Multi-agent Social Choice for Dynamic Fairness-aware Recommendation

Robin Burke
robin.burke@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

Nicholas Mattei
nsmattei@tulane.edu
Department of Computer Science,
Tulane University
New Orleans, Louisiana, USA

Vladislav Grozin
vladislav.grozin@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

Amy Voida
amy.voida@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

Nasim Sonboli
nasim.sonboli@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

ABSTRACT

The pursuit of algorithmic fairness requires that we think differently about the idea of the "user" in personalized systems, such as recommender systems. The conventional definition of the user in such systems focuses on the receiver of recommendations, the individual to whom a particular personalization output is directed. Fairness, especially provider-side fairness, requires that we consider a broader array of system users and stakeholders, whose needs, interests and preferences may need to be modeled. In this paper, we describe a framework in which the interests of providers and other stakeholders are represented as agents. These agents participate in the production of recommendations through a two-stage social choice mechanism. This approach has the benefit of being able to represent a wide variety of fairness concepts and to extend to multiple fairness concerns.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Multi-agent systems**; • **Social and professional topics → User characteristics**.

KEYWORDS

recommender systems, fairness, computational social choice

ACM Reference Format:
Robin Burke, Nicholas Mattei, Vladislav Grozin, Amy Voida, and Nasim Sonboli. 2022. Multi-agent Social Choice for Dynamic Fairness-aware Recommendation. In *FairUMAP Workshop 2022*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3511047.3538032

1 INTRODUCTION

Recommender systems are personalized machine learning systems that support users' access to information in applications as disparate as rental housing, job seeking, social media feeds and online dating. The challenges of ensuring fair outcomes in such systems have been addressed in a growing body of research literature surveyed in [15].

The emergence of fairness as a consideration in personalized systems, including recommender systems, poses a challenge for the conception of the "user" in user modeing research. Conventionally, the user of a recommender system is understood to be the individual to whom recommendations are delivered and for whom they are personalized. However, as noted in [4, 10], some applications require that we expand our notion of who should be considered a user – an approach termed "post-userism". Consider the example of a recommender system embedded in a music streaming app. There are artists whose music appears on the platform and which may or may not be recommended to a listener. Artists have interfaces to the system as well, allowing them to upload new music, curate playlists, send posts to their fans, and see statistics about how often their music has been heard: the individuals might be considered as much users of the platform as their listeners. Except for some recent work [16, 17], there has been very little examination of creators' perspectives in personalized sytems.

In this work, we explore a type of post-userist user model in service of recommendation fairness. We seek to model the concerns around fairness that a variety of system stakeholders might have. We introduce an architecture for implementing multistakeholder fairness in recommender systems, in which *fairness concerns* are represented as agents participating in a dynamic social choice environment. We start from the assumption that multiple fairness concerns will be active at any one time, and that these fairness concerns can be relatively unrestricted in form. Secondly, we build the framework to be dynamic in that decisions are always made in the context of historical choices and results.

Our research in fairness examines concepts inspired by the application context of Kiva Microloans, which offers a platform (Kiva.org) for crowd-sourcing the funding of microloans, mostly in the developing world. Kiva's users (lenders) choose among the loan opportunities offered on the platform; microloans from multiple lenders that are aggregated and distributed through third party non-governmental organizations around the world. Kiva Microloans'

FairUMAP, July 4, Barcelona, Spain

ACM ISBN 978-1-4503-9232-7/22/07.
https://doi.org/10.1145/3511047.3538032

mission specifically includes considerations of "global financial inclusion"; as such, incorporating fairness in its recommendation of loans to potential users (lenders) is a key goal. We will use Kiva's platform as an example throughout this paper, but our analytic findings are not specific to this setting.

2 RELATED WORK

Our work addresses two key limitations in existing work on recommendation fairness and extending initial work found in [37]. The first limitation we see in current work is that researchers have generally assumed that the problem of group fairness can be reduced to the problem of ensuring equality of outcomes between a protected and unprotected group, or in the case of individual fairness, that there is a single metric of fairness to be addressed for all individuals.

We believe that the single-group limitation is severe and not representative of realistic recommendation tasks in which fairness is sought. US anti-discrimination law, for example, identifies multiple protected categories relevant to settings such as housing, education and employment including gender, religion, race, age, and others [3]. But even in the absence of such external criteria, it seems likely that any setting in which fairness is a consideration will need to incorporate the viewpoints of multiple groups.

Where fairness for multiple groups has been considered (e.g., [21, 38, 45]), it is defined in the same way for all groups. Seeking to avoid the "abstraction trap" noted in [35], we think it is essential to allow fairness to mean different things for different groups. Consider, for example, a system recommending news articles. Fairness might require that, over time, readers see articles that are geographically representative of their region: rural and urban or uptown vs downtown, for example. But fairness in presenting viewpoints might also require that any given day's set of headlines represent a range of perspectives. These are two different views of what fairness means, entailing different measurements and potentially different types of algorithmic interventions.

The second limitation that we see in current work is that fairness-aware interventions in recommender systems as well as many other machine learning contexts, have a static quality. In many applications, a system is optimized for some criterion and when the optimization is complete, it produces decisions or recommendations based on that learned state [31]. Similar to Freeman et al. [18], we think of fairness as a dynamic state, especially when what is of primary concern are fair outcomes. A recommender system's ability to produce outcomes that meet some fairness objective may be greatly influenced by context: what items are in inventory, what types of users arrive, how fair the most recent set of recommendations has been, and many others. A static policy that is not sensitive to context runs the risk of failing to capitalize on opportunities to pursue fairness when they arise and/or trying to impose fairness when its cost is high.

Freeman et al. [18] investigate what they call *dynamic social choice functions* in settings where a fixed set of agents select a single item to share over a series of time steps. The work focuses on overall utility to the agents instead of considering the multiple sides of the recommendation interaction. Their problem is fundamentally a voting problem since all agents share the result, whereas we are focused on personalized recommendation.

There have been a number of efforts that explicitly consider the multisided nature of fairness in recommendation and matching platforms. Patro et al. [33] investigate fairness in two-sided matching platforms where there are both producers and consumers. They note, as we do, that optimizing fairness for only one side of the market can lead to very unfair outcomes to the other side of the market. Patro et al. [33] also appeal to the literature on the fair allocation of indivisible goods from the social choice literature [41]. They devise an algorithm that guarantees Max-min share fairness of exposure to the producer side of the market and envy-free up to one item to the consumer side of the market. Their work is closest to the allocation phase of our algorithm. However, in contrast to our work they only use exposure on the producer side and relevance on the consumer side as fairness metrics, whereas our work aims to capture additional definitions. Also, we note that envy-freeness is only applicable when valuations are shared: a condition not guaranteed in a personalized system. It is possible for a user with unique tastes to receive low utility recommendations and still not prefer another user's recommendation lists. Also, our fairness formulation extends beyond the users receiving recommendations to providers of recommended items and envy-freeness provides no way to compare users who are getting different types of benefits from a system. In addition our fairness definitions are dynamic, a case not considered by [33].

Like Patro et al. [33], the work of Sühr et al. [40] investigates fairness in two-sided platforms, specifically those like Uber or Lyft where income opportunities are allocated to drivers. However, unlike our work and the work of Patro et al. [33], Sühr et al. [40] take proportionality as their definition of fairness, specifically proportionality with respect to time in a dynamic setting, and ensure that there is a fair distribution of income to the provider side of the platform.

The architecture presented here advances and generalizes the approach found in [37]. Like that architecture, fairness concerns are represented as agents and interact through social choice. However, in [37], the allocation mechanism selects only a single agent at each time step and the choice mechanism has a fixed, additive, form. We allow for a wider variety of allocation and choice mechanisms, and therefore present a more general solution.

Ge et al. [20] investigate the problem of long term dynamic fairness in recommendation systems. They, like our work, highlight the need to ensure that fairness is ensured as a temporal concept and not see recommendation as a static, one off, decision. To this end they propose a framework to ensure fairness of exposure to the producers of items by casting the problem as a constrained Markov Decision Process where the actions are recommendations and the reward function takes into account both utility and exposure. Ge et al. [20] propose a novel actor-critic deep reinforcement learning framework to accomplish this task at the scale of large recommender systems with very large user and item sets. Again, this work fixes definitions of fairness a priori, although their learning methodology may serve as inspiration to our allocation stage problems in the future.

Morik et al. [27] investigate the problem of learning to rank over large item sets while ensuring fairness of merit based guarantees to groups of item producers. Specifically, they adapt existing methods to ensure that the exposure is *unbiased*, e.g., that it is not subject to rich-get-richer dynamics, and *fairness* defined as exposure being proportional to merit. Both of these goals are built into the regularization of the learner. In essence the goal is to learn user preferences while ensuring the above two desiderata. In contrast, our work factors out the recommendation methodology and we encapsulate the desired fairness definitions as separate agents rather than embedded in the learning algorithm.

3 FORMALIZING FAIRNESS CONCERNS

A central tenet of our work is that fairness is a contested concept [29]. From an application point of view, this means that ideas about fairness will be grounded in specific contexts and specific stakeholders, and that these ideas will be multiple and possibly in tension with each other. From a technical point of view, this means that any fairness-aware recommender system should be capable of integrating multiple fairness concepts, arising as they may from this contested terrain.

A central concept in this work is the idea of a *fairness concern*. We define a fairness concern as a specific type of fairness being sought, relative to a particular aspect of recommendation outcomes, evaluated in a particular way. We believe it is important to reify this concept, since the objective of algorithmic fairness is often assumed to be obvious and conforming to one of a handful of standard definitions [35]. A fairness concern models specific preferences that a particular stakeholder may have concerning the outcomes the system produces. For example, a possible fairness concern in the microlending context might be group fairness relative to different geographical regions considered in light of the exposure of loans from these regions in recommendation lists.[1] Such a concern might arise from organizational stakeholders seeking to operationalize Kiva's organizational mission around global financial inclusion. The concern identifies a particular aspect of the recommendation outcomes (in this case, their geographical distribution), the particular fairness logic and approach (more about this below), and the metric by which fair or unfair outcomes are determined. We can think of the fairness concern as a type of model of stakeholder groups or individuals for the purposes of controlling recommendation generation.

The first consideration in building a fairness-aware recommender system is the question of what fairness concerns surround the use of the recommender system itself. Many such concerns may arise and like any system-building enterprise, there are inevitably tradeoffs involved in their formulation. An initial step in fairness-aware recommendation is for an organization to consult its institutional mission and its internal and external stakeholders with the goal of eliciting and prioritizing fairness concerns. An example of this kind of consultation can be seen in the WeBuildAI project [23] and its participatory design framework for AI.

In addition to addressing different aspects of system outcomes, fairness concerns may invoke different logics of fairness. Welfare economists have identified a number of such logics and we follow Moulin [28] who identifies four:

[1]We are currently conducting research to characterize fairness concerns appropriate to Kiva's recommendation applications. At this stage, we can only speculate about the fairness concerns that might arise in that work. None of the discussion here is intended to represent design decisions or commitments to particular concerns and/or their formulation.

Exogenous Right: A fairness concern is motivated by exogeneous right if it follows from some external constraint on the system. For example, the need to comply with fair lending regulations may mean that male and female borrowers should be presented proportionately to their numbers in the overall loan inventory.

Compensation: A fairness concern that is a form of compensation arises in response to observed harm or extra costs incurred by one group versus others. For example, loans with longer repayment periods are often not favored by Kiva users because their money is tied up for longer periods. To compensate for this tendency, these loans may need to be recommended more often.

Reward: The logic of reward is operational when we consider that resources may be allocated as a reward for performance. For example, if we know that loans to large cooperative groups are highly effective in economic development, we may want to promote such loans as recommendations so that they are more likely to be funded and realize their promise.

Fitness: Fairness as fitness is based on the notion of efficiency. A resource should go to those best able to use it. In a recommendation context, it may mean matching items closely with user preferences. For example, when loans have different degrees of repayment risk, it may make sense to match the loan to the risk tolerance of the lender.

It is clear that fairness logics do not always pull in the same direction. The invocation of different logics are often at the root of political disagreements: for example, controversies over the criteria for college admissions sometimes pit ideas of reward for achievement against ideas of compensation for disadvantage.

Recommender systems often operate as two-sided platforms, where one set of individuals are receiving recommendations and possibly acting on them (consumers), and another set of individuals is creating or providing items that may be recommended (providers) [9]. Consumers and providers are considered, along with the platform operator, to be the direct stakeholders in any discussion of recommender system objectives. Fairness concerns may derive from any stakeholder, and may need to be balanced against each other.

It is also worth noting that the platform may be interested in enforcing fairness, even when other stakeholders are not. For example, the average recommendation consumer might only be interested in the best results for themselves, regardless of the impact on others. Fairness concerns can also arise on behalf of other, indirect, stakeholders who are impacted by recommendations but not a party to them. An important example is *representational fairness*, which directs attention to the way the outputs of a recommender system operate to represent the world and classes of individuals within it: for example, the way the selection of news articles might end up representing groups of people unfairly [30]. (See [15] for additional discussion). As a practical matter, representational fairness concerns can be handled in the same way as provider-side fairness for our purposes here.

Finally, we have the consideration of group versus individual fairness. This dichotomy is well understood as a key difference across types of fairness concerns, defining both the target of measurement of fairness and the underlying principle being upheld. Group fairness requires that we seek fairness across the outcomes relative to predefined protected groups. Individual fairness asks whether each individual user has an appropriate outcome and assumes that users with similar profiles should be treated the same. Just as there are tensions between consumer and provider sides in fairness, there are fundamental incompatibilities between group and individual fairness. Treating all of the outcomes for a group in aggregate is inherently different than maintaining fair treatment across individuals considered separately. Friedler et al. offer a thorough discussion of this topic [19].

It is important to note that our goal in this work is not necessarily fair recommendation in the sense of a perfect score on every agent's fairness metric, but rather fairness-enhanced recommendation. Given the constraints of recommendation applications, in particular the need for personalization, it may not be possible to provide a successful recommendation experience at the same time as achieving fairness as defined by each operative concern. The goal is to enhance fairness relative to each concern, improving on what the base recommendation algorithm would deliver unassisted.

3.1 Fairness Agents

Our architecture SCRUF-D (Social Choice for Recommendation Under Fairness – Dynamic) builds on the SCRUF architecture introduced in [11, 37]. It is designed to allow multiple fairness concerns to operate simultaneously in a recommendation context. Fairness concerns, derived from stakeholder consultation, model these stakeholder preferences through form of fairness agents, each having three capabilities:

- **Evaluation:** A fairness agent can evaluate whether the current historical state is fair, relative to its particular concern. Without loss of generality, we assume that this capability is represented by a function m_i for each agent i that takes as input a history of the system's actions and returns an number in the range $[0, 1]$ where 1 is maximally fair and 0 is totally unfair, relative to the particular concern.
- **Compatibility:** A fairness agent can evaluate whether a given recommendation context represents a good opportunity for its associated items to be promoted. We assume that each agent i is equipped with a function c_i that can evaluate a user profile ω and associated information and return a value in the range $[0, 1]$ where 1 indicates the most compatible user and context and 0, the least.
- **Preference:** An agent can compute a preference for a given item whose presence on a recommendation list would contribute (or not) to its particular fairness concern. Again, without loss of generality, we assume this preference can be realized by a function that accepts an item as input and returns a preference score in $\mathbb{R}_+$ where a larger value indicates that an item is more preferred.

Putting all of these dimensions together gives us a three-dimensional ontology of fairness concerns in recommendation: fairness logic, consumer- vs provider-side, and group vs individual target.

3.2 Recommendation Process

We assume a recommendation generation process that happens over a number of time steps t as individual users arrive and recommendations are generated on demand. Users arrive at the system one at a time, receive recommendations, act on them (or not), and then depart. When a user arrives, a recommendation process produces a recommendation list ℓ_s that represents the system's best representation of the items of interest to that user, generated through whatever recommendation mechanism is available. We do not make any assumptions about this process, except that it is focused on the user and represents their preferences. A wide variety of recommendation techniques are well studied in the literature, including matrix factorization, neural embeddings, graph-based techniques, and others.

The first step to incorporating fairness into the recommendation process is to determine which fairness concerns / agents will be active in responding to a given recommendation opportunity. This is the *allocation phase* of the process, the output of which is a vector of non-negative weights $\vec{\beta}$, summing to one, over the set of fairness agents, indicating to what extent each fairness agent is considered to be allocated to the current opportunity.

Once the set of fairness agents have been allocated, they have the opportunity to participate in the next phase of the process, which is the *choice phase*. In this phase, all of the active (non-zero weighted) agents and their weights participate in producing a final list of recommendations for the user. We view the recommendation algorithm itself as being an agent that participates in this phase.

4 THE SCRUF-D ARCHITECTURE

The two phases of the SCRUF-D architecture are detailed in Figures 1 and 2. The original SCRUF framework [37] concentrated on the representation of user preferences, as computed by the recommender system, and fairness concerns, as derived from stakeholder consultation as discussed in Section 3.1, and their integration. SCRUF-D incorporates the history of system decisions and the fairness achieved over time to control the allocation of fairness concerns. We will first provide a high level overview of the system and describe each figure in detail with formal notation.

4.1 Overview

We can think of a recommender system as a two-sided market in which the recommendation opportunities that arise from the arrival of a user $u \in \mathcal{U}$ to the system, and each are allocated to a set of items $v \in \mathcal{V}$ from the system's catalog. This market has some similarities to various forms of online matching markets including food banks [1], kidney allocation [2, 25], and ride sharing [13], in that users have preferences over the items; however, in our case this preference is known only indirectly and with uncertainty through either the prior interaction history or a recommendation function. Additionally, the items are not necessarily consumable or rivalrous. For example, a loan can be recommended to any number of users – it is not "used up" in the recommendation interaction.[2] Even when

[2]Loans on Kiva's platform may be exhausted eventually through being funded, but recommending a loan is neither a necessary nor a sufficient condition for a user to decide to fund it. Many other objects of recommendation such as streaming media assets, news articles, or social media posts are effectively infinitely available.

an item has limited availability (such as a job posting where only a single person will be hired), it is still the case that the limit on recommendation delivery of such items is not a hard constraint, as it is in a matching market, but rather a soft constraint. If it is desirable that an item in limited supply not be recommended to too many users, what counts as "too many" is a function of how likely users are to select that item from among the recommendations they see.

This problem has some similarities with those found in computational advertising, where messages are matched with users in a personalized way [42, 43]. Because advertising is a paid service, these problems are typically addressed through mechanisms of monetary exchange, such as auctions. There is no counterpart to budgets or bids in our context, which means that solutions in this space do not readily translate to fair recommendation [14, 44, 46].

Once we have a collection of fairness agents we must solve two interrelated problems: (1) what agents are allocated to a particular recommendation *opportunity* and (2) how do we *balance* between the allocated agents and the user's individual preferences? Figure 1 shows the first phase of this process, allocation [7], in which we decide which fairness concerns / agents should be allocated to a particular fairness opportunity. This is an online and dynamic allocation problem where we must consider many factors including the history of agent allocations so far, the generated lists from past interactions with users, and how fair the set of agents believes this history to be. As described in Section 3.1, agents take these histories and information about the current user profile and calculate two values: *m*, a measure of fairness relative to their agent-specific concern, and *c*, a measure of compatibility between the current context and the agent's fairness concern. The allocation mechanism takes these metrics into account producing a probability distribution over the fairness agents that we call the *agent allocation*, which can be interpreted as weights in the choice stage or be used to select a single agent via a lottery, e.g., a randomized allocation scheme [8].

In the second phase, shown in Figure 2, the recommender system generates a list of options, considered to represent the user's preferences. The fairness concerns generate their own preferences as well: lists of items and scores. The choice mechanism combines these preferences of both the user and fairness agents, along with the allocation weights of the fairness agents, to arrive at a final recommendation list to be delivered to the user. The list itself, and possibly interactions the user has with it, become a new addition to the choice history and the process continues for the next user.

4.2 Formal Description

In our formalization of a recommendation system setting we have a set of users $\mathcal{U} = \{u_1, \dots u_n\}$ and a set of items $\mathcal{V} = \{v_1, \dots, v_m\}$. For each item $v_i \in \mathcal{V}$ we have a k-dimensional feature vector $\phi = \langle \phi_1, \dots \phi_k \rangle$ over a set of categorical features ϕ, each with finite domain. Some of these features may be sensitive, e.g., they are associated with one or more fairness agent concerns, we denote this set as ϕ^s. Without loss of generality, we assume that all elements in $\mathcal{V}$ share the same set of features ϕ. Finally, we assume that each user is associated with a profile of attributes $\omega = \langle \omega_1, \dots \omega_j \rangle$, of which some also may be sensitive $\omega^s \subseteq \omega$, e.g., they are associated with one or more fairness agents.

As in a standard recommendation system we assume that we have a recommendation mechanism that takea a user profile ω and items v and produces a predicted rating $\hat{r} \in \mathbb{R}_+$. We will often refer to a recommendation list, $\ell = \langle \{v_1, \hat{r}_1\}, \dots \{v_i, \hat{r}_i\} \rangle$, which is generated for user u by sorting according to $\hat{r}$, i.e., $sort(\mathcal{R}_i(u, \mathcal{V})) \rightarrow \ell$. Note that this produces a permutation (ranking) over the set of items for that user, i.e. a recommendation. As a practical matter, the recommendation results will almost always contain a subset of the total set of items, typically the head (prefix) of the permutation up to some cutoff number of items or score value. For ease of exposition we assume we are able to score all items in the database.

In the SCRUF-D architecture, fairness concerns map directly onto agents $\mathcal{F} = \{f_1, \dots, f_i\}$. In order for the agents to be able to evaluate their particular concerns, they take account of the current state of the system and voice their evaluation of how fairly the overall system is currently operating, their compatibility for the current recommendation opportunity, and their preference for how to make the outcomes more fair. Hence, each fairness agent i is described as a set, $f_i = \{m_i, c_i, \mathcal{R}_i\}$ consisting of a fairness metric, $m_i(\vec{L}, \vec{H}) \rightarrow [0, 1]$, that takes a choice history $\vec{L}$ and allocation history $\vec{H}$ and produces a value in $[0, 1]$ according to the agent's evaluation of how fair recommendations so far have been; a compatibility metric, $c_i(\omega) \rightarrow [0, 1]$, that takes a particular user profile u and produces a value in $[0, 1]$ for how compatible fairness agent i believes they are for user u; and a ranking function, $\mathcal{R}_i(u, v) \rightarrow \{v, \hat{r}\}$, that gives the fairness agent preferences.

In the allocation phase (Figure 1), we must allocate a set of fairness agents to a recommendation opportunity. Formally, this is an allocation function, $\mathcal{A}(\mathcal{F}, m_{\mathcal{F}}(\vec{L}, \vec{H}), c_{\mathcal{F}}(u)) \rightarrow \beta \in \mathbb{R}_+^{|\mathcal{F}|}$ that takes a set of fairness agents $\mathcal{F}$, the agents' fairness metric evaluations $m_{\mathcal{F}}(\vec{L}, \vec{H})$, and the agents' compatibility metric evaluations $c_{\mathcal{F}}(\omega)$ and maps to an agent allocation $\vec{\beta}$, where $\vec{\beta}$ is a probability distribution over the agents $\mathcal{F}$. The allocation function itself is allocating fairness agents to recommendation opportunities by considering both the fairness metric for each agent as well as each fairness agent's estimation of their compatibility.

The allocation function can take many forms, e.g., it could be a simple function of which ever agent voices the most unfairness in the recent history [37], or it could be a more complex function from social choice theory such as the probabilistic serial mechanism [6] or other fair division or allocation mechanisms. Note here that the allocation mechanism is directly comparing the agent valuations of both the current system fairness and compatibility. Hence, we are implicitly assuming that the agent fairness evaluations are comparable. While this is a somewhat strong assumption, it is less strong than assuming that fairness and other metrics, e.g., utility or revenue, are comparable as is common in the literature [47]. So, although we are assuming different evaluation of fairness are comparable, we are only assuming that fairness is comparable with fairness, and not other aspects of the system. We plan to explore options for the allocation function in our empirical experiments. We track the outputs of this function as the allocation history, $\vec{H} = \langle \vec{\beta}^1, \dots, \vec{\beta}^t \rangle$, an ordered list of agent allocations $\vec{\beta}$ at each time.

In the second phase of the system (Figure 2), we must take the set of allocated agents and combine their preferences (and

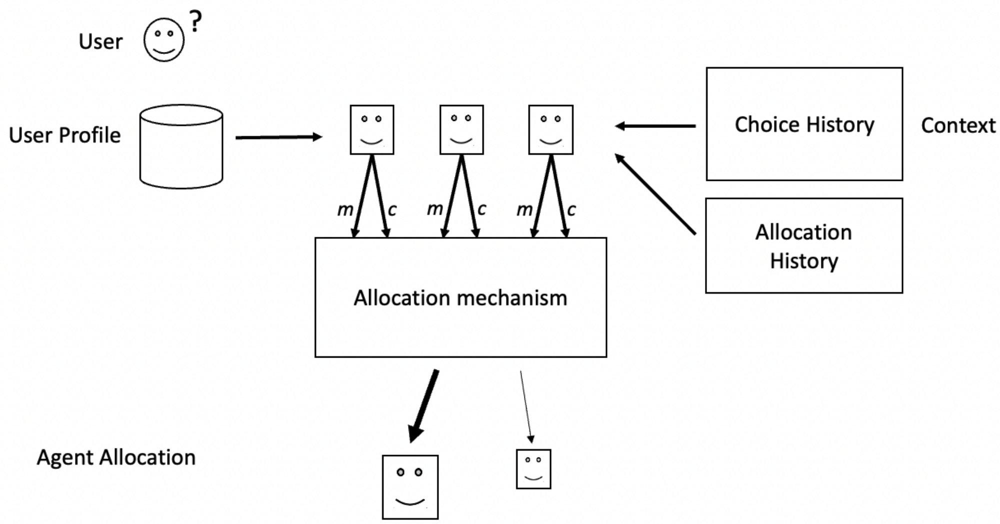

Figure 1: Allocation Phase: Recommendation opportunities are allocated to fairness concerns based on the context.

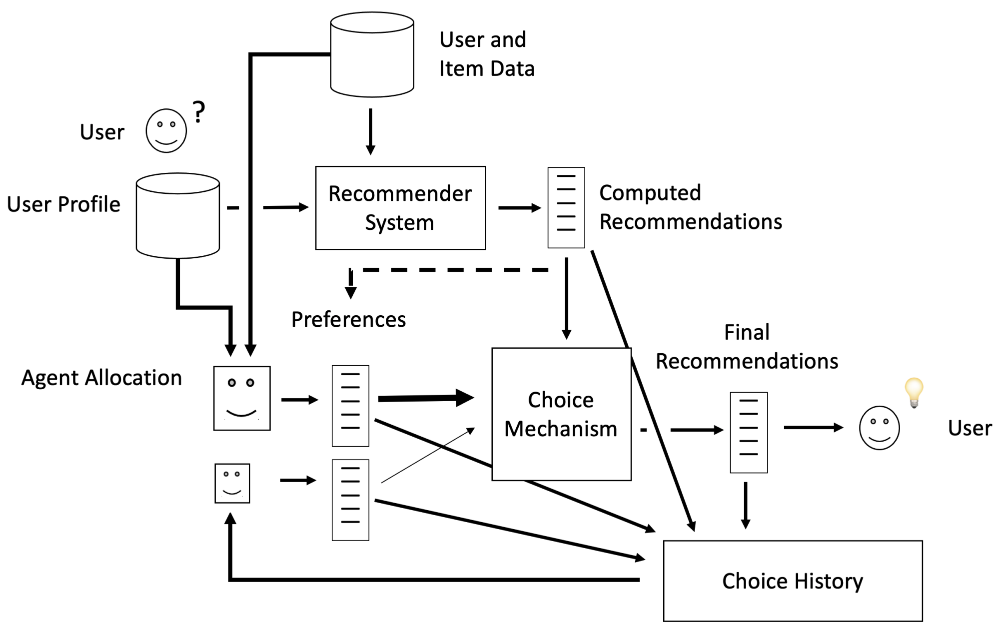

Figure 2: Choice Phase: The preferences from the recommender system and the fairness agents are integrated by the choice mechanism.

weights) with those of the current user u. To do this we define a choice function, $C(\ell, \beta, \ell_{\mathcal{F}}) \rightarrow \ell_C$, as a function from a recommendation list ℓ, agent allocation $\vec{\beta}$, and fairness agent recommendation list(s) $\ell_{\mathcal{F}}$ to a combined list ℓ_C. Each of the fairness agents is able to express their preferences over the set of items for a particular user, $\mathcal{R}_i(u, v) \rightarrow \{v, \hat{r}\}$, and we take this set of lists,$\ell_{\mathcal{F}} = \{\mathcal{R}_1(u, \mathcal{V}), \ldots, \mathcal{R}_i(u, \mathcal{V})\}$, as input to the choice function that generates a final recommendation that is shown to the user, ℓ_C.

We again leave this choice function unspecified as this formulation provides a large design space: we could use a simple voting rule, a simple additive utility function or something much more complicated like rankings over the set of all rankings [7]. Note that the choice function can use the agent allocation β as either a lottery to, e.g., select one agent to voice their fairness concerns, or as a weighting scheme. We will investigate a range of choice functions in further research. In order for the fairness agents to be able to evaluate the status of the system we also track the choice history, $\vec{L} = \langle \ell^t, \ell^t_{\mathcal{F}}, \ell^t_C \rangle$, as an ordered list of user recommendation list ℓ, agent recommendation list(s) $\ell_{\mathcal{F}}$, and choice function output list ℓ_C, indexed by time step t.

5 EXAMPLE

In this section, we work through a detailed example demonstrating the function of the architecture through several iterations of user arrivals. Consider the following set of fairness agents and their associated evaluations and preferences. We assume in this example that in all cases the agents' compatibility functions follow the pattern described in [38] where the entropy of the user profile relative to the sensitive feature is calculated and users with high entropy are determined to be good targets for fairness-enhancing interventions:

f_H**: Health** This agent is concerned with promoting loans to the health sector. Its evaluation function compares the proportion of loans in the database in the health sector against the proportion of health recommendations in the recommendation list history. Its preference function is binary: if the loan is in the health sector, the score is 1; otherwise, zero.

f_A**: Africa** This agent is concerned with promoting loans to Africa. Its evaluation function, however, is list-wise. It counts lists in the recommendation if they have a least one loan recommendation to a country in Africa, and consider a fair outcome one in which every list has at least one such loan. Its preference function will be similarly binary as the f_H agent.

f_G**: Gender Parity** This agent is concerned with promoting gender parity within the recommendation history. If, across the previously generated recommendation lists, the number of men and women presented is proportional to their prevalence in the database, its evaluation will return 1. However, it is preference function is more complex than those above. If the women are underrepresented in the history, it will prefer loans to female borrowers, and conversely for men.[3]

f_L**: Large** This agent is concerned with promoting loans with larger total amounts: over $5,000. Research has shown that such loans are often very productive because they go to cooperatives and have a larger local impact. However, internal research has shown that Kiva users are less likely to support them because each contribution has a smaller relative impact. This agent is similar to the f_A agent above in that it seeks to make sure each list has one of these larger loans.

Consider the contents of Table 1. For the sake of example, we will assume these loans, characterized by the Region, Gender, Section and Amount, constitute the set of loans available for recommendation.

For the sake of exposition, we posit two very simple mechanisms for allocation and choice. We will assume that our allocation mechanism is a single outcome lottery, e.g., a randomized allocation mechanism [8]. One agent will be chosen to participate in the choice mechanism, based on a random draw with probabilities based on the historic unfairness and user compatibility as measured by each agent. There are many more sophisticated algorithms for allocation and choice, which we will explore in future work.

[3]Note: Kiva's borrower database currently recognizes only binary gender categories.

	ϕ^s_1 : Region	ϕ^s_2 : Gender	ϕ^s_3 : Sector	ϕ^s_4 : Amount
v_1	Africa	Male	Agriculture	$5,000-$10,000
v_2	Africa	Female	Health	$500-$1,000
v_3	Middle-East	Female	Clothing	$0-$500
v_4	Central America	Female	Clothing	$5,000-$10,000
v_5	Central America	Female	Health	$0-$500
v_6	Middle-East	Female	Clothing	$0-$500

Table 1: Set of Potential Loans.

We assume that the recommendation lists are of size 3 and the choice mechanism uses a weighted voting / score-based mechanism [7] using a weighted sum of 0.75 on the personalized results for the recommender system and 0.25 on the output of the allocated fairness agent.

5.1 Users

At time t_1, **User** u_1 arrives at the system and the recommendation process is triggered. The user has previously supported small loans only in Central America and Middle East, but has lent to a wide variety of sectors and genders.

For the sake of example, we will assume that the agents measure their prior history relative to their objectives as equally unfair at 0.5, except the Gender Parity agent, which starts out at parity and therefore returns a value of 1. However, the compatibility functions for f_A and f_L returns lower scores because of the user's historical pattern of lending. This yields a lottery in which f_G has probability zero, f_A has a low probability, and f_H a higher one. The allocation mechanism chooses randomly, and we will assume that f_H, the health-focused agent, is picked.

The recommender returns the following list of items and predicted ratings $[\{v_6, 0.6\}, \{v_4, 0.5\}, \{v_5, 0.3\}, \{v_3, 0.3\}, \{v_1, 0.0\}, \{v_2, 0.0\}]$. The f_H agent gives a score of 1 to the health-related loans v_2 and v_5 and 0 to all others.

The choice mechanism combines these scores as described above and returns the final recommendation list $[\{v_5, 0.475\}, \{v_6, 0.45\}, \{v_4, 0.375\}]$. Note that the Health agent has successfully promoted its preferred item to the first position in the list. For the sake of example, we assume that the agents' evaluation functions are very sensitive. Therefore, when **User** u_2 arrives, the results of the previous recommendations have caused the evaluations to shift such that the Health f_H and Large f_L agents are now satisfied (note that v_4 is included in u_1's list and it was a large loan), the Gender parity agent f_G is now at 0.9 (note that there is only one male loan in the database) but the Africa agent f_A, which got nothing in u_1's list is now at 0.25. We assume that u_2 is similar to u_1 in profile and therefore compatibility, but f_A has a much worse fairness score than f_G, and therefore a high allocation probability. We will assume f_A is chosen.

Because this user has similar preferences to u_1, they get the same recommendations: $[\{v_6, 0.6\}, \{v_4, 0.5\}, \{v_5, 0.3\}, \{v_3, 0.3\}], \{v_1, 0.0\}, \{v_2, 0.0\}]$. The f_A agents scores the two loans from Africa (v_1 and

v_2) at 1 and the others at 0. So, after randomly breaking the tie between v_1 and v_2, the final recommendation list is $[\{v_6, 0.45\}, \{v_4, 0.375\}, \{v_1, 0.25\}]$.

In this iteration, all four agents find themselves scoring fairness at 1 over the evaluation window and so no agents are allocated when **User** u_3 arrives. The results from the recommendation algorithm pass through the choice mechanism unchanged.

6 DESIGN CONSIDERATIONS

Within this framework there are a number of important design considerations to take into account for any particular instantiation of the SCRUF-D architecture. We have left many of the particular design choices open for future investigation. We allow for any type of recommendation algorithm; fairness agents representing users' concerns may incorporate any type of compatibility function or fairness evaluation function. Similarly, we do not constrain the allocation or choice mechanisms. With SCRUF-D, we are able to explore many definitions of fairness and recommendation together in a principled uniform way. In this section, we discuss a few of the parameters that may be explored in future work.

6.1 Agent Design

As noted above, agents may have preferences over disjoint sets of items or they may be constrained only to have preferences over the items produced by the recommender system for the given user. This second option corresponds to a commonly-used *re-ranking* approach, where the personalization aspect of the system controls what items can be considered for recommendation and fairness considerations re-order the list [15]. If an agent can introduce any item into its preferences, then we may have the challenge of integrating items that are ranked by some agents but not others. Some practical work-arounds might include a constraint on the recommender system to always return a minimum number of items of interest to the allocated agents or a default score to assign to items not otherwise ranked.

Despite our terminology, it is clear that our architecture as described is sufficiently general that an agent could be designed that pushes the system to act in harmful and unfair ways rather than beneficial and fairness-enhancing ones. Thus, the importance of the initial step of stakeholder consultation and the careful crafting of fairness concerns. Because fairness concerns are developed within a single organization and with beneficence in mind, we assume that we do not need to protect against adversarial behavior, such as collusion among agents or strategic manipulation of preferences. The fact that the agents are all "on the same team" allows us to avoid constraints and complexities that otherwise arise in multi-agent decision contexts.

6.2 Agent Efficacy

The ability of an agent to address its associated fairness concern is non-deterministic. It is possible that the agent may be allocated to a particular user interaction, but its associated fairness metric may still fail to improve. One likely reason for this is the primacy of the personalization objective. Generally, we expect that the user's interests will have the greatest weight in the final recommendations delivered. Otherwise, the system might have unacceptably low accuracy, and fail in its primary objective of personalized information access.

One design decision therefore is whether (and how) to track agent efficacy as part of the system history. If the agent's efficacy is generally low, then opportunities to which it is suited become particularly valuable. Another aspect of efficacy is that relationships among item characteristics may mean that a given agent, while targeted to a specific fairness concern, might have the effect of enhancing multiple dimensions of fairness at once. Promoting an under-served region might also promote an under-served economic sector. Thus, the empirically-observed multidimensional impact of a fairness concern will need to be tracked to represent its efficacy. Efficacy may also be a function of internal parameters of the agent itself. A separate learning mechanism could then be deployed to optimize these parameters on the basis of allocation, choice and user interaction outcomes.

6.3 Mechanism Inputs

Different SCRUF implementations may differ in what aspects of the context are known to the allocation and/or choice mechanisms. Our hope is that we can leverage social choice functions in order to limit the complexity of the information that must be shared. However, if a sophisticated and dynamic representation of agent efficacy is required, it may be necessary to implement a bandit-type mechanism to explore the space of allocation probabilities and/or agent parameters as discussed above. Recent research on multidimensional bandit learning suggests possible approaches [26].

6.4 Agent Priority

As we have shown, agent priority in the allocation phase may be a function of user interests, considering different users as different opportunities to pursue fairness goals. It may also be a function of the history of prior allocations, or the state of the fairness concerns relative to some fairness metric we are trying to optimize. As the efficacy consideration would indicate, merely tracking allocation frequency is probably insufficient and it is necessary to tie agent priority to the state of fairness. Allocation priority is also tied to efficacy as noted above. It may be necessary to compute expected fairness impact across all dimensions in order to optimize the allocation.

We plan to leverage aspects of social choice theory to help ameliorate some of these issues. There is a significant body of research on allocation mechanisms that provide desirable normative properties including envy-freeness [12], e.g., the guarantee that one agent will not desire another's allocation (although see discussion above regarding envy-freeness) and Pareto optimality [6]. An important and exciting direction for research is understanding what allocation properties can be guaranteed for the SCRUF-D architecture overall depending on the allocation mechanism selected [7].

We note that in most practical settings the personalization goal of the system will be most important and therefore the preference of this agent will have topmost priority. It is always allocated and is not part of the allocation mechanism. Thus, we cannot assume that the preference lists of the agents that are input to the choice system are anonymous, a common assumption in the social choice literature on voting [7].

6.5 Bossiness

Depending on how the concept of agent / user compatibility is implemented, it may provide benefits to *bossy* users, those with very narrow majoritarian interests that do not allow for the support of the system's fairness concerns. Those users get results that are maximally personalized and do not share in any of the potential accuracy losses associated with satisfying the system's fairness objectives. Other, more tolerant users, bear these costs. A system may wish to ensure that all users contribute, at some minimal level, to the fairness goals. In social choice theory, a mechanism is said to be *non-bossy* if an agent cannot change their preferences so that they get the same allocation but some other agent gets a different (possibly worse) allocation [32]. Hence, as a design goal we want to think of mechanisms for both allocation and choice that are non-bossy over the dynamic settings we consider.

6.6 Fairness Types

We concentrate in this paper and our work with Kiva generally on provider-side group fairness, that is characteristics of loans where protected groups can be distinguished. However, it is also possible to use the framework for other fairness requirements. On the provider side, an individual fairness concern is one that tracks individual item exposure as opposed to the group as a whole. It would have a more complex means of assessing preference over items and of assessing fairness state, but still fits within the framework.

Consumer-side fairness can also be implemented through use of the compatibility function associated with each agent. For example, the example of assigning risk appropriately based on user risk tolerance becomes a matter of having a risk reduction agent that reports higher compatibility for users with lower risk tolerance.

7 CONCLUSION AND FUTURE WORK

We have introduced the SCRUF-D architecture for modeling the fairness concerns of a wide range of recommender system stakeholders. We integrate fairness into recommendation generation leveraging social choice. The design is general and allows for many different types of fairness concerns—involving multiple fairness logics and encompassing both provider and consumer aspects of the recommendation platform. The architecture is also general in that it makes few assumptions about the nature of the allocation and choice mechanisms by which fairness is maintained, allowing for a large design space. Future work will proceed in multiple research arcs. One arc of future work is to apply the architecture in more realistic settings, particularly with Kiva. We are working with Kiva stakeholders and beginning the process of identifying fairness concerns. In the meantime, we also plan to conduct experiments with a variety of data sets, exploring a range of different fairness concern formalizations and social choice options.

We have made the mechanisms and the agents fairly simple by design. Further experimentation will show how effective this structure is for maintaining fairness over time and allowing a wide variety of fairness concerns to be expressed. However, there are some areas of exploration that we can anticipate.

A key feature of the recommendation context is that the decisions of the recommender system only influence the exposure of protected items. There is no guarantee that a given user will show any interest in an item just because it is presented. In some settings and for some fairness concerns, exposure might be enough. But in cases where utility derives from usage rather than exposure, there would be some value in having the system learn about the relationship between exposure and a given agent's fairness objective. This setting has the attributes of a multi-objective bandit learning problem [26], where the fairness concerns represent different classes of rewards.

Even when we consider exposure as our main outcome of interest, it is still the case that the allocation of different agents may result in differential improvements in fairness, as noted in Section 6.2 above. The current architecture does not make any assumptions about the distribution of user characteristics. That is, suppose fairness concern f_i is "difficult" to achieve in that users with an interest in related items appear rarely. In that case, we should probably allocate f_i whenever a compatible user arrives, regardless of the state of the fairness metrics. This example suggests that the allocation mechanism could be adapted to look forward (to the distribution of future opportunities) as well as backwards (over fairness results achieved). This would require a model of opportunities similar to [34], and others studied in computational advertising settings.

The current architecture envisions fairness primarily in the context of group fairness expressed over recommendation outcomes. We believe that the architecture will support other types of fairness with additional enhancements. For example, a representational fairness concern would be incompatible with the assumption that fairness can be aggregated over multiple recommendation lists. It would not be acceptable for a recommender system to deliver highly non-representative results at times, even if those results were balanced out in some overall average across users. Representational fairness therefore imposes a stricter constraint than those considered here, effectively requiring allocation for the associated concern, and possibly overriding personalization altogether.

The model expressed here assumes that fairness agents have preferences only over items. But it is also possible to represent agents as having preferences over recommendation lists. This would allow agents to express preferences for combinations of items, which cannot be expressed simply in terms of scores associated with items. Agents would have to become more complex in their ability to reason about and generate such preferences, and the choice mechanism would become more like a combinatorial optimization problem. It is possible that we can characterize useful subclasses of the permutation space and avoid the full complexity of arbitrary preferences over subsets.

Another interesting direction for research is more theoretical in nature. Much of the research in social choice focuses on providing guaranteed normative properties of mechanisms. However, the models used in traditional social choice theory do not take into consideration dynamics as most mechanisms are designed to work in one-off scenarios. It will be important to understand the properties of existing mechanisms for allocation and choice when deployed in these dynamic contexts and to develop new methods with good properties.

Finally, we note that there is very little user-centered research on algorithmic fairness. (See [5, 22, 24, 36, 39] for some recent work.) There are many unresolved questions including how to provide explanations and other transparency mechanisms for fairness-aware algorithms. Research around appropriate modeling of non-end-user stakeholders in personalized systems is still in its infancy.

ACKNOWLEDGMENTS

Authors Burke, Grozin, Voida and Sonboli were supported by the National Science Foundation under grant awards IIS-1911025 and IIS-2107577. Nicholas Mattei was supported by NSF Grant IIS-2107505.

REFERENCES

[1] M. Aleksandrov, H. Aziz, S. Gaspers, and T. Walsh. 2015. Online Fair Division: Analysing a Food Bank problem. In *Proc. 24th International Joint Conference on AI (IJCAI)*. IJCAI, 2540–2546.
[2] P. Awasthi and T. Sandholm. 2009. Online Stochastic Optimization in the Large: Application to Kidney Exchange.. In *Proc. 21st International Joint Conference on AI (IJCAI)*. IJCAI, 405–411.
[3] Solon Barocas and Andrew D Selbst. 2016. Big Data's Disparate Impact. *California law review* 104, 3 (2016), 671. https://doi.org/10.15779/Z38BG31
[4] Eric PS Baumer and Jed R Brubaker. 2017. Post-userism. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 6291–6303.
[5] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, Article 377, 14 pages.
[6] Anna Bogomolnaia and Hervé Moulin. 2001. A new solution to the random assignment problem. *Journal of Economic theory* 100, 2 (2001), 295–328.
[7] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice*. Cambridge University Press, Cambridge, UK.
[8] Eric Budish, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom. 2013. Designing random allocation mechanisms: Theory and applications. *American economic review* 103, 2 (2013), 585–623.
[9] Robin Burke. 2017. Multisided Fairness for Recommendation. In *Workshop on Fairness, Accountability and Transparency in Machine Learning (FATML)*. Halifax, Nova Scotia, 5 pages. https://arxiv.org/abs/1707.00093
[10] Robin Burke. 2022. Personalization, Fairness, and Post-Userism. In *Perspectives on Digital Humanism*. Springer Nature, Bern, 145.
[11] Robin Burke, Amy Voida, Nicholas Mattei, and Nasim Sonboli. 2020. Algorithmic Fairness, Institutional Logics, and Social Choice. In *Harvard CRCS Workshop on AI for Social Good at 29th International Joint Conference on Artificial Intelligence (IJCAI 2020)*. IJCAI, 5 pages.
[12] Yuga J Cohler, John K Lai, David C Parkes, and Ariel D Procaccia. 2011. Optimal envy-free cake cutting. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*. AAAI, 626–631.
[13] John P Dickerson, Karthik A Sankararaman, Aravind Srinivasan, and Pan Xu. 2018. Allocation Problems in Ride Sharing Platforms: Online Matching with Offline Reusable Resources. In *Proc. Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*. AAAI, 1007–1014.
[14] Benjamin Edelman, Michael Ostrovsky, and Michael Schwarz. 2007. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American economic review* 97, 1 (2007), 242–259.
[15] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. arXiv:2105.05779 [cs.IR]
[16] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the loop: Gender imbalance in music recommenders. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. ACM, New York, 249–254.
[17] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. What is fair? Exploring the artists' perspective on the fairness of music streaming platforms. In *IFIP Conference on Human-Computer Interaction*. Springer, 562–584.
[18] Rupert Freeman, Seyed Majid Zahedi, and Vincent Conitzer. 2017. Fair social choice in dynamic settings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. International Joint Conferences on Artificial Intelligence, Marina del Rey, CA, 4580–4587.
[19] Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2021. The (Im)possibility of fairness: different value systems require different mechanisms for fair decision making. *Commun. ACM* 64, 4 (April 2021), 136–143. https://doi.org/10.1145/3433949
[20] Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, et al. 2021. Towards long-term fairness in recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. ACM, New York, 445–453.
[21] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. arXiv:1711.05144 [cs.LG]
[22] Min Kyung Lee, Nina Grgić-Hlača, Michael Carl Tschantz, Reuben Binns, Adrian Weller, Michelle Carney, and Kori Inkpen. 2020. Human-centered approaches to fair and responsible AI. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 1–8.
[23] Min Kyung Lee, Daniel Kusbit, Anson Kahng, Ji Tae Kim, Xinran Yuan, Allissa Chan, Daniel See, Ritesh Noothigattu, Siheon Lee, and Alexandros Psomas. 2019. WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–35.
[24] Q Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: informing design practices for explainable AI user experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 1–15.
[25] N. Mattei, A. Saffidine, and T. Walsh. 2018. An Axiomatic and Empirical Analysis of Mechanisms for Online Organ Matching. In *Proceedings of the 7th International Workshop on Computational Social Choice (COMSOC)*. 24 pages.
[26] Rishabh Mehrotra, Niannan Xue, and Mounia Lalmas. 2020. Bandit based Optimization of Multiple Objectives on a Music Streaming Platform. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, 3224–3233.
[27] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling fairness and bias in dynamic learning-to-rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, 429–438.
[28] Hervé Moulin. 2004. *Fair division and collective welfare*. MIT press.
[29] Deirdre K Mulligan, Joshua A Kroll, Nitin Kohli, and Richmond Y Wong. 2019. This thing called fairness: disciplinary confusion realizing a value in technology. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–36.
[30] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How search engines reinforce racism*. NYU Press.
[31] Cathy O'Neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
[32] Szilvia Pápai. 2000. Strategyproof assignment by hierarchical exchange. *Econometrica* 68, 6 (2000), 1403–1433.
[33] Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P Gummadi, and Abhijnan Chakraborty. 2020. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. In *Proceedings of The Web Conference 2020*. ACM, New York, 1194–1204.
[34] Claudia Perlich, Brian Dalessandro, Rod Hook, Ori Stitelman, Troy Raeder, and Foster Provost. 2012. Bid optimizing and inventory scoring in targeted online advertising. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, New York, 804–812.
[35] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. ACM, New York, 59–68.
[36] Avital Shulner-Tal, Tsvi Kuflik, and Doron Kliger. 2022. Fairness, explainability and in-between: understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics and Information Technology* 24, 1 (2022), 1–13.
[37] Nasim Sonboli, Robin Burke, Nicholas Mattei, Farzad Eskandanian, and Tian Gao. 2020. "And the Winner Is...": Dynamic Lotteries for Multi-group Fairness-Aware Recommendation. arXiv:2009.02590 [cs.IR]
[38] Nasim Sonboli, Farzad Eskandanian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-Aspect Fairness through Personalized Re-Ranking. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 239–247. https://doi.org/10.1145/3340631.3394846
[39] Nasim Sonboli, Jessie J Smith, Florencia Cabral Berenfus, Robin Burke, and Casey Fiesler. 2021. Fairness and transparency in recommendation: The users' perspective. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, 274–279.
[40] Tom Sühr, Asia J Biega, Meike Zehlike, Krishna P Gummadi, and Abhijnan Chakraborty. 2019. Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, New York, 3082–3092.
[41] William Thomson. 2011. Fair allocation rules. In *Handbook of Social Choice and Welfare*. Vol. 2. Elsevier, 393–506.
[42] Jun Wang, Weinan Zhang, and Shuai Yuan. 2017. Display Advertising with Real-Time Bidding (RTB) and Behavioural Targeting. arXiv:1610.03013 [cs.GT]
[43] Shuai Yuan, Ahmad Zainal Abidin, Marc Sloan, and Jun Wang. 2012. Internet Advertising: An Interplay among Advertisers, Online Publishers, Ad Exchanges

and Web Users. arXiv:1206.1754 [cs.IR]
[44] Shuai Yuan, Jun Wang, and Xiaoxue Zhao. 2013. Real-time bidding for online advertising: measurement and analysis. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*. ACM, New York, 3.
[45] Meike Zehlike, Tom Sühr, Ricardo Baeza-Yates, Francesco Bonchi, Carlos Castillo, and Sara Hajian. 2022. Fair Top-k Ranking with multiple protected groups. *Information Processing & Management* 59, 1 (2022), 102707.
[46] Weinan Zhang, Shuai Yuan, and Jun Wang. 2014. Optimal real-time bidding for display advertising. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, New York, 1077–1086.
[47] Ziwei Zhu, Xia Hu, and James Caverlee. 2018. Fairness-aware tensor-based recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, New York, 1153–1162.