Algorithmic Fairness, Institutional Logics, and Social Choice

Robin Burke$^1$, Amy Voida$^1$, Nicholas Mattei$^2$, Nasim Sonboli$^1$, and Farzad Eskandanian$^1$

$^1$University of Colorado, Boulder
$^2$Tulane University
$^3$DePaul University

{robin.burke, amy.voida}@colorado.edu, nsmattei@tulane.edu, nasim.sonboli@colorado.edu, feskanda@depaul.edu

Abstract

Fairness, in machine learning research, is often conceived as an exercise in constrained optimization, based on a predefined fairness metric. We argue that this abstract model of algorithmic fairness is a poor match for the real world, in which applications are likely to be embedded within a larger context involving multiple classes of stakeholders as well as multiple social and technical systems. We may expect multiple, competing claims around fairness coming from various stakeholders, especially in applications oriented towards social good. We propose computational social choice as a promising framework for the integration of multiple perspectives on system outcomes in fairness-aware systems and provide an example in the application of personalized recommendation for a non-profit.

1 Introduction

A substantial body of research on fairness in machine learning, especially in classification settings, has emerged in the past ten years, including formalizing various definitions of the concept of fairness [Chouldechova, 2017; Dwork et al., 2012; Hardt et al., 2016; Narayanan, 2018] and offering algorithmic techniques to mitigate unfairness under these definitions [Kamiran et al., 2010; Pedreshi et al., 2008; Zemel et al., 2013; Zhang and Wu, 2017]. While many cases of problematic systems appear in popular literature, e.g., [O’Neil, 2016], only a small number of studies of deployed systems exist, e.g., [Chouldechova et al., 2018; Mehrotra et al., 2018; Beutel et al., 2019b]. All too often, detailed studies of the impacts of these systems are hampered by their commercial nature, although there has been some recent sharing of this kind of experience [Holstein et al., 2018; Cramer et al., 2018].

One side effect of this lack of empirical grounding in real-world, deployed systems is that researchers tend to rely on highly simplified concepts of fairness in their metrics and algorithms. Generally, a single protected attribute and a single binary distinction define the problem. There is little recognition of the interaction of multiple fairness definitions and dimensions, although recent work has noted the benefits of combining multiple fairness definitions [Beutel et al., 2019a]. Most existing research considers only a single protected class, and even in cases where multiple groups are considered, e.g., [Buolamwini and Gebru, 2018; Hebert-Johnson et al., 2018; Kearns et al., 2017; Zhu et al., 2018], fairness is conceived using the same definition for all groups. Complex, multi-vocal notions of fairness arising from multiple stakeholders do not appear, in spite of decades of research in sociology and organizational studies demonstrating the complexity of such values in practice. Hence, we believe it is necessary to accommodate different definitions of fairness from different stakeholders, all of which must be integrated in a single framework. This nuanced understanding of the value of fairness is essential for capturing the richness of this social construct.

We are keenly interested in the application of algorithmic fairness in contexts where the requirements for fairness arise from an organization’s mission, in contrast to a legalistic orientation, where the requirements for fairness take the form of legal/regulatory requirements imposed externally. A legalistic focus concentrates research effort on trying to ensure that a system will produce legally defensible results, based on particular standards. Although possibly influenced by regulatory regimes, mission-oriented fairness concerns are endogenous to an organization and thus the goal of fairness is held by internal stakeholders. We expect that such an orientation will be particularly important for non-profit organizations striving to achieve social good, but this is not the only application. Consider the work by Mehrotra et al. [2018] in which the music streaming service Spotify details its attempts to balance recommendation of highly-popular artists with the promotion of lesser-known names. There is no legal requirement for Spotify to promote less-popular artists, but to do so is beneficial to their platform ecosystem and their business model.

In accepting the complexity of fairness in mission-oriented contexts, we aim to widen the lens of algorithmic fairness towards the societal and organizational processes through which decisions are made and in which different groups contest for their own sense of what constitutes fair treatment. Social and political mechanisms decide what constitutes fairness, with results that differ widely by historical and social context. After all, student and senior citizen discounts constitute a form of ageist discrimination, but are considered acceptable (and legal) in the United States. By extending our analysis back into this contested territory, we can render algorithmic fairness research more relevant to the contexts where it is most needed.
2 Institutional Logics and Values

Scholars in organizational studies widely concur that broader belief systems shape the way that stakeholders operate and make decisions within organizations, referred to as institutional logics [Thornton et al., 2012]. Numerous studies have characterized the ways in which, in practice, multiple institutional logics are operating simultaneously within organizations—in competition and contestation with each other (e.g., [Besharov and Smith, 2014; Greenwood et al., 2010; Pache and Santos, 2013; Reay and Hinings, 2009]). In the nonprofit context, for example, organizations are commonly required to negotiate the competing and often conflicting institutional logics that are part and parcel of being a mission-driven organization, e.g., prioritizing services to clients, and institutional logics derived from their often-public sector funders, such as fiscal efficiency and accountability (e.g., [Evers, 2005; Binder, 2007; Mullins, 2006]).

Connecting this body of research to the domain of computing, Voida et al. [2014] further found that technologies also embody institutional logics in the ways they instantiate particular values. Even when organizational stakeholders agree on the importance of a particular value, such as fairness, that value may not be operationalized in the deployed technology in a way that is harmonious with stakeholders’ often-heterogeneous assumptions about and orientations towards how to put those values into practice. Such a mismatch can create significant challenges for organizations and the clients they serve. This research, then, suggests that fairness-aware machine learning systems will need to be able to harmonize across multiple, conflicting logics.

2.1 Example Application

Consider the loan recommendation situation of Kiva.org—a peer-to-peer micro-lending platform. The end users of Kiva are lenders who support entrepreneurs, typically from developing countries, by lending small amounts of money. The organization has the goal of providing equitable access to capital for all entrepreneurs who request loans, regardless of their geographic location, economic sector, gender, etc. This mission is instantiated through an online platform on which end-users search for and select entrepreneurs to fund. In this domain, a recommendation system that could promote entrepreneurship in certain underfunded areas, i.e., be more fair, would likely better serve the organization’s mission of equity.

Yet, within Kiva, there are numerous classes of stakeholders, many working from different institutional logics and different—sometimes conflicting or competing—understandings of what equity means and/or how best to enable it in this complex, real-world context. Fairness to an entrepreneur might mean that the quality of their business plan is privileged in the recommendation process. Fairness to the non-governmental organizations that help to serve as mentors and fiscal liaisons to the entrepreneurs might mean that each organizations’ entrepreneurs are funded at equivalent rates. Fairness to Kiva’s global mission might mean prioritizing funding entrepreneurs from systemically underfunded regions. And so on. In this context, then, one would need to conduct an analysis of the ways in which different stakeholders involved in the loan recommendation scenario understand fairness; and the development of any such system would need to consider how to prioritize and harmonize these multiple fairness concerns for the real-world system. We should not assume that all fairness concerns are cut from the same cloth or can be measured in the same way [Hutchinson and Mitchell, 2019].

3 Social Choice and Fairness

Fairness has been a central concern of numerous technical and philosophical disciplines for centuries. Among these is the economic discipline of social choice and computational social choice, which study of how groups make decisions when each member is endowed with their own preferences [Arrow et al., 2010; Brandt et al., 2016]. We propose to re-conceptualize algorithmic fairness in mission-oriented contexts as an application of computational social choice, specifically a version of the assignment problem, and that doing so solves a number of key problems in algorithmic fairness. In a traditional assignment setting we have a finite set of agents $N = \{1, \ldots , n\}$ and a finite set of alternatives $A = \{1, \ldots, m\}$. Each agent $i \in N$ has a preference $\succ_i$ over the alternatives. Typically these preferences are expressed as a binary relation over the set $A$. The social choice mechanism must find an assignment of a subset of items in $A$ to each agent $N$. Common additional constraints include maximum and minimum capacities for one or both sides of the market.

Fair allocation [Thomson, 2016] is a cornerstone topic in social choice and has found practical application in a number of areas including allocating courses in schools [Budish and Cantillon, 2012], papers to reviewers [Lian et al., 2018], and numerous other settings [Roth, 2015; Aziz, 2019]. Adopting a social choice perspective for both matching/allocation and voting have started to appear in the fairness literature including for finding fair matches in ride-sharing [Sühr et al., 2019], finding fair group recommendations through viewing them as elections [Chakraborty et al., 2019], and most closely to our setting, building a fair recommendation system through viewing it as matching market [Patro et al., 2020]. However, none of this research considers multiple fairness concerns on the provider side as required in the Kiva case.

3.1 Example Framework

In our algorithmic fairness scenario as shown in Figure 1, the agents are not actors in a traditional sense (like individual voters in a democracy) but rather fairness concerns that emerge from institutional logics within an organization. The alternatives over which these fairness concerns contend will differ depending on the organization, what resources it has, and how these resources may be allocated. The only capability agents need is the ability to compute their preferences over decision outcomes in the classic social choice sense. Combining those preferences is the task of a social choice function in the figure that produces the final result.

For Kiva, the fairness concerns may center around different aspects of the loans that may be recommended: the associated entrepreneurs, the sectors of the economy in which they work, and other considerations. The point of decision is the moment at which recommendations must be delivered to an
end-user. Within this decision, the system is ranking different loans to present and it is therefore natural to think of loans being ordered with respect to different preferences, including the end-user’s own personalized interests.

The formulation we propose is not limited to personalized recommendation or to micro-lending. Consider a classic fairness-aware machine learning task, the college admissions scenario studied in [Friedler et al., 2016]. The concerns for different aspects of diversity or fairness may be considered as having preferences over which set of students should receive an “admit” label in the classification operation. A fair admissions process is one in which the interplay between these different concerns resolves to a final outcome representing, as best as possible, the collective preference so expressed. Here we can leverage research from the preference handling community [Mattei, 2020] as well as work in participatory design to address issues in eliciting and formalizing complex preferences. The work in [Lee et al., 2019] is particularly inspiring as an effort that uses both participatory design and social choice to create an algorithmic system that incorporates the preferences of multiple stakeholders.

Conceptualizing multi-aspect fairness as social choice does not mean that algorithmic fairness will suddenly be solved. Many social choice problems are known to be NP-hard or to have no solutions when even reasonable constraints are imposed. For example, using social choice mechanisms like multi-winner voting to find proportional rankings is computationally hard [Skowron et al., 2017]. Researchers who adopt this perspective will still have plenty of problems to keep them occupied. However, what this move does offer is a more natural way to derive and express fairness concerns, an expansive formalism for concerns defined in different ways, and a better way to explain the operation of such fairness-aware systems using, e.g., axiomatic analysis of the choice rules.

### 3.2 Non-deterministic Ranking

As an example of the kind of solution that a social choice perspective on algorithmic fairness makes possible, we outline here an approach to implementing fairness in personalized recommender systems. Recognizing the tension between group and individual fairness [Friedler et al., 2016], we will be assuming a setting in which group fairness / non-discrimination is sought. We assume that the fairness concerns consist of groups over which recommendation results should be made fair, according to some fairness metric, and under some metric of recommendation outcomes. We also assume that the system is delivering a large number of recommendation outcomes over time. Thus, our objective is not that each individual list meets some fairness target but rather that our system can be fair in expectation over some time period.

We can think of this as a repeated choice environment, which lends itself to the use of probabilistic social choice methods, which have the advantage of greater tractability [Brandl et al., 2016]. Because all of our fairness concerns are derived through a deliberative process within one organization and should, in principle, be aspects of the organizational mission, strategic aspects of preference disclosure are less significant than in other social choice contexts [Conitzer et al., 2007]. There is a vast literature on repeated social choice and many algorithmic options, depending on specific problem characteristics. As an example, consider the simple non-deterministic mechanism in the random serial dictator model: the algorithm chooses randomly and with equal probability among the agents (in our case the fairness concerns) and the chosen agent gets to impose their preferred ranking on the outcome [Brandt et al., 2016].

Our recent work has examining this and other non-deterministic social choice mechanisms for integrating multiple fairness concerns within the micro-lending domain and others [Sonboli et al., 2020].

## 4 Conclusion

We have argued for an approach to algorithmic fairness that does not take fairness as an externality but rather assumes that fairness concerns arise from business models, organizational missions, and stakeholder diversity. It is natural, therefore, to expect multiple institutional logics to be operative and multiple fairness concerns to arise. Thus we need a flexible and general characterization of fairness objectives and a way to allow multiple such objectives to interact in deriving outcomes.

The field of computational social choice provides an avenue for re-conceptualizing algorithmic fairness in a way that foregrounds the multiple and contested definitions likely to arise in practice. Social choice and welfare economics more generally have a long history of grappling with and reasoning about problems of fairness, and with balancing the concerns and preferences of multiple groups.

We have shown that a natural formulation of algorithmic fairness is to represent fairness concerns as actors with preferences over system outcomes. As a benefit, some of the thorny issues of social choice are ameliorated in this setting since the number of such concerns will be tractably small and, as noted above, the incentives for strategic “gaming” of the algorithm minimal. We anticipate that this formulation of algorithmic fairness will offer rich opportunities for research and system development in both machine learning and social choice.

### Acknowledgments

Authors Burke and Sonboli were supported in part by the National Science Foundation under Grant No. 1911025.