

Algorithmic Fairness and its Limits in Group-Formation

Long paper

Otto Sahlgren¹[0000-0001-7789-2009] and Arto Laitinen²[0000-0002-4514-7298]

¹ Department of Philosophy at Tampere University, Tampere, Finland

² Department of Philosophy at Tampere University, Tampere, Finland
otto.sahlgren@tuni.fi

Abstract. Algorithmic group formation has become a flourishing research area in the computer sciences, and more recently in the field of data mining and fair machine learning. Application domains for algorithmic solutions to grouping span wide, from team-recommendation and formation in work settings to ability-grouping in education. Recent work has also focused on fairness in group-formation. We briefly review literature on algorithmic team-formation and consider fairness in different group-formation contexts. We articulate different dimensions and constraints that are relevant for fair group-formation and discuss the tension between utility and fairness. Many problems and limitations regarding formal definitions of fairness explicated in the fair machine learning literature apply also in the context of group-formation. We suggest some limits to the relevance of fairness in general and algorithmic fairness, in particular. We argue that algorithmic fairness is less relevant to some groups because of the way they come to existence or because fairness is not a central value for them. Other central values are subjective rights; autonomy or liberty; legitimacy and authority; solidarity; and diversity, each of which can be in tension with optimal fairness-and-utility. But within acceptable limits, we argue that fairness is indeed a valuable goal that may be in tension with maximization of the relevant types of utility.

Keywords: algorithmic group-formation, team-recommendation, machine learning, fairness, political philosophy, ethics

1 Introduction

Algorithmic group formation has become a flourishing research area in the computer sciences, and more recently in the field of data mining and fair machine learning (ML). Application domains for algorithmic solutions to grouping span wide, from team-recommendation and formation in work settings (Chen & Lin, 2004; Lappas et al., 2009; Barnabò et al., 2019) to ability-grouping in education (see Carmel & Ben-Shahar, 2017). Relevant questions in this field include: How do we form optimal groups algorithmically while adhering to a specific criterion or desiderata, such as utility or

diversity? How do we ensure that the process or its outcome is fair? What is fairness in this context? Little attention has been paid to other morally significant aspects of algorithmic group-formation, however, such as questions of rights and entitlements, autonomy and solidarity. This paper bridges this gap by drawing lessons from political philosophy that are relevant for algorithmic group-formation.

In Section Two we briefly review literature on algorithmic team-formation and consider fairness in different group-formation contexts. We articulate dimensions and constraints relevant for fair algorithmic group-formation and discuss the tension between utility and fairness. We consider different metrics for fairness and argue that many of the problems and limitations regarding the formal definitions of fairness explicated in the fair ML literature apply also in the context of group-formation.

In Section Three we discuss key concepts and ideas from political philosophy relevant for fairness in algorithmic group-formation, and also introduce novel viewpoints to algorithmic group-formation alongside considerations of fairness and utility. Recent research has drawn lessons from political philosophy concerning fairness (see Binns, 2018; Narayanan, 2018). This paper will go further and draw also on debates on group-formation (sometimes under the heading of teams, communities, even societies), which suggest some limits to the relevance of fairness in general and algorithmic fairness in particular. We argue that fairness is less relevant to many forms of groups, either because of the way the groups come to existence (making algorithmic solutions redundant) or because fairness is not the only central value for them (making fairness-optimization unjustified). Other central values are subjective rights; autonomy or liberty; legitimacy and authority; solidarity; and diversity, each of which can be in tension with optimal fairness-and-utility. But within acceptable limits, we argue that fairness (which can come in many forms) is indeed a valuable goal that may be in tension with maximization of the relevant types of utility (which can also come in many forms).

2 Fairness in algorithmic group-formation

The Team Formation Problem was first formulated in Lappas et al. (2009), who offered algorithmic solutions for determining an optimal team given a task, a set of individuals and a set of relations between them capturing their compatibility. Majumder et al. (2012) build on this approach, adding a capacity constraint which requires that no individual is overloaded by an assigned task. More recently, Bulmer et al. (2020) consider scenarios where instead of one task there are multiple tasks to be assigned to multiple teams, and full coverage of project tasks is not required (i.e., optimal coverage of project tasks is pursued).

Fairness in group-formation and team-recommendation systems has also gained increasing interest (see Anagnostopoulos et al, 2012; Machado & Stefanidis 2019). We consider different formal definitions and metric for fairness that are applicable in the context of group-formation below, alongside their limitations and blind spots. First, however, we consider some aspects and constraints of grouping.

2.1 Constraints in group-formation

Decision-making is always constrained by external affordances, circumstances and contextual factors. Group-formation does not differ in this regard, in that approaches to group-formation are defined and constrained by the nuanced goals, resources and affordances, as well as limitations of context, in addition to possible ethical and legal considerations. We will next explicate some dimensions and constraints of group-formation. They include, but are not limited to the following: (i) number of tasks/projects/activities/roles to be distributed, (ii) amount of individuals to be grouped, (iii) desired amount of groups, (iv) team-composition and role flexibility (e.g., whether an individual can have several roles, perhaps in several teams); (v) appropriate criteria of inclusion or criteria according to which membership is distributed, (vi) costs and benefits, and (vii) group-level properties, such as diversity, that ought to be exhibited by some or all groups that are formed.

First, a team formation task involves one or more (sets of) objectives, tasks, resources, activities, etc., to be assigned. Sometimes there is one task, such as a single case in a law firm to be assigned to a legal team. In other situations, there may be numerous assignments or projects, e.g., different essay topics on a philosophy course.

Second, the number of individuals to be grouped, or from which a subset of individuals is to be assigned to a single task, may vary (e.g., size of overall personnel, department sizes, course attendees).

Third, there may be a constraint as to how many groups can (or should) be formed. This will usually, but not necessarily, depend on the number of tasks or projects available as well contextual affordances. In certain contexts, there may be an optimal or desirable number of groups (e.g., number of groups in child day-care) due to resource constraints (e.g., personnel and facility affordances). This may be more or less flexible, however (e.g., there is no principled constraint on how many learners can be grouped for a certain project in a philosophy class).

Fourth, depending on the situation, individuals may be grouped into only one group at a time or they may belong to several groups simultaneously. For example, in some cases, it might be required that workers split their time between different projects, and in others, workers may be assigned to only one group/task at a time.

Fifth, the inclusion criteria for group-membership may vary from context to context. In product development tasks, a given project may require a certain heterogeneous set of attributes, such as range of different skills. Workers with rare skills or high-level expertise may be in a position of advantage in the competition for a place in the team. Sometimes criteria for inclusion may not be skill related. In education, group essay topics, for example, might be assigned to groups based not on their skills but group-members' (aggregated) preferences. The question regarding the selection of inclusion criteria is, of course, significant from the point of view of fairness, provided that certain individual attributes, such as 'race', are commonly excluded from the criteria for ethical and/or legal reasons.

Sixth, there may be varying costs and benefits associated with specific aspects of group-formation. Different kinds of errors may bring about different costs or harms in algorithmic decision-making. As such it may be appropriate to prioritize the mitigation or elimination of certain errors, such as false positives in recidivism risk assessment or false negatives in algorithmic medical diagnosis.¹ This applies also in group formation: assigning individuals into groups that are a poor fit for them or excluding them from groups that they would have been best suited for may bring about different costs. In algorithmic team-formation, coordination or communication costs are often estimated when forming collaborative groups (see, e.g., Anagnostopoulos et al, 2012; Lappas et al., 2009).

Furthermore, certain group compositions may be more optimal and/or just than others. There is a long debate in the educational sciences regarding ability-grouping, for example, which has centered around the question of whether homogeneous or heterogeneous learning groups are most beneficial for learning (e.g., Steenbergen-Hu 2016; see also Wichmann et al., 2016). Empirical questions regarding diversity in groups (what kinds of groups bring about the best utility) are accompanied by ethical ones – e.g., should groups be diverse along certain lines even if this results in costs in terms of utility or productivity?

Lastly, the grouping process, and the quality of its outcomes, may be evaluated along several lines. Ounnas et al. (2007) offer a metrics framework which includes the following sets of metrics: (i) *formation metrics* that tell us how well individual or all groups in the cohort are formed under given constraints (includes metrics for perceived satisfaction of the formation), (ii) *productivity metrics* that tell us how well the formed groups perform with respect to the task they are assigned to (i.e., outcome quality), and (iii) *goal satisfaction metrics* that indicate how well an individual goal within a given collaborative task or project, or the set of all goals, are satisfied at the group or cohort level. Okimoto et al. (2015) also present a property of singular groups, which may be desired in several contexts, called *k-robustness*: A team is *k-robust* if the team can accomplish a given task albeit a *k* number of members are removed from the team. In addition to these metrics, one can of course evaluate grouping and the outcomes of grouping, i.e., the formation, from the point of view of fairness and diversity, and other morally significant aspects which we consider below.

This brief overview of relevant considerations provides different viewpoints to fairness considerations in grouping, although it is not by any means exhaustive. Furthermore, algorithmic fairness cannot be considered meaningfully without considering a variety of sociotechnical factors, such as contexts of development and use, institutional practices, accountability and transparency, and the moral value and justification of particular technologies (see Selbst et al., 2019).

2.2 Bias in group-formation

Historical and technical biases in training data and in algorithms may lead to unfair outcomes and discrimination in algorithmic decision-making processes (Angwin et al.,

¹ For discussion on the subject see Corbett-Davies et al. (2017) and Hellman (2019).

2016; d’Alessandro et al., 2017). Furthermore, once built into the system, biases may be continuously reproduced and amplified through feedback-loops (see, e.g., Lum & Isaac, 2016). Given that algorithmic solutions to group-formation rely on data which may be susceptible to introducing bias into the model, issues with bias and unfairness prove worrisome in algorithmic group-formation as well. The formation of work teams for projects may utilize data on individuals’ work experience (e.g., past projects), assessments of work performance (e.g., employee evaluations), and social connection information (see, e.g., Liu et al., 2014). While often natural choices for such a task, these types of information are prone to contain biases due to discrimination at the workplace (see, e.g., Barocas & Selbst, 2016), either by employers or co-workers, and thus risk unfairness in team-recommendation if not appropriately scrutinized, checked for bias, and sanitized.

In addition to possible bias, the fairness of group-formation also depends on whether the used data is suitable to predict task-performance, suitability, or other relevant qualities in the first place. For example, team-formation in the context of work has been conducted on the basis of individuals’ personality profiles as measured with the Myers-Briggs Type Indicator (MBTI) (see Chen & Lin, 2004; Yilmaz et al., 2015). While inter-individual compatibility within groups plausibly affects the performance or productivity of groups, there are limits to the utility of certain types of information with respect to the objective of group-formation. There is, for example, significant controversy regarding MBTI’s reliability and validity as well as its value in the context of work-related issues, such as career planning (Pittenger, 1993).

2.3 Definitions, metrics, and limitations

A plethora of metrics and formal definitions for fairness have been introduced in the fair ML literature for measuring unwanted biases in processing and outcome distributions. Common approaches to measuring fairness in ML models include estimating outcome class equity, error rate disparities, and the illegitimate or undesired effects sensitive attributes have on the generated outputs (Verma & Rubin 2018). Most of these metrics are formulated with contexts in mind that involve allocating goods or opportunities through ranking or scoring individuals and assigning them into positive or negative outcome classes accordingly (e.g., binary classification).

In team-formation, when the goal is to form a single team from a set of individuals (e.g., one team of experts), several metrics introduced in the fair ML literature prove applicable. This is because the formation process is, in these cases, formally analogous to those described above (e.g., binary classification). When this is the case, we can treat equity in the formed group as a commodity to be (fairly) distributed between the individuals, akin to other goods or benefits. (The desirability of equity in the group may vary depending on whether being assigned to the group is a burden or a benefit.) When there are multiple groups to be formed and multiple goods or activities to be distributed to those groups (e.g., in multiclass classification or clustering tasks), there are other considerations, such as diversity or proportionality of shares both *within* and *across* groups, that matter for fairness.

Before we move on to consider distinct metrics and definitions for fairness, it should be noted that the objective (or desirable outcome) of group-formation is in many cases maximal *utility*. In work and education, utility may be understood, e.g., as teams' efficiency in solving a task, their aggregated skill-level or the coverage of a required set of skills – or, more generally, productivity or efficiency. This is the case with many approaches in the computer science literature, which do not necessarily consider fairness in team-formation or diversity within and across teams. We consider some exceptions below. In Section Three, we consider other aspects of group-formation which bear on the normative considerations regarding algorithmic grouping (and grouping in general).

Blindness to sensitive information

A first definition for fairness to be considered here – and plausibly one aligning with common intuitions about non-discrimination – is *fairness through unawareness* (see Kusner et al., 2017). Essentially, this approach requires that information about sensitive attributes is not used in algorithmic decision-making. In this view, a team-formation process or outcomes thereof are fair given that sensitive information remains unconsidered. As information about sensitive attributes may be redundantly encoded in seemingly neutral data, and as other factors may function as proxies for sensitive traits, unawareness is not often sufficient to ensure fairness, however (Dwork et al., 2012). To take an example from the context of education, data on students' language skills may correlate with immigrant status and ethnicity, and proficiency in the use of technology may correlate with socio-economic status by functioning as a proxy for access to technology (see, e.g., Carmel & Ben-Shahar, 2017). The use of such data in team-recommendations or ability-grouping may lead to discrimination or undue denials of access to educational opportunities for members of disadvantaged groups even if the grouping process is nominally blind to sensitive attributes.

The underlying intuition behind *fairness through unawareness* may be more appropriate to understand in terms of conditional independence or counterfactuals. For example, we might want that an individual's predicted outcome will not differ in a possible world where she does not belong to a sensitive group (other factors remaining identical). This notion of fairness is captured by *counterfactual fairness* (Kusner et al., 2017). Similarly, we might hold that a fair process or outcome of group-formation is one unaffected by information about sensitive attributes neither directly nor by proxy. However, it may be that some factors, which sensitive attributes *do* have an effect on (in the model), are reasonable or legitimate criteria for inclusion (see Kilbertus et al., 2017; Nabi & Shpitser 2018). While gender may affect, e.g., an individual's choice of education, there are arguably cases where information about educational background is relevant for group-formation – e.g., in forming interdisciplinary collaboration teams.

Equity and diversity

Unawareness and/or conditional independence of sensitive attributes may be laudable in many cases, but it may be inconsistent with explicitly promoting diversity; equitable

group-formation may require accounting for individuals' sensitive attributes. *Statistical parity* - a simple definition for equal equity in outcome classes - is satisfied when members of the relevant groups (typically protected groups, e.g., gender groups) are equally likely to receive a positive prediction (see Dwork et al., 2012). This approach to fair team-formation is taken in Barnabò et al. (2019) where fair single-team formation is defined through equitable outcomes; each protected group should have equal equity in the formed team. This requirement for equity could be extended to multi-team team-formation settings; across the entire set of formed groups, each group should satisfy statistical parity.

Enforcing statistical parity may result in outcomes which are “blatantly unfair” for individuals, however, as it does not guarantee that the most suitable individuals are selected (Dwork et al., 2012, p. 215). Equal equity, thus, may come with a cost for accuracy, recommendation relevance or other measure of utility. Furthermore, statistical parity with respect to unitary protected attributes does not ensure that intersectional groups enjoy comparatively equal representation in the formed group. Naïve adherence to a parity metric may lead to what Kearns and colleagues call ‘fairness gerrymandering’ (2017); the quota for some value of a protected attribute (e.g., women) may be filled, but some subset of the individuals in a given group (e.g., women of color) may still be comparatively underrepresented within that quota.

Notably, equity or diversity is a placeholder term in the formal sense; i.e., there are multiple dimensions or “currencies” of diversity in any given group. Typically, diversity is discussed mainly in terms of ethnicity or gender, perhaps due to a lack of diversity in those regards being a persistent problem in many contexts. The currency of diversity may, nonetheless, also be one of skill, vocation, etc. In other words, the notion of diversity in itself has no predetermined content but, rather, functions as a formal property of a formation. Hence, a group may be diverse along one dimension but not in another. A product development team may be diverse in terms of skill (e.g., it is interdisciplinary), but homogeneous in terms of gender (e.g., exclusively males), and vice versa. As such, the relevant question regarding fairness is whether group-formations are diverse in morally significant ways – i.e., what kinds of similarities and dissimilarities we ought to find within and across the formed groups².

Fair allocations and burdens

Fairness in group-formation may range beyond considerations of equity and diversity. For many groups, diversity is no stringent moral requirement. It may not be morally required from, e.g., interest-groups that serve and promote rights of certain demographic or vocational groups. (However, diversity along some dimensions of identities may still be morally praiseworthy in these cases.) Alongside diversity, fair distributions of tasks both within and between groups seem significant from the point of view of justice. For fair team-formation in collaborative work it might be desirable, for example, to have an individual fairness constraint requiring that each individual has a reasonable workload (irrespective of the overall workload of her team), but also a

² See Green & Hu (2018) for a critical discussion on the concept of similarity in fair ML.

formation-level constraint requiring that different teams' burdens are roughly equal (see, e.g., Anagnostopoulos et al, 2012).

Machado and Stefanidis (2019) define fairness in terms of teams' equal suitability to the projects they are assigned; project requirements to team skill-set suitability should be (roughly) equal across all formed pairs of teams and projects. This approach captures the idea that, when forming multiple teams, it seems appropriate to ensure that the formation process does not prioritize any group over another in terms of group-to-task match success. However, this notion of fairness can be satisfied even if every group is poorly suited to complete the task they are assigned to. Similarly, an ML model may be calibrated in that it will estimate the underlying probability (e.g., suitability/relevance) equally across different individuals, but still perform inadequately in terms of the overall accuracy of its predictions³. Furthermore, if different demographics differ in terms of their representation in the data, calibration is impossible to satisfy simultaneously with error rate parity (Chouldechova, 2017; Kleinberg et al., 2016). For group-formation this means that fair algorithmic assessments of individuals' suitability for different teams/tasks may conflict with equal error rates in team-recommendation. To provide a hypothetical example, if women workers' performance has been systematically downplayed in assessments by their employers (i.e., the data exhibits gender bias) this may lead to a higher amount of errors in team-recommendations for women.

Preferences

One can also approach fairness from the point of view of preference-satisfaction. Political philosophers leaning towards welfarist and utilitarianist stances, in particular, argue that the extent to which distributions of goods and opportunities are just depends on whether individuals' preferences are appropriately respected and/or satisfied (Cohen, 1989; see also Gosepath, 2011). Preference-satisfaction is a key factor in recommender systems where the utility of a model is often measured as the relevance of the recommendations it generates for users (utility).

Whether individuals' preferences are satisfied, and to what extent, is arguably significant when we consider how, e.g., product development teams or learning groups should be formed. We might deem it appropriate in an educational context that each group of students receives, from a bucket of different essay topics, a topic which they prefer the best. The extent to which individuals (and by extension, groups) are satisfied in the given group they are assigned to, or in the overall formation of groups, can be measured via self-assessment questionnaires, for example (see Ounnas et al., 2007).

Preference-satisfaction can also play a part in fairness considerations, as Dworkin (1981) argues defending a resource-focused theory of equality. Dworkin entertains a hypothetical auction in which individuals can, through an equal set of resources, acquire bundles of goods and commodities. If no individual is jealous of another's bundle (*envy-freeness*), the resulting distribution can be deemed just even if individuals' bundles are non-identical. Envy-freeness has been operationalized in the recommender

³ For discussion on calibration and fairness, see Pleiss et al. (2017).

system literature as a way of measuring the fairness of package-to-group recommendations at the level of both individuals and groups (Serbos et al., 2017). Envy-freeness approaches can be used to assess whether sets of tasks algorithmically distributed to groups are fair for individuals in those groups and across the overall formation.

A limitation of preference-based notions of fairness is that they are not sensitive to *entitlements*. Sometimes certain individuals may have priority to have their preference satisfied above someone else's. Institutional entitlements that come with, say, seniority or work-related merits, may warrant deviations from fair allocation (e.g., a priority in choosing which projects to work on). It may also be that some preferences should not be recognized as legitimate claims to membership in a group (e.g., a preference to work only with individuals of a certain gender or ethnicity).

3 Rights, autonomy, legitimacy, solidarity, and diversity

The phenomenon of group-formation has a long and rich history of debates in political philosophy. In the subsections that follow we start from these debates. We argue that algorithmic fairness is irrelevant to many forms of groups, either because of the way the groups come to existence or because fairness is not a central value for them.

To anticipate, the values that come closest to fairness and utility are solidarity and diversity. They can be seen as good-making features in group-formation, and, arguably, of both intrinsic and instrumental value: groups may perform better when high levels of solidarity and appropriate degree of heterogeneity or diversity is present (figure 1).

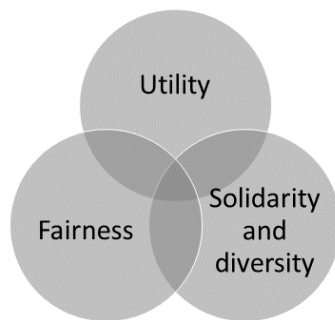


Figure 1. *Values to be jointly optimized*

But arguably, there are other considerations (from individual rights and liberties to legitimate authority) that either pre-empt or override questions of utility, fairness, solidarity and diversity, or concern the question of whose decision the group-formation is (figure 2).

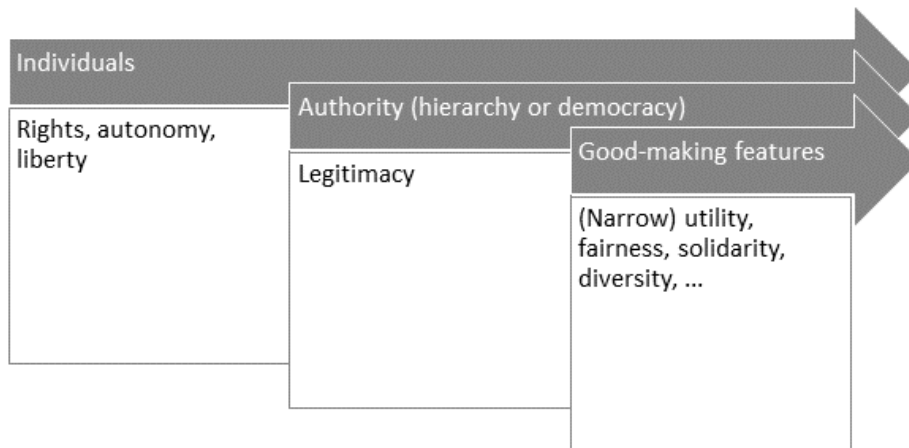


Figure 2. Normative considerations that complicate the picture

3.1 Kinds of groups

Groups come in many shapes and sizes. Some groups (call them “feature-based groups”) consist of people with a certain property - the group of all left-handed people includes people who are left-handed. Some groups (call them “emergent” groups) emerge out of interaction, say, a group of people who regularly meet for cigarettes outside a library or who share a train commute and chat at the station waiting for the train. For such groups, the problem of managing group-formation (algorithmically or otherwise) cannot directly arise (of course, managing the features or interactions will indirectly affect the constitution of these groups).

Only some groups are such that membership is independent of the features of individuals of their patterns of interaction. Some groups (call them “birthright groups”) we are born to (a family, a state in contemporary understandings; see below), and only some memberships are granted intentionally or entered into voluntarily later in life (call them “intentionally formed groups”). The issue of group-choice or managed team-formation is at home concerning such groups. Fairness, algorithmic or otherwise, can be action-guiding only in such contexts. (We can, nonetheless, in a backward-looking sense, assess the fairness or utility of membership-allocations even in contexts where they are not action-guiding: see section 3.6 below). It is however to be balanced with other considerations, such as rights, autonomy or liberty, solidarity, legitimate authority, and diversity even in those contexts. Thus, the conflict between (kinds of) utility and (kinds of) fairness does not capture the whole normative shape of matchmaking and group-formation.

3.2 The relevance of rights in group-formation

Michael Walzer (1983) in his theory of justice discusses *state membership* (citizenship) as the first good to be distributed justly, because other distributions depend heavily on this membership. Similarly, family membership (not a merely biological phenomenon, as it involves the status of a family member, as e.g. adoption or rejection suggest) is another high-stakes decision. The issue of fairness is not practically relevant for those who already are citizens of this rather than that state, or this rather than that family - questions of *birthright* silence any considerations of fairness.

For those who are stateless or children without a family, a question of *rights* is the most relevant consideration. They have the right to become a member of a state, to then have the package of rights that the state grants to citizens. Corresponding to the rights of the stateless individuals, each state may have a duty to naturalize stateless person, although there may be questions of fair sharing of bearing the burden.

Plato's *Republic* implicitly thematizes the conflict between fairness-and-utility and modern subjective rights, by neglecting the latter. Plato suggests that newborn children are collectively raised outside families and then distributed to societal tasks that are optimal to them, thus fulfilling justice (understood as everyone being in their right place). This violates such modern subjective rights as the right of parents to raise their children (and the corresponding rights of children), and the right of everyone to choose their career, and the right of everyone to freely choose their spouse. For example, Hegel in *Philosophy of Right* (1821) thinks that a rational whole (e.g. in terms of fairness and utility) should be formed so that also such modern subjective rights are respected - via the exercise of such rights. That prevents the state from being a totalitarian matchmaker allocating the right foster parents to the right children, or the right spouses to each other. That would violate subjective rights, and however optimal the allocation, it is practically irrelevant.

It is not only such high stakes decisions as career, family or state-membership that modern rights govern; there is a general freedom of association that modern constitutions declare. This takes us to the very closely related principle in group-formation: liberty or autonomy. Note how these values significantly lessen the practical relevance of optimal matchmaking in many contexts of modern life.

Many groups have, moreover, the right to choose their members. This has been discussed in the debates on group rights. The extent to which groups have a right to choose their members limits the extent to which outsiders have subjective rights to become members. It also, in a different way, may limit the relevance of utility and fairness: if a group has a right to choose its members, it has the right to do so in a non-optimal and to some extent unfair manner. And if two groups have a right to choose their members, they have a right to take into consideration the utility and fairness from their own perspective, not optimizing what is optimal for the whole system of two groups and a set of applicants. It is possible that the group's right is limited by the rights of individuals though.

It is important not to jump to the conclusion that utility or fairness would not matter: in some cases (outside birthrights) rights concern who has the right to make the

decision, whereas utility and fairness help determine what is the optimal decision that should be made.

The relevance of rights may thus be threefold: (i) birthrights may pre-empt any deliberation of how to intentionally form groups; (ii) sometimes subjective rights may override considerations of utility or fairness in group-formation; and (iii) subjective and group rights may point to a distinction between two questions: who has a right to make a decision and what decision should be made.

3.3 The relevance of autonomy and liberty in group-formation

One principle for group-formation is *voluntariness*: individuals are free to form new associations, and the membership profile of different groups is based on who freely enter or exit.⁴ Importantly, in modern contexts individuals have a freedom to choose their spouses and occupations (together with the choices of others). Similarly, within an established group, a new taskforce may be formed, and members chosen via volunteering for the task. This may lead to suboptimal formations: the volunteers may not be the most suitable, and unfairness may accumulate via voluntary choices. And this typically does not pre-empt the need to manage the team-formation: there may be a different number of volunteers than slots available. And so, fairness and utility may be action-guiding after questions of voluntariness have been taken into account. Voluntariness need not always override utility and fairness - if a leader or manager has the call, they may decide to go for the optimal solution even when volunteering has not been at stake.

So, in addition to pre-empting and overriding (in cases where rights are at stake), the considerations (fairness, utility, and voluntariness) may be in conflict as separate values to be jointly optimized. The manager may choose to prioritize voluntariness even when the individual would not strictly speaking have a corresponding right; such decisions may be motivated by the expected utility via motivational patterns etc.

3.4 Legitimacy, authority, and coercion in matchmaking

In political philosophy, the difference between legitimacy and justice (or fairness) has been fruitfully discussed in recent decades. The interim upshot is that the question of legitimate governance depends on input legitimacy (whether the governor has been chosen in a legitimate procedure, and whether the governor has legitimate authority to make decisions in some realm), and output legitimacy (whether the decisions respect individual rights, and are sufficiently good in terms of utility and fairness), and some add also throughput legitimacy (how the decision-making process proceeds). (Schmidt & Wood, 2019.)

One special case of authority is the democratic authority of the group to make binding decisions concerning itself: the decisions are illegitimate if they do not respect

⁴ Some liberties are afforded by memberships (citizenship, organizational role, marital status), some may be independently grounded human rights.

the rights of individuals. But to the extent that they are legitimate, they bind everyone - even those that voted against. (Christiano, 2008).

Another special case is a hierarchical organization such as military order of rank. If a captain has authority over a group of soldiers, then the captain's decision gives the soldiers an overriding reason to act accordingly. Even if a soldier would know a better route from A to B, it is irrelevant, if the captain has ordered to take another route. The captain's command is a stronger reason. The whole point of authority-relations is to enable such orders being overriding reasons (Raz, 1990).

There is ongoing debate about the extent to which normal workplaces should be like democracies or like military units in that respect, but what is common to both is that questions of legitimate authority are *different* from questions of fairness and utility. They concern the question of *who is to make a decision*, and not the question of which decision would be optimal. The flipside of the coin is that they do not diminish the relevance of fairness and utility in assessing which decision to make.

3.5 Solidarity, diversity, and functioning groups

The concept of solidarity has a long history in the vocabularies of political philosophy and the social sciences (see, e.g., Laitinen & Pessi, eds. 2015). In a liberal tradition individual rights, liberty and fairness (justice or equality) have been much more central. Solidarity (of fraternity) has remained a less central concept, but it is very relevant for the functioning of groups. Solidarity and the value of community are more often stressed by critics of the social contract -tradition, who often emphasize organic formations.

Solidarity typically exists when individuals have shared interests and needs, alongside possible responsibilities and inclinations to promote and partake in projects of others they identify with (Smiley, 2017, Sec. 3). For modern philosophers, such as Rousseau, the key factor driving the formation of *societies* is normatively motivated solidarity (Wrong, 1994). Of course, there is little reason to think that solidarity-based group-formation would not occur with regard to smaller communities and groups; e.g., interest-groups for people with disabilities bring together people who share an interest in supporting those with disabilities and their families, and in pursuing justice for them.

In many cases, however, group solidarity is an emergent feature of a group with a history, and thereby more like a desideratum than an active principle of membership-distribution. Of course, higher levels of solidarity can be anticipated and worked for, and especially known hostilities and obstacles for solidarity should be taken into account in group-formation.

Our suggestion is that for the purposes of this paper, solidarity is to be treated of one aspect of "utility" – one aspect of functioning teams, rather than a principle that might override or conflict with such aim. It may be in tension with other aspects of utility, nonetheless, stressing the importance of pluralism in defining utility: utility-and-fairness-and-high-level-of-solidarity can be a meaningful goal.

The concept of diversity is to some extent similar (see van Parijs, ed, 2003). If solidarity sometimes points towards homogeneous groups, diversity is an important value in itself. So, the aim of group formation should be something like utility-and-

fairness-and-high-level-of-solidarity-and-appropriate-diversity. And it should be pursued within the limits of rights, liberties, and legitimate authority.

3.6 Practically (ir)relevant assessability

In this Section, we have aimed to show that there are other values than utility narrowly conceived and fairness to be accounted for in group-formation. Solidarity and diversity are important aims in themselves, complementing the aims of fairness and narrow utility. Sometimes we can assess the optimality of matchmaking in all these respects, but it is practically irrelevant because the kind of group-formation is determined in some other ways than deliberate management (e.g., birth or spontaneous emergence).

Further normative considerations, such as rights, liberties, autonomy, legitimate authority, may concern the questions of *whose* the decision is, rather than what is the optimal decision to make. They may also be considerations that pre-empt (e.g. in the case of birthrights) or override (in case of some other rights) questions of fairness, utility, solidarity or diversity. It does not mean that the fairness or utility of the group compositions cannot be assessed in such cases. Analogously how natural beauty of sunsets or wildernesses can be assessed even though these values have not made a practical difference in their formation, we can assess fairness and optimality in groups that emerge while not being guided by such values. It would be wrong to say that utility or fairness does not matter in such cases; what can be said is that they are not practically and effectively guiding the group-formation.

4 Conclusions

In this paper we discussed algorithmic fairness in different group-formation contexts, and the limits that other normative considerations (rights, autonomy and liberty, legitimate authority, solidarity, diversity) pose. Extending the discussion of fairness to algorithmic group-formation, we argued that many technical notions of fairness introduced in the fair ML literature, such as equal equity, counterfactual fairness and envy-freeness, can be applied in the context of group-formation. The specific nature and setting of a group-formation task as well as the respective computational approach will determine the extent of their applicability. Assessment at the level individuals and comparisons both within and between groups may be relevant, and the limitations of technical fairness approaches – the informational scope of distinct metrics and trade-offs between values of interest – should be recognized and carefully considered.

Furthermore, drawing on debates in political philosophy, we argued that fairness is not action-guiding in all contexts of group-formation because questions of rights and legitimacy may override claims to fairness or maximal utility or render them less relevant. Even if not action-guiding in all cases, fairness is a value to be jointly optimized alongside others and may serve as a lens for assessing decisions regarding the formation of groups.

References

- Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., & Leonardi, S. (2012). Online team formation in social networks. *Proceedings of the 21st international conference on World Wide Web*, 839-848.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. Retrieved from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. *Proceedings of Machine Learning Research*, 81, 1–11.
- Barnabò, G., Fazzone, A., Leonardi, S., & Schwiegelshohn, S. (2019). Algorithms for Fair Team Formation in Online Labour Marketplaces. *Companion Proceedings of The 2019 World Wide Web Conference*, 484-490.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104, 671–732.
- Bulmer, J., Fritter, M., Gao, Y., & Hui, B. (2020). FASTT: Team Formation Using Fair Division. *Canadian Conference on Artificial Intelligence*, 92-104.
- Carmel, Y. H., & Ben-Shahar, T. H. (2017). Reshaping Ability Grouping Through Big Data. *Vanderbilt Journal of Entertainment & Technology Law*, 20, 87–128.
- Chen, S. J., & Lin, L. (2004). Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. *IEEE transactions on Engineering Management*, 51(2), 111-124.
- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 5(2), 153-163.
- Christiano, T. (2008) *The Constitution of Equality: Democratic Authority and its Limits*. Oxford UP.
- Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 99(4), 906-944.
- Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797-806.
- d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious classification: A data scientist's guide to discrimination-aware classification. *Big data*, 5(2), 120-134.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd innovations in theoretical computer science conference*, 214-226.
- Dworkin, R. (1981). What is equality? part 1: Equality of welfare. *Philosophy & public affairs*, 185–246.
- Gosepath, S. (2011). Equality. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), Retrieved from <https://plato.stanford.edu/archives/spr2011/entries/equality/>. [Accessed 31.5.2020]

- Green, B., & Hu, L. (2018). The myth in the methodology: Towards a recontextualization of fairness in machine learning. *Proceedings of the Machine Learning: The Debates Workshop*.
- Hegel, G.W.F. (1991 [1821]) *The Elements of Philosophy of Right*. Cambridge UP. Cambridge.
- Hellman, D. (2019). Measuring Algorithmic Fairness. Virginia Public Law and Legal Theory Research Paper No. 2019-39; Virginia Law and Economics Research Paper No. 2019-15; Virginia Law Review, Forthcoming. Retrieved from <https://ssrn.com/abstract=3418528>.
- Kearns, M., Neel, S., Roth, A., & Wu, Z. S. (2017). Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. arXiv preprint arXiv:1711.05144.
- Kilbertus, N., Rojas-Carulla, M., Parascandolo, G., Hardt, M., Janzing, D., and Schölkopf, B. (2017). Avoiding Discrimination Through Causal Reasoning. *Advances in Neural Information Processing Systems*.
- Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. arXiv preprint arXiv:1609.05807.
- Kusner, M. J., Loftus, J. R., Russell, C., and Silva, R. (2017). Counterfactual Fairness. *Advances in Neural Information Processing Systems*, 1-11.
- Laitinen, A & Pessi, A.B., eds. (2015) *Solidarity. Theory and Practice*. Lanham: Lexington Books.
- Lappas, T., Liu, K., & Terzi, E. (2009). Finding a team of experts in social networks. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 467-476.
- Liu, H., Qiao, M., Greenia, D., Akkiraju, R., Dill, S., Nakamura, T., Song, Y., & Nezhad, H. M. (2014). A machine learning approach to combining individual strength and team features for team recommendation. *2014 13th International Conference on Machine Learning and Applications*, 213-218.
- Lum, K., & Isaac, W. (2016). To predict and serve?. *Significance*, 13(5), 14-19.
- Machado, L., & Stefanidis, K. (2019). Fair team recommendations for multidisciplinary projects. *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 293-297.
- Majumder, A., Datta, S., & Naidu, K. V. M. (2012). Capacitated team formation problem on social networks. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1005-1013.
- Nabi, R., & Shpitser, I. (2018). Fair inference on outcomes. *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Narayanan, A. (2018). Translation tutorial: 21 fairness definitions and their politics. *Proceedings of the Conference on Fairness, Accountability and Transparency*. New York, USA.
- Okimoto, T., Schwind, N., Clement, M., Ribeiro, T., Inoue, K., & Marquis, P. (2015). How to form a task-oriented robust team. *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, 395-403.

- Ounnas, A., Millard, D. E., & Davis, H. C. (2007). A metrics framework for evaluating group formation. *Proceedings of the 2007 international ACM conference on Supporting group work*, 221-224.
- Pittenger, D. J. (1993). Measuring the MBTI... and coming up short. *Journal of Career Planning and Employment*, 54(1), 48-52.
- Plato (2000) *The Republic*. Cambridge: Cambridge University Press.
- Pleiss, G., Raghavan, M., Wu, F., Kleinberg, J., & Weinberger, K. Q. (2017). On fairness and calibration. *Advances in Neural Information Processing Systems*, 5680-5689.
- Rawls, J. (1971) *A Theory of Justice*. Harvard UP.
- Raz, J. (1990). *Practical Reason and Norms*. Oxford UP: Oxford.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59-68.
- Serbos, D., Qi, S., Mamoulis, N., Pitoura, E., & Tsaparas, P. (2017). Fairness in package-to-group recommendations. *Proceedings of the 26th International Conference on World Wide Web*, 371-379.
- Schmidt, V. & Wood (2019) Conceptualizing throughput legitimacy. *Public administration*, 97(4), 727-740.
- Smiley, M. (2017). Collective Responsibility. *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), Retrieved from <<https://plato.stanford.edu/archives/sum2017/entries/collective-responsibility/>>. [Accessed 31.5.2020]
- Steenbergen-Hu, S., Makel, M. C. & Olszewski-Kubilius, P. (2016). What one hundred years of research says about the effects of ability grouping and acceleration on K–12 students' academic achievement: Findings of two second-order meta-analyses. *Review of Educational Research*, 86(4), 849-899.
- van Parijs, P., et al. (eds.) (2003), *Cultural Diversity Versus Economic Solidarity: Is There a Tension? How Must it be Resolved?* De Boeck Supérieur.
- Verma, S., & Rubin, J. (2018). Fairness definitions explained. *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)*, 1-7.
- Walzer, M. (1983) *Spheres of Justice*. Basic books.
- Wichmann, A., Hecking, T., Elson, M., Christmann, N., Herrmann, T., and Hoppe, H. U. (2016). Group Formation for Small-Group Learning: Are Heterogeneous Groups More Productive? *Proceedings of the 12th International Symposium on Open Collaboration (OpenSym '16)*, 1–4.
- Wrong, D. (1994) *The Problem of Order*. Macmillan.
- Yilmaz, M., Al-Taei, A., & O'Connor, R. V. (2015). A machine-based personality oriented team recommender for software development organizations. *European Conference on Software Process Improvement*, 75-86.