee w minimizing $\max_{v \in V} d_{\mathrm{H}}(v, w)$.

Proportional approval voting (PAV). The PAV score of a committee w is defined as

$$\sum_{\substack{v \in V, \\ v \cap w \neq \emptyset}} \left(1 + \frac{1}{2} + \cdots + \frac{1}{|v \cap w|}\right).$$

A winning k-committee is an one with the maximum score.

Sequential proportional approval voting (seq-PAV). This rule provides an approximation solution to PAV rule. It selects k winners in k rounds, one in each round. Precisely, initially we let $w = \emptyset$. Assume that we have an i-committee w after round $i < k$. Then, in the next round, we find a candidate c which offers the maximum PAV score of $w \cup \{c\}$, and we extend w by resetting $w := w \cup \{c\}$. After k rounds, w contains exactly k candidates.

Chamberlin–Courant approval voting (CCAV). This rule is a variant of CC rule for approval-based voting. In particular, a voter satisfies with a committee if and only if this committee contains at least one of her approved candidates. This rule selects a k-committee that satisfies the maximum number of voters.

Monroe's approval voting (MonAV). This is a variant of Monroe's rule for approval-based voting and is similar to CCAV. In CCAV, a candidate can satisfy all voters who approve this candidate. However, in MonAV, we require that each candidate is assigned to at most $\lceil \frac{n}{k} \rceil$ voters approving this candidate and, moreover, each voter can be assigned to at most one candidate. The MonAV score of a committee is the maximum number of voters who are satisfied by this committee and fulfill the above conditions.

In addition to the above rules, a class of important rules, coined by Phragmén, have been studied. These rules determine the winners in a reverse-thinking approach. Particularly, assume that we know the k winners. The rules assume that each of this winners has a unit point which is distributed over all voters approving this candidate in a way to achieve some objective (Phragén's rules differ only at the objectives). Then, the selected winners should be those that yield the optimal objective over all subsets of k candidates. We give the formal definitions below.

A *load distribution* is a two-dimensional array $\mathbf{x} = (x_{v,c})_{v \in V, c \in C}$ satisfying the following conditions.

1. For each $v \in V$ and $c \in C$ it holds that

$$0 \leq x_{v,c} \leq 1.$$

2. For every $c \in C$ and $v \in V$, if $c \notin v$, then

$$x_{v,c} = 0.$$

This corresponds to winner that is only distributed over voters approving that winner.

3. It holds that

$$\sum_{v \in V, c \in C} x_{v,c} = k.$$

That is, there are in total k pointes to be distributed.

4. For every $c \in C$, it holds that
$$\sum_{v \in V} x_{v,c} \in \{0, 1\}.$$
This together with the previous restriction ensure that exactly k candidates have points to distribute.

For a load distribution $\mathbf{x}$ and a vote v, let $x_v = \sum_{c \in C} x_{v,c}$. Particularly, x_v is referred to as the voter load of v. Due to the last two conditions in the definition of load distribution, we know that each load distribution $\mathbf{x}$ gives us a unique k-committee
$$f(\mathbf{x}) = \left\{ c \in C : \sum_{v \in V} x_{v,c} = 1 \right\}.$$
Note that for a k-committee w, there can be multiple load distributions $\mathbf{x}$ such that $f(\mathbf{x}) = w$.

max-Phragmén. This rule first calculates a load distribution $\mathbf{x}$ such that $\max_{v \in V} x_v$ is minimized. Then, $f(\mathbf{x})$ is the winning committee.
var-Phragmén. This rule first calculates a load distribution $\mathbf{x}$ such that $\sum_{v \in V} x_v^2 = \sum_{v \in V} (\sum_{c \in C} x_{v,c})^2$ is minimized. Then, $f(\mathbf{x})$ is the winning committee.
seq-Phragmén. This rule takes k rounds to select the winners, one for each round. For a candidate c, let $V_c = \{v \in V : c \in v\}$ be the set of voters approving c. Initially, let $w = \emptyset$. Let $x_v^{(j)}$ denote the voter loads after round j. At first, all voters have a load of 0, i.e., $x_v^{(0)} = 0$ for all $v \in V$. As a first candidate, we select one $c \in C$ that receives the most approvals and add c into w. Then, the voter load of each voter approving this selected candidate is increased to $\frac{1}{|V_c|}$. In the next round, we choose a candidate that induces a (new) maximal voter load that is as small as possible, but now we have to take into account that some voters already have a non-zero load. The new maximal load if some candidate $c \in C$ is chosen as the $(j+1)$-st committee member is measured as
$$s_c^{(j+1)} = \frac{1 + \sum_{c \in v \in V} x_v^{(j)}}{|V_c|}.$$
In other words, if c is chosen, then we adjust the voter loads of all voters approving c, so that they have the same voter load afterwards. Let c be the candidate that minimizes $s_c^{(j+1)}$ among those that are not yet in w. Then we add c to w and set $x_v^{(j+1)} := s_c^{(j+1)}$ for all $v \in V_c$. After k rounds, the committee w consists of exactly k candidates. Note that we also obtain a load distribution $\mathbf{x}$ such that $f(\mathbf{x}) = w$.

The above rules have been extensively studied in the literature from different aspects [26–34]. However, the proportionality properties defined above of these rules have only received attention recently. Fernández et al. [24] proved that winner determination for all multi-winner voting rules that satisfy PR must be NP-hard. This directly implies that AV, SAV, and seqPAV do not fulfill PR as winner determination for these rules are polynomial-time solvable. Fernández and Fisteus [35] showed that MAV does not satisfy PR. Aziz et al. [23] showed that AV, SAV, MAV, and seqPAV do not satisfy JR. As PJR and EJR imply JR, it must be that AV, MAV, SAV, and seqPAV fail also PJR and EJR. So, none of AV, MAV, SAV, and seqPAV satisfy any properties studied in this section. The proportional properties of other rules have also been studied in the literature and we summarize them in Table 4.

Table 4. A summary of proportional properties of important approval-based multi-winner voting rules and the complexity of winner determination for these rules. In the table, "N" means that the rule in the corresponding row does not satisfy the property in the corresponding column, and "Y" means that the rule satisfies the property.

	EJR	PJR	JR	PR	Complexity
AV	N [23]	N [23,24]	N [23]	N [24,35]	P (trivial)
SAV	N [23]	N [23,24]	N [23]	N [24,26,35]	P [26]
seqPAV	N [23]	N [23,24]	N [23]	N [24,26,35]	P [26]
MAV	N [23]	N [23,24]	N [23]	N [23]	NP-complete [36]
CCAV	N [23]	N [24]	Y [23]	Y [35]	NP-complete [37]
MonAV	N [23]	N [24]	Y [23]	Y [24]	NP-complete [21]
var-Phragmén	N [38]	N [38]	Y [38]	Y [38]	NP-complete [38]
seq-Phragmén	N [38]	Y [38]	Y [38]	N [38]	P [38]
max-Phragmén	N [38]	Y [38]	Y [38]	Y [38]	NP-complete [38]
PAV	Y [23]	Y [24]	Y [23]	N [24]	NP-complete [26]

With the help of Table 4, we know that the answer to the following important question is in the negative:

> *Is there a natural rule (or an algorithm) whose outcome always provide JR, EJR, PJR, and PR simultaneously?*

But do we still have some hope? The answer is unfortunately in the negative again. In fact, Fernández et al. [24] proved that there are no voting rules whose outcome always provides both PR and EJR. In particular, they construct an election instance where none of the PR committees provides EJR (Theorem 4 in the work by the authors of [24]). This negative result is in fact their motivation to propose the PJR property. Due to the fact that EJR implies PJR, and PJR implies JR, and the above impossibility result, our question then breaks down to the following two questions. First, is there any natural EJR rule? Second, is there any natural rule whose outcome always provide PR and PJR? The results in Table 4 provide a comprehensive answer: among the rules in the table, PAV is the only one that provides EJR, and thus provides JR and PJR too, and max-Phragmén is the only one that guarantees PJR and PR, and thus provides JR too. However, an obvious disadvantage of PAV and max-Phragmén is that computing a winning committee for them turned out to be a computationally hard problem. To overcome this dilemma, we need to explore alternative rules that satisfy these properties. Max-Phragmén seems unlikely to have any proper alternative to remedy the disadvantage, since it has been shown that computing any PR committee is NP-complete [24]. For PAV, there do exist good alternatives. In particular, very recently, Aziz et al. [25] crafted two polynomial-time algorithms (multi-winner voting rules) whose outcome always provides EJR, and thus provides JR and PJR as well. It should be pointed out that other approaches to overcoming the dilemma include designing fixed-parameter algorithms or polynomial-time algorithms for some domain-restricted elections. This work has been conducted for PAV in recent years (see, e.g., [34,39,40]).

2.3. Fairness for Candidates with Sensitive Attributes

In the previous two sections, we mainly survey fairness properties designed for certain groups of voters. In some real-word applications, candidates have sensitive attributes. In these applications, fairness for groups of candidates has to be imposed into the decision-making procedure to avoid discrimination. In this section, we survey the recent progress of the study on this topic. We still assume that a fixed number k of winners shall be selected.

Ceils, Huang, and Vishnoi [41] studied fairness in the setting where candidates are in a number of groups, each of which corresponds to a sensitive attribute such as gender, ethnicity, etc. Notably, each candidate may have several attributes and hence the groups may be non-disjoint. They studied a quite general framework which requires that a winning committee should be a one that

maximizes the score with respect to some defined scoring function, and fulfills the restriction that for each group of candidates with a specific attribute, a prescribed fraction of the group members must be selected.

Formally, let $f : 2^C \to \mathbb{R}_{\geq 0}$ be a scoring function. Moreover, let $C_1, C_2, \ldots, C_t \subseteq C$ be subsets of candidates (it may that $C_i \cap C_j \neq \emptyset$), and for each C_i, $1 \leq i \leq t$, let ℓ_i and u_i be two integers such that $0 \leq \ell_i \leq u_i \leq |C_i|$. Then, the goal of an f-multi-winner voting rule is to select a k-committee $w \subseteq C$ with the maximum score under the restriction that for each C_i, $1 \leq i \leq t$ it holds that $\ell_i \leq |w \cap C_i| \leq u_i$.

The framework is so general that it only stipulates the maximum and the minimum numbers of candidates that should be selected from each group in general but leaves the settings of these two values to ad hoc applications. Particularly, ℓ_is and u_is can be constants, or any function of the number of candidates in the groups, the total number of candidates, etc. As argued by the authors, the framework generalizes several important proportional fairness properties studied in the literature including such as fully proportional representation [42], fixed-degressive proportionality [43], flexible proportionality [44], etc.

Given that the framework is so general, it is not surprising that computing a winning committee is a computationally hard problem. Given this negative result, the authors explored numerous approximation algorithms for calculating committees satisfying the above fairness constraints. Their results largely depend on the maximum number of groups each candidate is included. For example, they showed that if everyone belongs to exactly one group, i.e., $(C_1, C_2, \ldots, C_t)$ form a partition of C, there is a $(1 - 1/e)$-approximation algorithm, and they showed that this is probability optimal. However, in the case where some candidate belongs to at least three groups, checking whether there is a feasible solution is already NP-hard, and even in the case where feasible solutions exist, finding a solution with approximation factor $\omega(\log \triangle / \triangle)$ remains NP-hard, where $\triangle$ is the maximum number of groups each candidate belongs to. For many other interesting theoretical results, we refer to Tables 1 and 2 in [41]. The authors also conducted an experimental work to show that for many rules, the constrained version outputs a committee which is very close to the unconstrained version. Table 4 in [41] summarizes their findings regarding this issue.

Almost at the same time Ceils, Huang, and Vishno [41] posted their paper on Arix (https://arxiv.org/abs/1710.10057); Bredereck et al. [45] posted on Arix (https://arxiv.org/abs/1711.06527) a paper investigating a similar model. However, they mainly focused on the parameterized complexity and computational complexity of the winner determination problem. Similar constraints have been also considered in party-based voting (a apportionment problem) [46], where each party nominates several candidates and a total number of k seats should be distributed to these parties based on the preferences of voters to parties.

2.4. Stable Fairness

Cheng et al. [47] recently put forward a notion of group fairness inspired by the concept of core in cooperative game theory. In general, it says that a committee is fair to a group of voters if they cannot obtain a committee of proportional size that is strictly better for all members by deviating. Formally, for two committees $S \subseteq C$ and $S' \subseteq C$, let $V(S, S')$ be the number of voters prefer S' to S. We say that S' blocks S if and only if

$$V(S, S') \geq \frac{|S'|}{k} \cdot n$$

where n is the number of voters and k is the desired winning committee size. A committee S is i-stable for some integer i, such that $1 \leq i \leq k$ if and only if there does not exist a committee S' of size at most i which blocks S. Cheng et al. [47] showed that their notion generalizes some previous studied notions such as justified representation. They also extended their notion to Stable Lotteries and Approximate Stability, and studied the existence of these stable solutions and how efficient they can be calculated. We refer to the work by the authors of [47] for the details.

3. Machine Learning Algorithms

Machine learning (ML) algorithms have gained a lot of attention in recent years due to their growing predictive capabilities. In this paper, we mainly cover supervised machine learning. The other two classes of machine learning, namely, unsupervised machine learning and reinforcement machine learning algorithms, have gained comparatively lesser research attention with respect to fairness and also remains beyond the scope of our review.

Supervised machine learning algorithms are provided a set of input "features" denoted by $x^{(i)} \in \mathcal{X}$ and output "target" labels $y^{(i)} \in \mathcal{Y}$, which is jointly called "training set" $(\mathcal{X}, \mathcal{Y})$. Given the training set, supervised machine learning algorithms learn a function $h : \mathcal{X} \mapsto \mathcal{Y}$ such that $h(x)$ is a "good" predictor for the corresponding value of y (for an unknown x), where h denotes "hypothesis". Based on the distribution of $\mathcal{Y}$, such a task could either be "regression" (where $y^{(i)} \in \mathcal{Y}$ is continuous) or "classification" (where $y^{(i)} \in \mathcal{Y}$ is a discrete class). A machine learning algorithm is evaluated based on its ability to correctly predict label y' for an unseen data point (x'). Notably, such algorithms represent automated data-driven decision-making which functions by learning from historical decisions, often taken by humans. The utility of such systems (both classification and regression) is optimized by minimizing the errors while training and prediction over given training set. When given an initial set of alternatives, such tasks could represent clustering or classifying a set of alternatives into disjoint groups. Arguably, it is possible that when being trained and optimized for making such decisions (especially for individuals belonging to different protected classes), some classes might be unfairly treated with respect to the outcome and the error rates of the algorithmic decision-making. To account for and avoid such unfairness, the studies in fairness in machine learning has introduced various notions of unfairness. In the next sections, we provide a brief review on various such definitions (Section 3.1) and mechanisms (Section 3.2) of fair machine learning algorithms.

3.1. Fairness Notions

The literature on fair ML algorithms has predominately drawn on the concepts and definition of fairness from legal domain. Popular concepts such as direct discrimination (or "disparate treatment") and indirect discrimination (or "disparate impact") are based on various antidiscrimination laws that prohibit unfair treatment of individuals based on sensitive attributes such as gender, race, etc. [4]. Disparate treatment occurs when the decision an individual user receives is prone to change with respect to changes in her corresponding sensitive attribute information. Similarly, disparate impact occurs when the decision outcomes disproportionately benefit or hurt members of certain sensitive attribute groups. More formally,

Disparate Treatment. Given dataset $D = (A, X, Y)$, with a set of sensitive attributes A (such as race, gender, etc.), remaining attributes X, and binary class to be predicted Y, predicted binary class $\hat{Y}$, disparate treatment is said to exist in data D if

$$Pr(\hat{Y}|X) \neq Pr(\hat{Y}|X, A).$$

Disparate Impact. Given dataset $D = (A, X, Y)$, with a set of sensitive attributes A (such as race, gender,etc), remaining attributes X, and binary class to be predicted Y, disparate impact is said to exist in data D if

$$\frac{Pr(Y = 1|A = 0)}{Pr(Y = 1|A = 1)} \leq \tau = 0.8$$

for positive outcome class 1 and majority protected attribute 1 where $Pr(Y = y|A = a)$ denotes the conditional probability that the class outcome in $y \in Y$ given sensitive attribute $a \in A$.

For the convenience of the readers, in Table 5, we provide concise summary of various definitions of fairness in literature.

In our review, we observed that a majority of recent studies have focused on design of automated decision-making systems that aim at avoiding one or both of these unfairness notions. For example, consider the work of Feldman et al. [48], who developed a test for disparate impact as well as methods by which data might be made unbiased. Luong et al. [49] provided a method of discrimination discovery and prevention from a dataset of historical decisions by adopting a variant of k-NN classification. Zemel et al. [50] proposed a learning algorithm for fair classification that achieved both group fairness and individual fairness by formulating the fairness as an optimization problem of finding a good representation of the data with two competing goals: to encode data as well as possible while simultaneously obfuscating any information about membership in the protected group. Zafar et al. [51] introduced a flexible constraint-based framework to enable the design of fair margin-based classifiers which make use of a general and intuitive measure of decision boundary unfairness. In a more recent work, Zafar et al. [52] introduced an alternative notion of unfairness called **disparate mistreatment**. A classifier is said to suffer from disparate mistreatment if the misclassification rates for different groups of individuals having different values of the sensitive attribute A are different. Zafar et al. [52] proposed that disparate mistreatment in binary classification task can be specified with respect to various misclassification measures such as overall misclassification rate, false positive rate, false negative rate, false omission rate, and false discovery rate. We have also witnessed recent works drawing on fairness concepts from economics and social welfare such as equality, Gini distribution, etc., in its conceptualization of fairness.

Despite gaining tremendous research attention in recent years, a major feature of fair ML literature has been an extensive set of definitions of fairness to choose from and various empirical and theoretical findings suggesting the impossibility of satisfying various fairness definitions at the same time. Nevertheless, in our review, we aim to cluster fairness in machine learning algorithms studied in the extant literature into three main categories, namely, anticlassification, statistical parity, and calibration.

1. Anticlassification, also known as unawareness, seeks to achieve fairness in ML outcomes by excluding the use of protected features such as race, gender, or ethnicity from the statistical model. This notion is consistent with disparate treatment. Despite being intuitive, easy-to-use and having legal support, a crucial difficulty of this approach is that a protected feature might be correlated with many other unprotected features, and it is practically infeasible to identify all such covariate "proxies" and remove them from the statistical model. For example, protected class *race* might be correlated with various other features, such as education level, salary, life-expectancy, etc., and removing all these proxies from the statistical model could have detrimental effects in predictive performance. Consider we have a vector $x_i \in \mathbb{R}^t$ that represents the visible attributes of individual i such as race, gender, education level, age, etc. An algorithmic decision can be represented as a function $d : \mathbb{R}^t \mapsto \{0,1\}$, where $d(x) = k$, $k \in \{0,1\}$, means that action a_k is taken. Suppose that x can be partitioned into protected and unprotected features: $x = (x_p, x_u)$. Let X_p denote the set of all protected features. Then, anticlassification requires that decisions do not consider protected attributes, more formally,

$$d(x) = d(x') \text{ for all } x, x' \text{ such that } x_u = x'.$$

   Several other variants of anticlassification are also proposed in the literature [53,54].
2. Statistical parity (also known by the names of demographic parity, independence, statistical parity, and classification parity) requires that common measures of predictive accuracy and performance errors remain uniform across various groups segmented by the protected features. This includes notions such as statistical parity, equality of accuracy, equality of false positive/false negative rates, and equality of positive/negative predictive values [55–57]. The main idea of this notion is to quantify and equate benefit and harm of the impact of the ML prediction to groups segmented by protected attributes equally and to distribute the errors among different

stakeholders equally [55]. This notion of fairness has recently found application in criminal justice [58] and is consistent with disparate impact.

The measure of classification parity based on false positive rate and the proportion of decisions that are positive have received considerable attention in machine learning domain [55,59,60]. For formal definition, please refer to Table 5.

Recent research by Hu and Chen [61] suggests that the enforcement of statistical parity criteria in the short-term benefits building up the reputation of the disadvantageous minority in labor market in the long run. Note that, a critical flaw of notion of statistical parity is that it is easy to satisfy it by some arbitrary configuration, for example selecting best and qualified candidates from one group and random alternatives from the other group can still satisfy statistical parity. Moreover, the definition also ignores any possible correlation between positive outcome and protected attributes.

3. Calibration requires that ML outcomes remain independent of protected features after controlling for estimated risk. Calibration relates to the fairness of risk scores and requires that for a given risk score, the proportion of individuals re-offending remains uniform across protected groups. Calibration is beneficial as a fairness condition as it does not require much intervention in the existing decision-making process [62]. A major disadvantage of calibration is that it has been shown that risk score can be manipulated to appear calibrated by ignoring information about the favored group [63]. Formally, given risk scores $s(x)$, calibration is satisfied when

$$\Pr(Y = 1 \mid s(x), A) = \Pr(Y = 1 \mid s(x)).$$

Despite these multitude of notions measuring fairness from a diverse perspective, recent research has identified theoretical and empirical evidence that each of them suffer from significant statistical limitations [57]. The above-described notions of fairness only aim to ensure equality between group averages, particularly drawn from protected classes such as gender, race, etc. In contrast, "individual notion" takes into account additional characteristics of individual features and looks into differences between individuals rather than groups. Individual fairness is satisfied when similar individuals are treated similarly. Users are treated as individuals regardless of their group membership (either protected or unprotected group). Individual fairness is quantified by the distance between the predicted outcomes and the distance between the individual characteristics [64]. Josef et al. [65] introduced the study of fairness in multi-armed bandit problems which ensures that given a pool of individuals, a worse individual is never favored over a better one, despite a learning algorithm's uncertainty over the true payoff. A major drawback of the existing individual notion of fairness is the need to make strong initial assumptions. For instance, the notion coined by Dwork et al. [64] assumes the existence of prior agreed upon similarity metric which is nontrivial to compute and that of Joseph et al. [65] requires significant assumptions of the underlying functional form of the relationship between features and labels for any possible practical application. Another drawback pertains to the difficulty in selecting an appropriate metric function to measure the similarity of two inputs [66].

It is also beneficial to note that a new and emerging notion of fairness considers "causal" notion draws on literature on causal discovery and inference in its definitions [67–69]. Another emerging literature proposes that the right notion of fairness depends on the context right notion of fairness, which depends on the context [60,70]. Please refer to Table 5 for specific definition.

Table 5. Different types of fairness in recommender systems.

Fairness Definition	Description
Equalized Odds	Predicted outcome $\hat{Y}$ satisfies equalized odds with respect to protected attribute A and true outcome Y, if $\hat{Y}$ and A are independent conditional on Y, more specifically $P(\hat{Y} = 1 \mid A = 0, Y = y) = P(\hat{Y} = 1 \mid A = 1, Y = y)$ [55]
Equal Opportunity	A binary predictor $\hat{Y}$ satisfies equal opportunity with respect to A and Y if $P(\hat{Y} = 1 \mid A = 0, Y = 1) = P(\hat{Y} = 1 \mid A = 1, Y = 1)$ [55]
Statistical Parity	A predictor $\hat{Y}$ satisfies demographic parity if $P(\hat{Y} \mid A = 0) = P(\hat{Y} \mid A = 1)$ [64]
Counterfactual Fairness	For a given causal model (U, V, F) where $V \equiv A \cup X$, predictor $\hat{Y}$ is said to be "counterfactually fair" if under any context $X = x$ and $A = a$, $P(\hat{Y}_{A \leftarrow a})(U) = y \mid X = x, A = a) = P(\hat{Y}_{A \leftarrow a'}(U) = y \mid X = x, A = a)$, for all y and for any value a' attainable by A [68]
Fairness through awareness	An algorithm is fair if it gives similar predictions to similar individuals. Any two individuals who are similar with respect to a similarity metric defined for a particular task should be classified similarly [64].
Individual fairness	Let $\mathcal{O}$ be a measurable space and $\delta(\mathcal{O})$ be the space of the distribution over $\mathcal{O}$. If $M : \mathcal{X} \mapsto \delta(\mathcal{O})$ denotes a map that maps each individual to a distribution of outcomes, the formulation of individual fairness is then $D(M(\mathcal{X}), M(\mathcal{X}')) \leq d(\mathcal{X}, \mathcal{X}')$, where $\mathcal{X}, \mathcal{X}' \in \mathbb{R}^d$ are two metric functions on the input space and the output space, respectively [64].

3.2. Fairness Mechanisms

Next we turn to discuss the three fairness mechanisms clustered on the timing of the application of debiasing mechanism into preprocessing, in-processing, and postprocessing.

A. Preprocessing. Preprocessing methods deal with removing the protected features or their covariates before training the model. Similar to anticlassification, this method come with severe disadvantages as the protected feature might be correlated with many other unprotected features, and it is practically infeasible to identify all such covariates and exclude them without losing a lot on predictive accuracy. Kamiran and Calders [71] suggest a set of data processing techniques aimed at ensuring fairness for classification tasks. These include suppression, massaging the dataset, reweighting, and sampling.

> **Suppression.** In this process, exactly like anticlassification, all the features that correlate with the protected set of features X_p are first identified which are then removed from the classification model.
> **Massaging the dataset.** In this process, labels of some data points are manipulated in order to remove existing discrimination from the training data. In order to find a good set of labels to change, Kamiran and Calders [71] proposed a combination of ranking and learning.
> **Reweighting.** Instead of changing the labels, in this method the tuples in the training dataset are assigned asymmetric weights in order to overcome the bias
> **Sampling.** Kamiran and Calders [71] introduced "uniform sampling" and "preferential sampling", where the training data is sampled with the help of a ranker as a debiasing method.

Kamiran and Calders [71] found that suppression of the protected attributes does not always result in the removal of bias and massaging and preferential sampling techniques performed best for debiasing with a minimal loss in accuracy.

Another idea developed in preprocessing is to learn a new representation of the data such that it removes the information correlated to the sensitive attribute [50,72,73]. The central algorithm such as classification then use the cleaned data. An advantage of this method is that the analyst can avoid the need to modify the classifier or access sensitive attributes during test time.

B. In-processing. In this method, the optimization procedure is modified to incorporate cost of unfairness. This is typically done by addition of a constraint to the optimizing problem or

addition of cost of fairness as a regularizer. For example, Agarwal et al. incorporate cost-sensitive classification into their original objective function [59]. Given a dataset, $\{(x_i, c_i^0, c_i^1)\}_{i=1}^n$, where c_i^0 is the cost of predicting 0 on x_i and c_i^1 is the cost of predicting 1 on x_i, a cost-sensitive classification algorithm given the dataset outputs

$$\hat{h} = \arg\min_{h \in H} h(x_i)c_i^1 + (1 - h(x_i))c_i^0.$$

where $h(x_i)$ represents the original objective function without cost sensitivity.

More generally, the reduction approach by Agarwal et al. suggests the reduction of training with fairness constraints and solving a series of cost-sensitive classifications using off-the-shelf methods [59].

An important advantage of this method is that there is no need to access sensitive attributes at test time. This method also provides higher flexibility in terms of trade-off between accuracy and fairness measures. An important disadvantage is that this method is task specific and requires modification of classifier which can often exponentially increase the computational complexity.

The method to optimize counterfactual fairness also falls into this category. Kusner et al. [68] propose "counterfactual fairness" that explicitly specifies the assumptions about the data generating process. This can be done by adding a linear or convex surrogate for the fairness constraint in the learning models. For example, consider a predictive problem with fairness considerations, where A, X, and Y represent the protected attributes, remaining attributes, and the output of interest, respectively.

C. Postprocessing. Postprocessing methods require editing the posteriors in order to satisfy the fairness constraints. The method searches for a proper threshold using the original score function for each group. We refer to Hard et al. [55] for more details on this postprocessing method. This method requires test-time access to the protected attribute and lacks flexibility in terms of trade-off between accuracy and fairness. However, this method benefits from being general and applicable to any classifier without any modification.

Besides these, there are also some work on fairness in unsupervised learning. In their recent paper, Bolukbasi et al. [74] analyzed the unfairness present in word embeddings, a popular framework used to represent text data as vectors and quantitatively demonstrate that word-embeddings contain biases in their geometry that reflect stereotypes present in our society (for example words like "programmer" was closer to male names as compared to "homemaker", which was closer to female names). Additionally, the authors also introduce various debiasing methods to deal with detrimental effects of such gender bias. In the similar line, Zhao et al. [75] investigated various datasets and models associated with multilabel object classification and visual semantic role labeling, and found that various datasets for these tasks contain significant gender bias which are amplified by the models trained on these datasets. As an example, the authors found that activities like cooking are highly associated with females as compared to males. Following these works, a large number of subsequent work has been devoted to debiasing techniques biases embedded in word embeddings [76–79].

4. Recommender Systems

Recommender systems are among the most pervasive applications of algorithmic decision-making in industry, with many services using them to support users in finding products or information that are of potential interest [80]. Such systems find applications in various online platforms such as Netflix, Linkedin, Amazon, etc., where the alternative set of items is much larger which needs to be filtered (and a smaller set of items is to be designed) before being presented to the user. There are various approaches for recommender systems available, such as collaborative filtering [81], content-based filtering [82], and knowledge-based recommendation [83], or some hybrid combinations of these. First, collaborative filtering algorithms are based on the assumption of word-of-mouth, that is,

decisions of users are influenced by other users who are closer to her (such as family and friends). User-based collaborative filtering [81] identifies the *k*-nearest neighbors of the focal user and based on these nearest neighbors calculates a prediction of the focal user's rating for a specific item. In contrast to user-based collaborative filtering. Item-based collaborative filtering [84] searches for items rated by focal user that received similar ratings as items currently under investigation in order to estimate the probability of its utility. Second, content-based collaborative filtering [82] is based on the assumption of monotonicity of personal interests. In content-based filtering, the content of already consumed items are compared with those of the new items that can potentially be recommended to the user. Based on some "similarity" measure of such comparisons, items that are likely to be of interest to the focal user are recommended. Third, knowledge-based recommendation [83] also draws on deeper knowledge (such as semantic knowledge) about the items in addition to ratings and textual item descriptions that the first two approaches use.

The study of bias and fairness in recommender systems is an emerging research area that is receiving increasing attention. This is further fueled by evidences of detrimental consequences of popularity bias in recommender systems where recommenders typically emphaize popular items over other "long-tail", less popular ones that may only be popular among small groups of users [85]. Notably, a majority of recommender algorithms can be considered as a subset of machine learning algorithms. Notwithstanding, we discuss them separately here due to their unique importance and application pertaining to fairness, and because studying fairness in recommender systems is considered to be challenging and complex as they often consist of multiple models, must balance multiple goals, and are difficult to evaluate due to sparsity and dynamism.

Like algorithmic fairness in general, the definition of fairness in recommender systems is as well challenging. In traditional recommender systems, the optimization only takes place on the accuracy of performance, that is, how well the algorithm predicts whether a user will like an item or not based on the utilities of users. Literature in fairness of recommender systems adds in constraints or additional objectives in order to ensure sufficient item coverage, fairness or diversity when it comes to item recommendation. Recommender systems with such constraints can better facilitate their adoption and purchase and fairly deal with the wishes and preferences of all classes/groups of users [86].

Similar to the accuracy–fairness trade-off in machine learning, recommender systems as well suffer from utility-fairness conundrum as making the recommendations fair will likely reduce utility of the entire system. Moreover, recommendation systems also suffer from some unique shortcomings as compared to machine learning fairness in general [87]. For instance, in a recent paper, Farnadi et al. [88] defined two primary types of bias drawn from imbalance in data. First, observation bias appears due to the feedback loop in the recommender systems, as item displayed by the recommender system gets further reinforced in the choice by the agent over the period of time, leading to the increase in probability for the item to be retained in the system. Moreover, items similar to such an item also get more weightage by the system to be further recommended. Second, biases that come from imbalance in the data are caused when a systematic bias is present in the data/ experience due to societal or historical features. Literature has explored approaches towards handling such biases by increasing the diversity of recommendations [89,90]. Additionally, a more recent line of research looks at fairness in recommender systems through the use of various metrics. For instance, Yao and Huang [87] adopt five different fairness metrics in their exploration of fair recommender systems based on matrix factorization. Burke [91] introduces fairness via neighborhood balancing with a space linear method.

Although these early works have played a vital role in increasing our understanding of fairness in recommender systems, most of the existing work in fair recommender systems focus on fairness in supervised learning setting, and only very recently are researchers moving towards fairness in unsupervised tasks such as clustering and ranking (See, for example, the works by the authors of [87,91–93].). Below, we provide a more elaborate overview of the existing literature divided into three main clusters: (1) fairness for users and group of users (Section 4.1) where we look at research

work aimed at introducing fairness for users or their groups; (2) fairness for items (Section 4.2) where fairness is introduced from the side of recommended items, and, finally; (3) multi-stakeholder fairness (Section 4.3) where fairness incorporates various stakeholders at the same time. In Table 6, we provide a concise summary of these three domains.

Table 6. Different types of fairness in recommender systems.

Type of RecSys Fairness	Focus	References
User & Group Fairness	Ensure fairness for individual or a group of individuals, protected group incurs rating prediction errors in parity with the nonprotected group.	Yao and Huang [87] Ning and Karypis [94]
Item Fairness	Fairness among item categories when recommended to users	Steck [95] Tsintzou et al. [96]
Multiple Stakeholder Fairness	Fairness for multiple parties involved	Burke [91] Abdollahpouri et al. [97] Mehrotra et al. [98]

4.1. Fairness for Users and Groups of Users

These methods are aimed at ensuring fairness for individual or a group of users. Similar to the classification or statistical parity discussed in Section 3.1, fairness for users consider group fairness in which protected group incurs rating prediction errors in parity with the nonprotected group. Yao and Huang [87] studied fairness in collaborative-filtering settings and identify new fairness metrics that can be optimized by adding fairness terms to the learning objective. They also show via experiments that their new metrics can better measure fairness than the baseline and are effectively useful in reducing bias. In another paper, Ning and Karypis [94] aimed to achieve the same notion of user fairness by adding a regularization term to the collaborative filtering objective function that measures the deviation with respect to the total weight assigned to the protected and nonprotected group member.

A related but separate line of work looks at individual fairness in group recommendation, where the goal is to design systems that recommend to a group of users while respecting the individual preferences of the group members. In such a setting, the objective is not only to maximize the overall satisfaction among group members but also to ensure that the recommendations are fair in terms of minimizing the feeling of dissatisfaction among group members. Earlier work in this line mainly view fairness issues from the perspective of game theory and voting theory by treating the group decision process either as non-cooperative game or as a voting campaign without clearly modeling the trade-off between overall satisfaction and fairness of users [99–102].

In a more recent work, Lin et al. [103] investigated the group recommendation problem from a computational lens. Their method tries to maximize the satisfaction of each group member while minimizing the unfairness between them. The authors conceptualize such fairness-aware group recommendation as a multiobjective optimization problem consisting of two independent objectives: individual fairness and social welfare. In a similar line of research, user fairness is modeled in terms of satisfaction of the user with the group recommendation. Qi et al. [104] propose probabilistic models that capture the preference of a group towards a recommended package, and incorporate fairness into it by ensuring further devouring so that no user is consistently slighted by the item selection in the package. This idea has been further developed in subsequent papers [105,106]. For example, in [105], Serbos et al. develop fairness measures for package recommendation based on "proportionality" and "envy-fairness".

Proportionality. Given a package , P, and a parameter , $\triangle$, we say that a user u likes an item $i \in P$ if i is ranked in the top-$\triangle$% of the preferences of u over all items. Consequently, for a user , u, and a package , P, we say that P is m-proportional for u, for $m \geq 1$, if there exists at least m items in P, which are liked by u.

Envy-freeness. Given a group G, a package P, and a parameter $\triangle$, we say that a user $u \in G$ is *envy-free* for an item $i \in P$, if $r(u,i)$ is in top -$\triangle$% of the preferences in the set $\{r(u,i) : v \in G\}$. Consequently, for a user u, a package P and a group G, we say that the package P is m-envy-free for u, for $m \geq 1$, if u is envy-free for at least m items in P.

The authors develop algorithms that can construct a package of items for a group of users satisfying either proportionality or envy-freeness.

A separate but related line of work looks at individual fairness in group recommendation. Sacharidis [107] looks into the minimum utility a group member receives as the notion of fairness. The author further proposes a technique that is able to rank the items by considering all admissible ways in which a group might reach a decision.

4.2. Fairness for Items

Although recent research has been focused on the importance of identifying fairness and diversity in terms of aspects of user preferences as a quality of recommendations, growing research attention is also being received by fairness in terms of groups of items. For example, researches have looked into algorithms that guarantee fairness among item categories when recommended to users. Steck [95] looked into the application of movie recommendations and suggested that item fairness should ensure that the various (past) areas of interest of a user need to be reflected with their corresponding proportions when making current recommendation. For a particular set of recommendations to be fair, it must contain items from various groups with a ratio that is equal to the group ratio present in the subject's input preferences. To ensure such fairness, the authors propose a greedy iterative re-ranking (postprocessing) algorithm that can construct a list that balances the utility of the objects selected and the list's deviation from the input preferences.

In a similar vein, Tsintzou, Pitoura, and Tsaparas [96] presented another re-ranking method that achieves fairness by recompiling a set of objects such that the ratio of objects from various groups (output bias) is the same as the ratio present in the subject's input preferences (input bias). Such a method is able to avoid amplifying existing biases in the input by iteratively swapping a low-utility nonprotected object with a high-utility protected object.

4.3. Multiple Stakeholder Fairness

A unique characteristic of recommender systems is in facilitating mapping or transaction between parties, such as producers and consumers—a perspective now popularly known as multi-stakeholder recommendation or two-sided markets. Such platforms benefit from integrating the preferences of multiple parties into recommendation generation and evaluation. They are now of common occurrence in online market places designed in a variety of industries such as music (Spotify, Soundcloud, and Pandora), recruitment (LinkedIn), content and entertainment (Dailymotion and Youtube), transportation and housing (Airbnb and Uber), etc. A commonality for all these platforms is that they provide a common place where providers and users congregate and make some form of transactions. While traditional recommender systems focused specifically towards satisfaction of consumer by providing a set of relevant content, these multi-sided recommender systems face the problem of additionally optimizing preferences for providers as well as for platform. Fairness requires multiple parties to gain or lose equally with respect to the recommendations made. Such a system is known as multi-stakeholder recommender system and is gaining a lot of recent research attention [108].

A recent paper by Burke [91] provides a great starting point for research in multi-stakeholder fairness. Burke's framework divides the stakeholders of a given recommender system into three categories—consumers, provides, and platforms—and introduces measures that take into consideration such multisided fairness. In a similar vein, Abdollahpouri et al. [97] describe origins of multistakeholder recommendation, and the landscape of system designs providing illustrative examples of current research. This line of research distinguishes itself from fairness consideration in

earlier works where fairness in recommender systems is typically evaluated on their ability to provide items that satisfy needs and interests of the end user. In the same line of research, Mehrotra et al. [98] propose a conceptual computational framework applying counterfactual estimation techniques in order to understand and evaluate different recommendation policies surrounding the trade-off between relevance and fairness in the absence of A/B tests, a popularmethod of comparing two versions of same method against each other to determine which one performs better.

5. Conclusions

Algorithms are taking increasingly prominent decision-making roles in various applications in societal, organizational, and individual lives. Algorithmic decision-making has proliferated everywhere from legal to medical and from social media to employee recruitment in firms. As algorithmic decisions find themselves in major areas of societal impact, it becomes imperative to ensure that they guarantee some level of fairness and trust, more so when individuals and groups that represent minorities or protected classes in terms of gender, race, etc., are exposed to the detrimental consequences of algorithmic decisions. Motivated by the growing attention and interest of public and academia into fairness in algorithmic decision-making, this article endeavored to collect, survey and synthesize emerging and existing research aimed at introducing fairness in algorithmic decision-making. In this work, we provide a useful and simplified taxonomy of the current state of research in algorithmic fairness with a particular focus on decision-making as an application. Such a taxonomy and framework for analyzing algorithmic fairness research, we believe should be beneficial for future research.

5.1. Challenges and Future Research Directions

Our review also identified various challenges with respect to existing research on fair algorithmic decisions. First, our review identified multiple definitions of what is a fair decision-making algorithm and diverse approaches to ensuring fairness in algorithmic decisions. This becomes particularly true as fairness being a social construct gets measured in various notions that often correspond to differing lens in social sciences, justice, economics, and moral philosophy. Such diverse (and often uncorrelated) definitions and methods on the one hand provides a variety of tools to address different manifestations of bias and discrimination embedded in data. On the other hand existence of different definitions has led the research community into diverging path of research endeavors leading to a defragmented domain of science.

For instance, consider the two salient measures of fairness, (a) algorithmic fairness that requires the score that an algorithm produces to be equally accurate for all members vs. (b) algorithmic fairness that requires that the algorithm produces the same percentage of errors in terms of prediction for each group under consideration. Even though there exists normative commonality across these measures, there is so far no algorithmic solution to achieve parity in both these dimensions. Moreover, there is no consensus in literature on what is the best definition of fairness under a given circumstance. Theoretical and empirical evidence showing that different definitions of fairness cannot be satisfied at once makes it even difficult endeavour for policy-makers. To this end, evaluating each definition and method to decide on which definition and method to consider for a given task is a daunting task. Therefore, it is important for the algorithmic fairness community to move towards a converging path. For instance, a unified framework by Speicher et al. [109] is an important and encouraging first step in this direction. To this end, we hope that this article provides a broad overview for such an effort to successfully be accomplished. Moreover, our view is that such a unified framework should cross domains of algorithms and not just remain limited to machine learning.

Second, our review also discovered that a large majority of above reviewed work is centered on the development of statistical definitions of fairness and methods to expose and remove the corresponding biases. Research efforts need to be directed to bridge the gap between mathematical and algorithmic research in academia and their application in practice. See, for example Veale, Kleek, and Binns for such

a work [110]. In order to make fair algorithms accessible to practitioners, easy-to-use and off-the-shelf tools need to be developed. We have already seen encouraging first steps in this direction from the Human–Computer Interaction (HCI) community (see, e.g., the works by the authors of [111,112]). Future work should therefore aim at linking the definition of fairness studied in research to the definition of fairness based on user's perception. For example, in their recent work Srivastava, Heidari, and Krause [46] found that most simplistic mathematical definitions of fairness (i.e., demographic parity) most closely matches the people's idea of fairness in practice. This association remains true even when the participants were explicitly informed about the existence of other more complicated notions of fairness [46]. In the same vein, Holstein et al. [113] conducted a systematic investigation of commercial product teams' challenges and needs for support in developing fairer ML systems. The study identified various areas of alignment and disconnect between the challenges faced by teams in practice and the solutions proposed in the fair ML research literature. Similar research associating the work in academia and practice should be beneficial in making the fairness in algorithms literature more realistic and more easily accessible to practitioners.

Third, empirical evidence and mathematical proofs have by now extensively established the prevalence of inherent trade-offs between the constraints imposed with the notions of fairness and performance accuracy of algorithms [63,114]. This has practical implications as the designer of the system and the user need to decide on level of performance accuracy that they are willing to forgo in order to ensure fairness constraints. Design of algorithms that aim at handling such trade-offs in a systematic way could be beneficial and further explored.

Fourth, our review has also discovered that fairness in machine learning and recommender systems is excessively focused on supervised learning. Though there has been some progress in unsupervised learning such as word embedding and clustering [5,115–117], it is limited as compared to supervised setting. Future work should further advance fair decision-making with respect to unsupervised learning algorithms.

5.2. Limitations

Note that, due to lack of space and a choosen design to keep the discussion focused, in this article we only focus on the fairness in algorithmic decision-making in three main domains, namely multi-winner voting, machine learning, and recommender systems. It is important to note that, by design, we have not given enough attention to a large and perhaps equally important work on peripheral topics such as fairness in natural language understanding, resource allocation, representation learning, causal learning, etc. This we leave open for the future research to survey.

Author Contributions: Y.R.S. and Y.Y. contributed equally to this work and are arranged in alphabetical order

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rejab, F.B.; Nouira, K.; Trabelsi, A. Health Monitoring Systems Using Machine Learning Techniques. In *Intelligent Systems for Science and Information*; Springer: Cham, Switzerland, 2014; pp. 423–440.
2. Chalfin, A.; Danieli, O.; Hillis, A.; Jelveh, Z.; Luca, M.; Ludwig, J.; Mullainathan, S. Productivity and Selection of Human Capital with Machine Learning. *Am. Econ. Rev.* **2016**, *106*, 124–127.
3. Waters, A.; Miikkulainen, R. GRADE: Machine Learning Support for Graduate Admissions. *AI Mag.* **2014**, *35*, 64–75.
4. Barocas, S.; Selbst, A.D. Big Data's Disparate Impact. *Calif. Law Rev.* **2016**, *104*, 671–732.
5. Caliskan, A.; Bryson, J.J.; Narayanan, A. Semantics Derived Automatically from Language Corpora Contain Human-Like Biases. *Science* **2017**, *356*, 183–186.
6. Angwin, J.; Jeff, L.; Surya, M.; Kirchner, L. Machine Bias. *ProPublica*, 23 May 2016.

7. Epstein, R.; Robertson, R.E. The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E4512–E4521.
8. Conover, M.D.; Ratkiewicz, J.; Francisco, M.R.; Gonçalves, B.; Menczer, F.; Flammini, A. Political Polarization on Twitter. In Proceedings of the ICWSM 2011 5th International Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
9. Garimella, V.R.K.; Weber, I. A Long-Term Analysis of Polarization on Twitter. In Proceedings of the ICWSM 2017 11th International Conference on Web and Social Media, Montréal, QC, Canada, 15–18 May 2017; pp. 528–531.
10. Leavy, S. Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning. In Proceedings of the 1st International Workshop on Gender Equality in Software Engineering, Gothenburg, Sweden, 28–28 May 2018; pp. 14–16.
11. Smith, J.H. Aggregation of Preferences with Variable Electorate. *Econometrica* **1973**, *41*, 1027–1041.
12. May, K.O. A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision. *Econometrica* **1952**, *20*, 680–684.
13. Woodall, D. Properties of Preferential Election Rules. *Voting Matters* **1994**, *3*, 8–15.
14. Skowron, P.; Faliszewski, P.; Slinko, A. Axiomatic Characterization of Committee Scoring Rules. *J. Econ. Theory* **2019**, *180*, 244–273.
15. Dummett, M. *Voting Procedures*; Oxford University Press: Oxford, UK, 1984.
16. Aziz, H.; Lee, B.E. The Expanding Approvals Rule: Improving Proportional Representation and Monotonicity. *arXiv* **2017**, arXiv:1708.07580.
17. Elkind, E.; Faliszewski, P.; Skowron, P.; Slinko, A. Properties of Multiwinner Voting Rules. *Soc. Choice Welf.* **2017**, *48*, 599–632.
18. Tideman, N. The Single Transferable Vote. *J. Econ. Perspect.* **1995**, *9*, 27–38.
19. Tideman, N.; Richardson, D. Better Voting Methods Through Technology: The Refinement-Manageability Trade-off in the Single Transferable Vote. *Public Choice* **2000**, *103*, 13–34.
20. Lu, T.; Boutilier, C. Budgeted Social Choice: From Consensus to Personalized Decision Making. In Proceedings of the IJCAI 2011 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011; pp. 280–286.
21. Betzler, N.; Slinko, A.; Uhlmann, J. On the Computation of Fully Proportional Representation. *J. Artif. Intell. Res.* **2013**, *47*, 475–519.
22. Sánchez-Fernández, L.; Fernández, N.; Fisteus, J.; Basanta-Val, P. Some Notes on Justified Representation. In Proceedings of the M-PREF 2016 10th Multidisciplinary Workshop on Advances in Preference Handling, New York, NY, USA, 9–11 July 2016.
23. Aziz, H.; Brill, M.; Conitzer, V.; Elkind, E.; Freeman, R.; Walsh, T. Justified Representation in Approval-Based Committee Voting. *Soc. Choice Welf.* **2017**, *48*, 461–485.
24. Fernández, L.S.; Elkind, E.; Lackner, M.; García, N.F.; Arias-Fisteus, J.; Basanta-Val, P.; Skowron, P. Proportional Justified Representation. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 670–676.
25. Aziz, H.; Elkind, E.; Huang, S.; Lackner, M.; Fernández, L.S.; Skowron, P. On the Complexity of Extended and Proportional Justified Representation. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 902–909.
26. Aziz, H.; Gaspers, S.; Gudmundsson, J.; Mackenzie, S.; Mattei, N.; Walsh, T. Computational Aspects of Multi-Winner Approval Voting. In Proceedings of the AAMAS 2015 14th International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey, 4–8 May 2015; pp. 107–115.
27. Brams, S.J.; Kilgour, D.M. Satisfaction Approval Voting. In *Voting Power and Procedures*; Fara, R., Leech, D., Salles, M., Eds.; Springer: Berlin, Germany, 2014; pp. 323–346.
28. Brams, S.; Fishburn, P. Approval Voting. *Am. Political Sci. Rev.* **1978**, *72*, 831–847.
29. Janson, S. Phragmén's and Thiele's Election Metholds. *arXiv* **2016**, arXiv:1611.08826.
30. Zhou, A.; Yang, Y.; Guo, J. Parameterized Complexity of Committee Elections with Dichotomous and Trichotomous Votes. In Proceedings of the AAMAS 2019 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, Canada, 13–17 May 2019; pp. 503–510.
31. Brams, S.J.; Kilgour, D.M.; Sanver, M.R. A Minimax Procedure for Electing Committees. *Public Choice* **2007**, *132*, 401–420.

32. Yang, Y.; Wang, J. Complexity of Additive Committee Selection with Outliers. In Proceedings of the AAMAS 2019 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, Canada, 13–17 May 2019; pp. 2291–2293.
33. Yang, Y.; Wang, J. Multiwinner Voting with Restricted Admissible Sets: Complexity and Strategyproofness. In Proceedings of the IJCAI 2018 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 576–582.
34. Yang, Y. On the Tree Representations of Dichotomous Preferences. In Proceedings of the IJCAI 2019 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019.
35. Fernández, L.S.; Fisteus, J.A. Monotonicity Axioms in Approval-based Multi-winner Voting Rules. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2019, Montreal, QC, Canada, 13–17 May 2019, pp. 485–493.
36. LeGrand, R. *Analysis of the Minimax Procedure*; Technical Report; Department of Computer Science and Engineering, Washington University: St. Louis, MO, USA, 2004.
37. Procaccia, A.D.; Rosenschein, J.S.; Zohar, A. On the Complexity of Achieving Proportional Representation. *Soc. Choice Welf.* **2008**, *30*, 353–362.
38. Brill, M.; Freeman, R.; Janson, S.; Lackner, M. Phragmén's Voting Methods and Justified Representation. In Proceedings of the AAAI 2017 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 406–413.
39. Peters, D. Single-Peakedness and Total Unimodularity: New Polynomial-Time Algorithms for Multi-Winner Elections. In Proceedings of the AAAI 2018 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 1169–1176.
40. Yang, Y.; Wang, J. Parameterized Complexity of Multi-winner Determination: More Effort Towards Fixed-Parameter Tractability. In Proceedings of the AAMAS 2018 17th International Conference on Autonomous Agents and MultiAgent Systems, Stockholm, Sweden, 10–15 July 2018; pp. 2142–2144.
41. Celis, L.E.; Huang, L.; Vishnoi, N.K. Multiwinner Voting with Fairness Constraints. In Proceedings of the IJCAI 2018 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 144–151.
42. Monroe, B.L. Fully Proportional Representation. *Am. Political Sci. Rev.* **1995**, *89*, 925–940.
43. Koriyama, Y.; Macé, A.; Treibich, R.; Laslier, J.F. Optimal Apportionment. *J. Political Econ.* **2013**, *121*, 584–608.
44. Brill, M.; Laslier, J.F.; Skowron, P. Multiwinner Approval Rules as Apportionment Methods. In Proceedings of the AAAI 2017 31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 414–420.
45. Bredereck, R.; Faliszewski, P.; Igarashi, A.; Lackner, M.; Skowron, P. Multiwinner Elections with Diversity Constraints. In Proceedings of the AAA 2018 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 933–940.
46. Srivastava, M.; Heidari, H.; Krause, A. Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning. In Proceedings of the KDD 2019 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, Alaska, 4–8 August 2019.
47. Cheng, Y.; Jiang, Z.; Munagala, K.; Wang, K. Group Fairness in Committee Selection. In Proceedings of the EC 2019 20th ACM Conference on Economics and Computation, Phoenix, AZ, USA, 24–28 June 2019; pp. 263–279.
48. Feldman, M.; Friedler, S.A.; Moeller, J.; Scheidegger, C.; Venkatasubramanian, S. Certifying and Removing Disparate Impact. In Proceedings of the KDD 2015 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; pp. 259–268.
49. Luong, B.T.; Ruggieri, S.; Turini, F. k-NN As an Implementation of Situation Testing for Discrimination Discovery and Prevention. In Proceedings of the KDD 2011 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 502–510.
50. Zemel, R.; Wu, Y.; Swersky, K.; Pitassi, T.; Dwork, C. Learning Fair Representations. In Proceedings of the ICML 2013 30th International Conference on International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013; pp. 325–333.
51. Zafar, M.B.; Valera, I.; Gomez-Rodriguez, M.; Gummadi, K.P. Fairness Constraints: A Flexible Approach for Fair Classification. *J. Mach. Learn. Res.* **2019**, *20*, 1–42.

52. Zafar, M.B.; Valera, I.; Gomez Rodriguez, M.; Gummadi, K.P. Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification Without Disparate Mistreatment. In Proceedings of the WWW 2017 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 1171–1180.
53. Bonchi, F.; Hajian, S.; Mishra, B.; Ramazzotti, D. Exposing the Probabilistic Causal Structure of Discrimination. *Int. J. Data Sci. Anal.* **2017**, *3*, 1–21.
54. Grgić-Hlača, N.; Zafar, M.B.; Gummadi, K.P.; Weller, A. Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning. In Proceedings of the AAAI 2018 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 51–60.
55. Hardt, M.; Price, E.; Srebro, N. Equality of Opportunity in Supervised Learning. In Proceedings of the NIPS 2016 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 3315–3323.
56. Kleinberg, J.M.; Mullainathan, S.; Raghavan, M. Inherent Trade-Offs in the Fair Determination of Risk Scores. In Proceedings of the ITCS 2017 8th Innovations in Theoretical Computer Science Conference, Berkeley, CA, USA, 9–11 January 2017; pp. 43:1–43:23.
57. Corbett-Davies, S.; Goel, S. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. *arXiv* **2018**, arXiv:1808.00023.
58. Berk, R.; Heidari, H.; Jabbari, S.; Joseph, M.; Kearns, M.J.; Morgenstern, J.; Neel, S.; Roth, A. A Convex Framework for Fair Regression. *arXiv* **2017**, arXiv:1706.02409.
59. Agarwal, A.; Beygelzimer, A.; Dudík, M.; Langford, J.; Wallach, H.M. A Reductions Approach to Fair Classification. In Proceedings of the ICML 2018 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 60–69.
60. Chouldechova, A. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *arXiv* **2017**, arXiv:1703.00056 .
61. Hu, L.; Chen, Y. A Short-term Intervention for Long-term Fairness in the Labor Market. In Proceedings of the WWW 2018 World Wide Web Conference on World Wide Web, Lyon, France, 23–27 April 2018; pp. 1389–1398, doi:10.1145/3178876.3186044.
62. Barocas, S.; Hardt, M.; Narayanan, A. Fairness and Machine Learning. *arXiv* **2017**, arXiv:1712.03586
63. Corbett-Davies, S.; Pierson, E.; Feller, A.; Goel, S.; Huq, A. Algorithmic Decision Making and the Cost of Fairness. In Proceedings of the KDD 2017 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 797–806.
64. Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; Zemel, R.S. Fairness Through Awareness. In Proceedings of ITCS 2012 3rd Innovations in Theoretical Computer Science, Cambridge, MA, USA, 8–10 January 2012; pp. 214–226.
65. Joseph, M.; Kearns, M.J.; Morgenstern, J.H.; Roth, A. Fairness in Learning: Classic and Contextual Bandits. In Proceedings of the NIPS 2016 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 325–333.
66. Kim, M.P.; Reingold, O.; Rothblum, G.N. Fairness Through Computationally-Bounded Awareness. In Proceedings of the NIPS 2018 Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, Montreal, QC, Canada, 3–8 December 2018; pp. 4847–4857.
67. Kilbertus, N.; Rojas-Carulla, M.; Parascandolo, G.; Hardt, M.; Janzing, D.; Schölkopf, B. Avoiding Discrimination through Causal Reasoning. In Proceedings of the NIPS 2017 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 656–666.
68. Kusner, M.J.; Loftus, J.R.; Russell, C.; Silva, R. Counterfactual Fairness. In Proceedings of the NIPS 2017 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4069–4079.
69. Nabi, R.; Shpitser, I. Fair Inference on Outcomes. In Proceedings of the AAAI 2018 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 1931–1940.
70. Kleinberg, J.M. Inherent Trade-Offs in Algorithmic Fairness. In Proceedings of the SIGMETRICS 2018 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Irvine, CA, USA, 18–22 June 2018; p. 40.
71. Kamiran, F.; Calders, T. Data Preprocessing Techniques for Classification without Discrimination. *Knowl. Inf. Syst.* **2012**, *33*, 1–33.

72. Louizos, C.; Swersky, K.; Li, Y.; Welling, M.; Zemel, R.S. The Variational Fair Autoencoder. In Proceedings of the ICLR 2016 4th International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
73. Gordaliza, P.; Barrio, E.D.; Fabrice, G.; Loubes, J.M. Obtaining Fairness using Optimal Transport Theory. In Proceedings of the ICML 2019 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 2357–2365.
74. Bolukbasi, T.; Chang, K.; Zou, J.Y.; Saligrama, V.; Kalai, A.T. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In Proceedings of the NIPS 2016 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 4349–4357.
75. Zhao, J.; Wang, T.; Yatskar, M.; Ordonez, V.; Chang, K. Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. In Proceedings of the EMNLP 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; pp. 2979–2989.
76. Garg, N.; Schiebinger, L.; Jurafsky, D.; Zou, J. Word Embeddings Auantify 100 Years of Gender and Ethnic Stereotypes. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3635–E3644.
77. Zhao, J.; Wang, T.; Yatskar, M.; Cotterell, R.; Ordonez, V.; Chang, K. Gender Bias in Contextualized Word Embeddings. In Proceedings of the NAACL-HLT 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 629–634.
78. Zhao, J.; Zhou, Y.; Li, Z.; Wang, W.; Chang, K. Learning Gender-Neutral Word Embeddings. In Proceedings of the EMNLP 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November, 2018; pp. 4847–4853.
79. Pitoura, E.; Tsaparas, P.; Flouris, G.; Fundulaki, I.; Papadakos, P.; Abiteboul, S.; Weikum, G. On Measuring Bias in Online Information. *SIGMOD Rec.* **2017**, *46*, 16–21.
80. Jannach, D.; Zanker, M.; Felfernig, A.; Friedrich, G. *Recommender Systems—An Introduction*; Cambridge University Press: Cambridge, UK, 2010.
81. Konstan, J.A.; Miller, B.N.; Maltz, D.; Herlocker, J.L.; Gordon, L.R.; Riedl, J. GroupLens: Applying Collaborative Filtering to Usenet News. *Commun. ACM* **1997**, *40*, 77–87.
82. Pazzani, M.; Billsus, D. Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Mach. Learn.* **1997**, *27*, 313–331.
83. Burke, R. Knowledge-Based Recommender Systems. *Encycl. Libr. Inf. Syst.* **2000**, *69*, 175–186.
84. Sarwar, B.; Karypis, G.; Konstan, J.A.; Riedl, J. Item-based Collaborative Filtering Recommendation Algorithms. In Proceedings of the WWW 2001 10th International World Wide Web Conference, Hong Kong, China, 1–5 May 2001; pp. 285–295.
85. Park, Y.; Tuzhilin, A. The Long Tail of Recommender Systems and How to Leverage It. In Proceedings of the RecSys 2008 2nd ACM Conference on Recommender Systems, Lausanne, Switzerland, 23–25 October 2008; pp. 11–18.
86. Koutsopoulos, I.; Halkidi, M. Efficient and Fair Item Coverage in Recommender Systems. In Proceedings of the IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress, DASC/PiCom/DataCom/CyberSciTech, Athens, Greece, 12–15 August 2018; pp. 912–918.
87. Yao, S.; Huang, B. Beyond Parity: Fairness Objectives for Collaborative Filtering. In Proceedings of the NIPS 2017 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 2925–2934.
88. Farnadi, G.; Kouki, P.; Thompson, S.K.; Srinivasan, S.; Getoor, L. A Fairness-aware Hybrid Recommender System. *arXiv* **2018**, arXiv:1809.09030 .
89. Wasilewski, J.; Hurley, N. Incorporating Diversity in a Learning to Rank Recommender System. In Proceedings of the FLAIRS 2016 29th International Florida Artificial Intelligence Research Society Conference, Key Largo, FL, USA, 16–18 May 2016; pp. 572–578.
90. Lu, F.; Tintarev, N. A Diversity Adjusting Strategy with Personality for Music Recommendation. In Proceedings of the RecSys 2018 5th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2018, co-located with ACM Conference on Recommender Systems, Vancouver, BC, Canada, 7 October 2018; pp. 7–14.

91. Burke, R. Multisided Fairness for Recommendation. *arXiv* **2017**, arXiv:1707.00093.
92. Zehlike, M.; Bonchi, F.; Castillo, C.; Hajian, S.; Megahed, M.; Baeza-Yates, R.A. FA*IR: A Fair Top-*k* Ranking Algorithm. In Proceedings of the CIKM 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 1569–1578.
93. Singh, A.; Joachims, T. Fairness of Exposure in Rankings. In Proceedings of the KDD 2018 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2219–2228.
94. Ning, X.; Karypis, G. SLIM: Sparse Linear Methods for Top-N Recommender Systems. In Proceedings of the ICDM 2011 11th IEEE International Conference on Data Mining, Vancouver, BC, Canada, 11–14 December 2011; pp. 497–506.
95. Steck, H. Calibrated Recommendations. In Proceedings of the RecSys 2018 12th ACM Conference on Recommender Systems, Vancouver, BC, Canada, 2–7 October 2018; pp. 154–162.
96. Tsintzou, V.; Pitoura, E.; Tsaparas, P. Bias Disparity in Recommendation Systems. *arXiv* **2018**, arXiv:1811.01461.
97. Abdollahpouri, H.; Adomavicius, G.; Burke, R.; Guy, I.; Jannach, D.; Kamishima, T.; Krasnodebski, J.; Pizzato, L.A. Beyond Personalization: Research Directions in Multistakeholder Recommendation. *arXiv* **2019**, arXiv:1905.01986.
98. Mehrotra, R.; McInerney, J.; Bouchard, H.; Lalmas, M.; Diaz, F. Towards a Fair Marketplace: Counterfactual Evaluation of the Trade-off Between Relevance, Fairness & Satisfaction in Recommendation Systems. In Proceedings of the 27th ACM Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 2243–2251.
99. Guzzi, F.; Ricci, F.; Burke, R.D. Interactive Multi-party Critiquing for Group Recommendation. In Proceedings of the RecSys 2011 5th ACM Conference on Recommender Systems, Chicago, CA, USA, 23–27 October 2011; pp. 265–268.
100. Dery, L.N.; Kalech, M.; Rokach, L.; Shapira, B. Iterative Voting under Uncertainty for Group Recommender Systems. In Proceedings of the RecSys 2010 4th ACM Conference on Recommender Systems, Barcelona, Spain, 26–30 September 2010; pp. 265–268.
101. Carvalho, L.A.M.C.; Macedo, H.T. Generation of Coalition Structures to Provide Proper Groups' Formation in Group Recommender Systems. In Proceedings of the WWW 2013 22nd International World Wide Web Conference, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 945–950.
102. Carvalho, L.A.M.C.; Macedo, H.T. Users' Satisfaction in Recommendation Systems for Groups: An Approach based on noncooperative games. In Proceedings of the WWW 2013 22nd International World Wide Web Conference, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 951–958.
103. Lin, X.; Zhang, M.; Zhang, Y.; Gu, Z.; Liu, Y.; Ma, S. Fairness-Aware Group Recommendation with Pareto-Efficiency. In Proceedings of the RecSys 2017 11th ACM Conference on Recommender Systems, Como, Italy, 27–31 August 2017; pp. 107–115.
104. Qi, S.; Mamoulis, N.; Pitoura, E.; Tsaparas, P. Recommending Packages to Groups. In Proceedings of IEEE 16th International Conference on Data Mining, Barcelona, Spain, 12–15 December 2016; pp. 449–458.
105. Serbos, D.; Qi, S.; Mamoulis, N.; Pitoura, E.; Tsaparas, P. Fairness in Package-to-Group Recommendations. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 371–379.
106. Qi, S.; Mamoulis, N.; Pitoura, E.; Tsaparas, P. Recommending Packages with Validity Constraints to Groups of Users. *Knowl. Inf. Syst.* **2018**, *54*, 345–374.
107. Sacharidis, D. Top-N Group Recommendations with Fairness. In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, Limassol, Cyprus, 8–12 April 2019; pp. 1663–1670.
108. Abdollahpouri, H.; Burke, R.; Mobasher, B. Recommender Systems as Multistakeholder Environments. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, Bratislava, Slovakia, 9–12 July 2017; pp. 347–348.
109. Speicher, T.; Heidari, H.; Grgic-Hlaca, N.; Gummadi, K.P.; Singla, A.; Weller, A.; Zafar, M.B. A Unified Approach to Quantifying Algorithmic Unfairness: Measuring Individual & Group Unfairness via Inequality Indices. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2239–2248.

110. Veale, M.; Kleek, M.V.; Binns, R. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In Proceedings of the 2018 ACM CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018.
111. Angell, R.; Johnson, B.; Brun, Y.; Meliou, A. Themis: Automatically Testing Software for Discrimination. In Proceedings of the 2018 ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, ESEC/SIGSOFT, Lake Buena Vista, FL, USA, 4–9 November 2018; pp. 871–875.
112. Galhotra, S.; Brun, Y.; Meliou, A. Fairness Testing: Testing Software for Discrimination. In Proceedings of the ESEC/FSE 2017 11th Joint Meeting on Foundations of Software Engineering, Paderborn, Germany, 4–8 September 2017; pp. 498–510.
113. Holstein, K.; Vaughan, J.W.; Daumé, H., III; Dudík, M.; Wallach, H.M. Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need? In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; p. 600.
114. Menon, A.K.; Williamson, R.C. The Cost of Fairness in Binary Classification. In Proceedings of the FAT 2018 Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 107–118.
115. Chierichetti, F.; Kumar, R.; Lattanzi, S.; Vassilvitskii, S. Fair Clustering Through Fairlets. In Proceedings of the NIPS 2017 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5036–5044.
116. Bera, S.K.; Chakrabarty, D.; Negahbani, M. Fair Algorithms for Clustering. *arXiv* **2019**, arXiv:1901.02393
117. Kleindessner, M.; Awasthi, P.; Morgenstern, J. Fair k-Center Clustering for Data Summarization. In Proceedings of the ICML 2019 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 3448–3457.