



Towards Substantive Conceptions of Algorithmic Fairness: Normative Guidance from Equal Opportunity Doctrines

Falaah Arif Khan
New York University
New York, NY USA
fa2161@nyu.edu

Eleni Manis
S.T.O.P.
New York, NY USA
eleni@stopspying.org

Julia Stoyanovich
New York University
New York, NY USA
stoyanovich@nyu.edu

ABSTRACT

In this work we use Equal Opportunity (EO) doctrines from political philosophy to make explicit the normative judgements embedded in different conceptions of algorithmic fairness. We contrast formal EO approaches that narrowly focus on *fair contests* at discrete decision points, with substantive EO doctrines that look at people's *fair life chances* more holistically over the course of a lifetime. We use this taxonomy to provide a moral interpretation of the impossibility results as the incompatibility between different conceptions of a *fair contest* — forward-facing versus backward-facing — when people do not have *fair life chances*. We use this result to motivate substantive conceptions of algorithmic fairness and outline two plausible *fair decision procedures* based on the luck egalitarian doctrine of EO, and Rawls's principle of fair equality of opportunity.

ACM Reference Format:

Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2022. Towards Substantive Conceptions of Algorithmic Fairness: Normative Guidance from Equal Opportunity Doctrines. In *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, October 6–9, 2022, Arlington, VA, USA. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3551624.3555303>

1 EQUALITY OF OPPORTUNITY

Equality of Opportunity (EO) is a philosophical doctrine that objects to morally arbitrary and irrelevant factors affecting people's access to desirable positions, and the social goods attached to them (such as opportunity and wealth). In an EO-respecting society, all people, irrespective of their morally arbitrary characteristics, such as socio-economic background, gender, race, or disability status, have comparable access to the opportunities that they desire. Similarly, in fair machine learning (fair-ML), we are usually interested in ensuring that the outputs of algorithmic systems, specially those used in critical social contexts, do not systematically skew along the lines of membership in protected groups based on gender, race, or disability. In so far as protected groups are constructed on the basis of morally arbitrary factors, the moral desiderata of EO doctrines from political philosophy align exactly with the fairness-related concerns in machine learning. In this work, we employ ideas from the rich EO literature from political philosophy [2, 3, 6, 14, 15, 23, 26, 30–33]

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

EAAMO '22, October 6–9, 2022, Arlington, VA, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9477-2/22/10...\$15.00

<https://doi.org/10.1145/3551624.3555303>

to clarify the normative foundations of fairness and justice-related interventions, and gauge the efficacy of current algorithmic approaches that attempt to codify these criteria.

1.1 Principles of EO

There are two broad principles of EO, namely, *the principle of fair contests* and *the principle of fair life chances*.

1.1.1 Fair contests. The principle of fair contests, commonly understood as the *nondiscrimination principle*, says that competitions for desirable positions should be open to all and should be adjudicated based on competitors' relevant merits, or qualifications. In any fair contest, the most qualified person wins. Conversely, fair contests do not judge competitors on the basis of irrelevant characteristics, especially excluding morally arbitrary factors such as gender, race, and socio-economic status that are not properly understood as qualifications at all.

The principle of fair contests has been very influential in fair-ML and has guided statistical measures and algorithmic interventions that conceptualize *fairness* as *nondiscrimination*.

1.1.2 Fair life chances. The principle of fair life chances says that people's chances of success over a lifetime should not depend on morally arbitrary factors. It takes a holistic view of equal opportunity by comparing the *opportunity sets* that people have over the course of a lifetime, and is popularly understood as a principle that *levels the playing field*.

The principle of fair life chances has been almost entirely overlooked in fair-ML, and this omission explains some of the limitations in current approaches, as we will discuss shortly.

1.2 Domains of EO

According to Fishkin [15], there are, broadly, three domains of EO:

1.2.1 Fairness at a specific decision point. The first domain comprises the discrete points at which social goods, such as employment, admissions, and loan decisions are distributed. EO doctrines compel us to think about whether outcomes of decision-making at discrete decision points are influenced by morally arbitrary factors.

1.2.2 Equality of developmental opportunities. The second domain comprises educational and other foundational opportunities that shape people's ability to compete for desirable positions in the first domain. EO doctrines are also concerned with whether people had comparable developmental opportunities to build up their qualifications ahead of competitions.

1.2.3 Equality of opportunities over the course of a lifetime. Lastly, EO doctrines also compel us to look more broadly at the *opportunity*

sets to which people have access over the course of a lifetime, and whether these bundles are comparable.

1.3 Roadmap

Different interpretations of the two principles discussed in Section 1.1, and to which domain they apply, give rise to different conceptions of EO. We fix notation in Section 2. We then discuss Formal EO doctrines—which emphasize the principle of *fair contests* at discrete decision points—in Sections 3 and 4. We go on to provide a moral interpretation of the impossibility results in fair-ML as the incompatibility between a forward-facing vs. backward-facing conception of a fair contest in Section 5. In Section 6, we discuss Substantive EO doctrines—which emphasize the principle of *fair life chances* and target all three domains—as they are classically understood. Next, in Section 7, we provide modern re-interpretations of these doctrines that are more amenable to real-world decision-making. In Section 8, we arrange the moral desiderata of different EO doctrines into a *Fairness as Equal Opportunity* taxonomy. In Section 9, we demonstrate how our EO-based framework can provide normative guidance in practical contexts using a hypothetical example of college admissions, and the real-world case study of COMPAS. We compare our framework with contemporary work in Section 10, and then conclude with a discussion about the importance of grounding algorithmic approaches in strong normative foundations, as well as the limitations in guidance that EO doctrines can provide towards this end.

2 NOTATION

Before discussing different EO doctrines and their codification in fair-ML, let us fix some notation: Algorithmic decision-making involves predicting an outcome y' given a set of observations (X, y) , where the X s are covariates/features and the y s are the targets. In a given context, we assume that covariates X can be partitioned into “morally relevant” attributes A , and “morally arbitrary” attributes (such as gender, race, age, or disability status) S . “Fair” decision-making is concerned with satisfying some moral desiderata C , with respect to the set of attributes S , to which we commonly refer as the “sensitive attributes”.

We can apply EO doctrines to predictive problems if the target y indicates a social outcome; where people receive or don’t receive some desirable social good, the covariates A measure some notion of merit, and protected-groups are constructed on the basis of a morally arbitrary and irrelevant characteristic or characteristics $s \in S$. For example, predicting the risk of loan default can be posed as the problem of allocating a positive or negative lending decision, and the co-variables measure “financial qualifications” like income, repayment behavior, or net worth. Racial discrimination is illegal in lending, and so the morally arbitrary feature set in this example would include race.

For the rest of the discussion, we restrict ourselves to decision-making of this type, where the predicted outcome y' is a real valued score used to make a distributive/allocative decision. We stress that EO doctrines are only suitable to predictive contexts that can be posed as the problem of distribution of some social good, on the basis of some relevant qualifications/merit. We emphasize that EO doctrines

are inapplicable to algorithmic contexts that cannot be posed as distributive problems [30].

3 FORMAL EO

Formal EO doctrines specify the moral desiderata of *fair contests*: they say that no person should be excluded from a competition for a desirable position on the basis of morally arbitrary criteria. Further, in a *fair contest*, people should only be judged on the basis of their relevant merit, and so, people with comparable relevant qualifications should get the same outcome.

Formal EO, commonly known as *careers open to talents*, is only concerned with the first domain of equal opportunity—fairness at a discrete decision point [32]. Formal EO is not attentive to whether people had comparable access to developmental opportunities to build qualifications leading up to the *fair contest* (the second domain of EO), nor whether people will have comparable opportunity sets over the course of their lifetimes (the third domain of EO).

In Bernard Williams’s famous example of a warrior society, formal EO is achieved when warrior positions are open to all, and all are allowed to compete — not just the children of warrior parents [32]. However, formal EO fails to prevent privilege from being converted into qualifications in advance of the competition, whereby children from non-warrior families have no realistic chance of winning the contest without the resources and training that is afforded to children of warrior parents.

3.1 Formal EO as Fairness Through Blindness

Decision-making that is blind to irrelevant characteristics is consistent with formal EO. In fair-ML, a prominent codification of formal EO is *fairness through blindness* [13], where protected (and morally irrelevant) attributes are removed from the data, and a group-blind classifier is produced. To the extent that irrelevant characteristics (and their proxies) can be successfully excluded from an algorithm’s pipeline, formal EO can make progress toward its aim of rejecting the use of morally irrelevant features as the basis for awarding privileged outcomes. In practice, however, formal EO is often too weak to enhance fairness, as has been demonstrated in both the digital and analog age [1, 27].

Take the example of the U.S. “Ban the Box” campaign, which was aimed at passing legislation that required employers to be *blind* to candidates’ criminal histories during initial assessments of qualifications¹. The campaign was aimed at eliminating the check box on job applications that asked applicants to indicate whether they have a criminal history. Excluding criminal history from initial screenings of candidates captures formal EO’s conception of a fair contest because it attempts to ensure that justice-involved persons are judged fairly on the basis of their qualifications and not dismissed out of hand. However, this *formally fair* policy ended up having the opposite effect in practice. Field studies showed that in the absence of individual information about applicants’ criminal histories, employers end up making group-level assumptions about prior criminal justice involvement [1]. This meant that applicants with no justice involvement who belonged to groups with higher (perceived) conviction rates, such as young black males, were adversely affected,

¹<https://bantheboxcampaign.org>

while white applicants with criminal justice involvement received the benefit of the doubt.

Similar statistical discrimination due to the exclusion of group-level information is seen when formal EO is encoded into algorithmic decision-making systems. For example, Lipton et al. [27] demonstrate a “gender-blind” algorithm that discriminates on the basis of “inferred” gender at the group level when gender information at the individual level is excluded. As a result, the algorithm adversely treats applicants that it perceives as women (including men with long hair) and favors candidates that it infers to be men (including women with short hair).

The limitations of fairness through blindness are well-appreciated in fair-ML [13], so we instead turn to an alternative, stronger conception of formal EO.

3.2 Formal EO as Calibration

A well-calibrated test satisfies formal EO’s conception of a fair contest because it ensures that the likelihood of getting a positive outcome does not depend upon morally arbitrary group membership ($s \in S$):

$$P(y = 1 | y' = c, s = 0) = P(y = 1 | y' = c, s = 1).$$

Put differently, if two individuals have the same predicted score y' (relevant merit) and only differ on group membership s (morally irrelevant factors) then they are likely to get the same outcome from a well-calibrated test.

3.3 Formal EO as Predictive Parity

A test that satisfies predictive parity at threshold p is formal-EO compliant because it ensures that the likelihood of getting a positive outcome is the same for all high-performing individuals, irrespective of morally arbitrary group membership ($s \in S$):

$$P(y = 1 | y' > p, s = 0) = P(y = 1 | y' > p, s = 1).$$

Intuitively, formal EO mandates that all people who have job-relevant qualifications ($y' > p$) should have the same chance of receiving a positive outcome ($y = 1$), irrespective of their irrelevant, morally arbitrary attributes (s), and predictive value parity as a fairness criterion reflects exactly this.

4 FORMAL-PLUS EO

The strength of formal EO as a moral framework to design fair contests relies greatly on the ability to correctly measure candidates’ relevant merit. In the codifications of formal EO as calibration and predictive parity, we are making an assumption about the predicted score y' , namely, that it does, in fact, measure the applicant’s “relevant merit.” This is a strong assumption and one that does not hold in societies with historic systemic inequality. Fishkin [15] writes: “When the formal egalitarian argued that the warrior children have more merit than the non-warrior children, that view depended on a factual premise: that the warrior test did what it was designed to do and accurately predicted future warrior performance. What if it did not?”

For example, think of the SAT as a predictor of college success: When students can afford to do a lot of preparation, scores are an inflated reflection of their college potential. On the other hand, when students have limited access to preparatory material, the

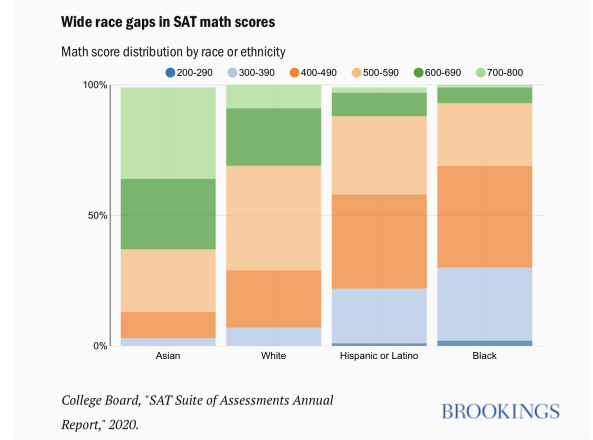


Figure 1: Distribution of SAT math scores by race or ethnicity

SAT underestimates their college potential. The SAT systematically over-predicts the future performance of more privileged students, while systematically under-predicting future performance of less privileged students: that is, the test’s validity as a predictor of college potential varies across groups. Such disparity in standardized test scores along the lines of race (and gender) has been observed in real-world contexts and is shown in Figure 1, reproduced from a Brookings 2020 report.²

In order to correct for this circular problem, where the measurement of relevant merit itself tracks morally irrelevant privilege and disadvantage, Fishkin [15] proposes a version of formal EO that he calls “formal-plus,” which adjusts test results for members of groups that are systematically underestimated by a test. From a formal-plus perspective, a *fair contest* is one in which test errors do not track (morally arbitrary) group membership.

4.1 Formal-plus EO as Error Rate Balance

A test with balanced error rates at a threshold p captures formal-plus EO’s conception of a fair contest because it ensures that test performance (i.e., false-positive rate and false-negative rate) does not skew with morally irrelevant group membership ($s \in S$):

$$P(y' > p | y = 0, s = 0) = P(y' > p | y = 0, s = 1) \text{ and}$$

$$P(y' \leq p | y = 1, s = 0) = P(y' \leq p | y = 1, s = 1)$$

4.2 Formal-plus EO as Equalized Odds

Fishkin [15] further explains that, in the absence of a “perfectly accurate test,” and with the understanding of which groups the test tends to underestimate, the formal-plus EO conception of a *fair contest* would “give compensatory bonus points” on the test to those whose future performance the test itself predictably underestimates. The idea of compensatory bonus points is simply to “make more accurate predictions about who, in the future, will actually be the best warriors.” The “equal opportunity” measure and algorithm from Hardt et al. [18] exactly captures formal-plus EO’s

²<https://www.brookings.edu/blog/up-front/2020/12/01/sat-math-scores-mirror-and-maintain-racial-inequity/>

moral desiderata, because it measures unfairness as the disparity in true-positive rates between groups, and randomly assigns positive outcomes when predicted scores fall between group-specific thresholds, to achieve the desired parity.

5 IMPOSSIBILITY RESULTS: FAIR CONTESTS WITHOUT FAIR LIFE CHANCES

Concurrent work by Chouldechova [9] and Kleinberg et al. [25] showed that it is impossible to simultaneously achieve parity in error rates and positive predictive value for different groups, if the prevalence (or base rates) differs between these groups. We will now provide a moral interpretation of these impossibility results, through the lens of the EO doctrines that we have discussed so far, as the incompatibility between two different conceptions of a *fair contest* — one that seeks to fairly reward *past* performance, versus one that seeks to fairly estimate *future* performance — when people do not have *fair life chances*. We use this result to motivate the need for substantive conceptions of algorithmic fairness, which we discuss in the rest of the paper.

Formal EO’s conception of a fair contest takes a moral *end-point* view [23] — one that rewards the qualifications that people have already developed. Formal EO codified as predictive parity mandates that whoever possesses the relevant qualification score $y' > p$ (estimated from past performance) should be given the positive outcome $y = 1$, irrespective of morally arbitrary group membership s . Kleinberg et al. [25]’s discussion of calibration as a fairness criterion is strikingly similar to the moral desiderata of formal EO. They write: “This [calibration] means we are justified in treating people with the same score comparably with respect to the outcome, rather than treating people with the same score differently based on the group they belong to.” This is exactly an end-point view of a fair contest — one that rewards people with a comparable qualification score (computed on existing merit, judged by past performance) comparably.

On the other hand, formal-plus EO’s conception of a fair contest takes a humane *starting-point* view [23] — one that aims to correctly estimate people’s likelihood of succeeding in the position on offer. Once again, Kleinberg et al. [25]’s discussion of balance for the positive and negative class as a fairness criterion is strikingly similar to Fishkin [15]’s conception of formal-plus EO: “The second [balance for the negative class] and the third [balance for the positive class] ask that if two individuals in different groups exhibit comparable *future behavior* (negative or positive), they should be treated comparably by the procedure.” This is exactly formal-plus EO’s starting-point view of a fair contest — one that treats people with comparable estimated future performance comparably.

Conversely, formal-plus EO would object to a test with unequal error rates across groups, because its estimate of people’s future performance skews along the lines of morally irrelevant privilege/disprivilege. This is mirrored in the discussion from Kleinberg et al. [25]: “In other words, a violation of, say, the second condition [balance for the negative class] would correspond to the members of the negative class in one group receiving consistently higher scores than the members of the negative class in the other group, despite the fact that the members of the negative class in the higher-scoring group have done nothing to warrant these higher scores.”

The prevalence or base rates in a particular population is the fraction of people who possess a certain quality/qualification or receive a positive outcome (based on the existence of that quality) [9, 25]. For example, in the context of hiring, the base rate in the female population is the fraction of female candidates who receive a positive outcome—a hiring offer—among all the female applicants. Fair life chances is a broad philosophical concept, but in the context of a discrete decision, we can think of base rates among populations as a proxy for their life chances. For example, if there was gender-equality in the workforce, and women had the same employment prospects as men, then we would expect equal proportions of women and men to receive a positive hiring outcome.

Putting this together, a philosophical interpretation of the impossibility results, through the lens of EO doctrines, says that it is impossible to design a *fair contest* that simultaneously rewards people’s past qualifications and accurately estimates people’s future prospects of success, if people did not have fair (comparable) life chances. This is simply because morally arbitrary and irrelevant factors weigh heavily on people’s achievement—as evidenced both in past performance and in the estimation of future performance.

The empirical results in critical domains such as criminal justice that led to the impossibility results in fair-ML [10, 25] are a stark demonstration of this fact: that people do not have comparable life chances and that morally arbitrary characteristics such as gender and race do weigh heavily on people’s prospects of success. Importantly, through the lens of EO doctrines, we can see the limitations of our current approaches in their narrow focus on designing fair contests at discrete decision points. We now discuss substantive EO doctrines, to pivot future directions of fair-ML research towards substantive conceptions of algorithmic fairness.

6 SUBSTANTIVE EO

In order to design interventions that improve people’s life chances we need substantive EO. Substantive EO focuses on giving people the opportunities to substantively build up their qualifications, so that when they do go on to compete for desirable social positions, they truly have a chance of winning. These qualifications-building opportunities fall into the second domain of EO, discussed in Section 1.2, and constitute *developmental opportunities*. The motivating reason behind both equality of developmental opportunities and EO over a lifetime is that morally arbitrary circumstances of birth, such as warrior parentage, should not determine people’s *life* prospects.

There are several conceptions of substantive EO, but in this paper we will limit our discussion to two highly influential doctrines that are relevant to fair-ML, namely, Rawls’s fair EO [30] and luck egalitarian EO [14, 31].

6.1 Limitations of Formal Doctrines

Before we move on to substantive EO, let us look at why formal EO doctrines fall short of satisfying all of our fairness-related concerns. Arbitrary and morally irrelevant privileges weigh heavily on the outcomes of formally *fair* competitions because people can leverage them to build qualifications in advance of competitions. This undermines the promise of formal doctrines—that only relevant skill will be rewarded—because it fails to stop gains from being distributed along the lines of privilege and disprivilege. We call

this the *before* problem: before competitions, people are allowed to exercise their privilege to develop relevant qualifications.

Formal doctrines also have an “after” problem. After formally *fair* competitions, winners are set up for even more success. A candidate that is hired for a job is consequently granted access to more training and job experience. This makes them even more competitive in the next competition for jobs. Conversely, those who lose at first, lose opportunities for skill development, leading to more losses. Formally fair competitions create a snowball effect, where early-on winners get further ahead while early-on losers fall further behind.

Formal EO’s “before” and “after” problems compound, with privileged candidates using their competitive advantages to win early on, thus securing more developmental opportunities, which enable further wins. Anderson [2] calls this phenomenon “discrimination laundering.” Formal doctrines cannot adequately address these problems, because they are limited to measuring people’s qualifications accurately, and to excluding irrelevant information.

To meaningfully correct for social inequalities we need substantive EO doctrines. Fishkin [15] writes: “The reason that the warrior society is interesting is that, per stipulation, it is not simply the case that children of warriors appear, through test-related artifice, most likely to be the best future warriors. The point of the example is that the children of warriors really are the most likely to grow into the best adult warriors as a result of their accumulated childhood advantages.” This is exactly the target of substantive EO doctrines: making sure that people have comparable “opportunity sets” over the course of a lifetime.

6.2 Rawls’s Fair EO

Rawls’s principle of fair equality of opportunity (FEO) says that *equally talented people should have equal prospects of success*. Rawls [30] writes: “Assuming that there is a distribution of natural assets, those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system.”

In setting out his theory of justice that embeds the FEO principle, Rawls identifies two distributive mechanisms: the “social lottery,” which distributes people their initial positions in the social system, and the “natural lottery,” which distributes people their native talent and ability. He writes [30]: “We do not deserve our place in the distribution of native endowments, any more than we deserve our initial starting place in society.” Rawls posits that the natural and social lotteries are not by themselves unjust, but it is the way that institutions have been set up that leads to inequality along the lines of morally arbitrary characteristics. His principles of justice are designed to help a society appropriately mitigate their effects.

With this in mind, Rawls’s theory of justice [30] posits the following principles that would regulate the distribution of primary social goods (including wealth and opportunity) by institutions in a just society:

- (1) Rights and liberties: Everyone has the same inalienable right to equal basic liberties.
- (2)(a) Principle of Fair EO: All offices and positions must be open to all under conditions of fair equality of opportunity.

- (b) Difference principle: Any social inequality must be applied in such a manner that they be of the greatest benefit to the least advantaged.

The principles are lexically ordered, in that people’s basic rights and liberties cannot be infringed upon while bringing about FEO, nor can the Difference Principle be applied in a way that violates FEO. People’s fundamental rights are given highest priority. Next, the principle of FEO regulates the distribution of desirable social positions such that people benefit from their arbitrary endowments from the natural lottery, but are not disadvantaged by the social lottery. Once this has been satisfied, the Difference Principle is applied, seeking to redistribute social inequality to the greatest benefit of the worst-off group, so as to limit the current and *future* effect of both lotteries.

Rawls’s principles, including his principle of fair EO, are regulatory and holistic in nature—they are to be applied iteratively to regulate the distribution of social goods by institutions. It is unclear how to port these principles, as currently understood, to discrete decision-making contexts—the kind that are the focus of fair-ML. In Section 7 we will propose a modern re-interpretation that allows for such a translation.

6.3 Luck Egalitarian EO

Luck egalitarian EO levels the playing field by making competitors’ opportunities comparable, and then allows individual choices and effort to determine the outcomes of competitions. Any resulting disparity in outcomes is morally acceptable because it is due to differential individual effort, not differential fortune.

At a discrete decision point (the first domain of EO), morally arbitrary circumstances have already weighed heavily on people’s abilities. The luck egalitarian conception of a fair contest partitions a person’s qualifications into two sets—matters of “option luck” or “choice luck” for which it is morally correct to hold the individual accountable, and effects of “brute luck” that are morally irrelevant. Luck egalitarian EO says that people’s outcomes (access to desirable positions) should only be affected by the former, and no matters of brute luck should affect the outcome of a fair contest.

The hard question now is how to make this correction. How do we separate the effects of brute luck (circumstance) from the effects of responsible choices (effort)? Roemer [31] proposed a version of EO that fulfills the moral desiderata of the luck egalitarian doctrine while bypassing the need to make an explicit separation between “responsible effort” and “arbitrary circumstance”. Instead, Roemer introduced the idea of “types”: people with the same morally arbitrary circumstance are of the same “type.” Now, for a certain matter of arbitrary circumstance (e.g., family income), the entire population can be partitioned into types (e.g., “high income,” “medium income,” and “low income”). Using this idea of circumstance-types, Roemer posits that, in comparing the effort of candidates, we should correct for the fact that those efforts are *drawn from different distributions*. In other words, effort distributions are characteristic of the type, and not of the individual, and this difference is due to a morally arbitrary factor of circumstance, for which individuals should not be held accountable. Now, in evaluating an individual’s qualifications (effort), we should only compare them to others of the same type (with the same circumstance).

For example, in Figure 1, we see a large disparity in the SAT score distributions broken down by race or ethnicity. In a luck egalitarian procedure, we would evaluate students' test scores based on where they placed within the score distribution of their particular race/ethnicity. From a moral standpoint, two individuals of different types are equally qualified for a desirable position if they lie at the same quantile of the effort distribution of their type. Hence, Roemer's conception of the a *fair contest* evaluates people by correcting for the unequal (unfair) life chances they've had in the past by ranking them in their effort-type distribution.

7 A MODERN RE-INTERPRETATION OF SUBSTANTIVE EO DOCTRINES

Substantive EO doctrines, discussed in Section 6, have much stronger moral desiderata than formal EO doctrines, discussed in Sections 3 and 4, but they also have severe shortcomings, which can preclude applying them to real-world contexts that are the focus of fair-ML. Rawls's theory has been the subject of much debate and criticism for being limited to ideal theorizing: There is very little guidance from Rawls about how to apply his principles of justice in practice, or how to bring about FEO in a world where people do not have comparable life prospects. Luck egalitarianism, specially strict interpretations of the doctrine, also has severe shortcomings such as issues of agency and autonomy in holding people responsible for certain types of luck and not for others.

In this section we provide a modern interpretation of luck egalitarian EO and Rawls's FEO, in a way that is both consistent with the original doctrines and amenable to providing normative guidance in practical contexts: We classify the luck egalitarian approach as a *backward-facing*, indirect approach to equalizing people's opportunity sets, and Rawls's as a *forward-facing*, direct approach to bringing about substantive EO.

Table 1 summarizes our classification of the normative approaches of different EO doctrines, and we elaborate on this in the remainder of the section.

Table 1: Classification of EO doctrines

	Backward-facing	Forward-facing
Fair contests	Formal	Formal-plus
Fair life chances	Luck egalitarian	Rawls

7.1 Luck Egalitarian EO as a Backward-Facing View of Fair Life Chances

The luck egalitarian view acknowledges that differences in people's qualifications at the point of competition are, at least in part, due to morally arbitrary circumstances (matters of brute luck) and so a fair competition should only evaluate candidates on the basis of their propensity to expend effort, and not on the qualifications that are built from this effort. Intuitively, the idea is that people who are disadvantaged by circumstances will have to put in far greater effort to reach the same level of ability as compared to people with advantageous circumstances, and so effort is the correct rubric of achievement, not ability.

From a practical standpoint, the luck egalitarian approach gives rise to a two-step procedure of substantive EO: first, control for people's unequal life chances, and then conduct fair contests on the basis of these adjusted qualifications. We interpret luck egalitarianism as a *backward-facing* conception: it corrects for *past* effects of brute luck, and then allows individual effort to decide future outcomes. It does improve people's life chances by distributing opportunities to which they likely would not have had access, had we not corrected for their unequal life chances in the past. However, this conception does not correct for the differential effort that will be required by people with different circumstances to excel in that position in the *future*.

7.2 Rawls's FEO as a Forward-Facing View of Fair Life Chances

Rawls's principle of fair equality of opportunity (FEO) says that *equally talented people should have equal prospects of success*. Let us unpack the principle into two components: the first part deals with identifying "equally talented" people, whereas the second part says that outcomes should be distributed in such a way that gives these "equally talented" people "equal prospects of success."

An implementation of the first part of the principle aligns with the luck egalitarian approach: we adjust our measurement of people's abilities after controlling for the effects of the social lottery in order to approximate their "native talent." The second part of the principle is where the two conceptions diverge: the luck egalitarian stops after correcting for *past* effects and simply distributes outcomes based on this corrected measurement. Rawls goes one step further, and distributes outcomes in a way that also makes people's *future* prospects of success (i.e., their prospects of succeeding in the next contest) comparable. Hence, we interpret Rawls's FEO as a *forward-facing* view of the principle of fair life chances.

We summarize our interpretation of EO doctrines in the next section, and go on to illustrate the distinction between practical applications of different EO doctrines using examples in Section 9.

8 FAIRNESS AS EQUAL OPPORTUNITY TAXONOMY

We summarize the moral desiderata and normative approaches of different EO doctrines in the *Fairness as Equal Opportunity* taxonomy given in Table 2. This gives us guidance about what value judgements our fairness interventions codify, and helps us design a suitable fairness intervention for a given context based on our normative judgements. For example, we can decide that outcomes correspond to rewards for past performance and choose a *formal* approach. Or, we can decide that the decision-making context requires selecting people who are most likely to succeed in the future and take a *formal-plus* approach instead. We can apply a two-step substantive approach following the luck egalitarian view by first adjusting the measurement of people's qualifications to correct for past effects of morally arbitrary factors, and then apply any suitable selection procedure on these adjusted qualification scores. Lastly, we can gauge that the decision-making context is a critical developmental opportunity, and choose Rawls's approach that forgoes maximum utility today in favor of improved equity tomorrow.

Table 2: Fairness as Equal Opportunity taxonomy

Doctrine	Moral desiderata	Normative approach
Formal	Fair contests should only measure morally relevant qualifications	Accurately measure past performance
Formal-plus	The performance of fair contests should not skew along the lines of morally irrelevant features	Accurately estimate future performance
Substantive: Luck egalitarian	Matters of brute luck should not affect people’s outcomes	Distribute outcomes on the basis of effort, after correcting for the past effects of morally arbitrary circumstances
Substantive: Rawls	Equally talented people should have equal prospects of success	Distribute outcomes to equalize future prospects of success of people who have the same native talent, irrespective of arbitrary circumstance

9 NORMATIVE GUIDANCE

9.1 Illustrative Example: College Admissions

We now present an example to illustrate how different EO doctrines conceptualize a *fair contest*. Suppose, for the sake of argument, that we are making a college admissions decision based on a single standardized score, shown in Figure 2. Each applicant belongs to one of two demographic groups, A and B, where group membership is based on some morally arbitrary characteristic, say race or gender. We can see that members of group A (the distribution to the left) have systematically lower scores than members of group B (the curve to the right), and we posit that this is due to the effect of morally arbitrary circumstances that differ between the two groups, and are not due to any innate difference in talents in the two groups. The question now is: How do we distribute the desirable outcome of a positive admissions decision to people from both groups, in a way that is *fair*?

Following formal EO, we would simply decide a threshold on the score, which corresponds to the level of past performance that we find deserving of a reward. This threshold is shown in the figure as “Formal threshold”. Positive outcomes are distributed to everyone who has a score (y') that is higher than this threshold. As we can see from the figure, although we did not explicitly set this threshold based on any morally irrelevant factors (we simply decided a cut-off on the standardized test score), nonetheless, this threshold is prohibitively high and our selection procedure effectively eliminates all of group A. This is a backward-facing view of a *fair contest*: one that distributes outcomes based on people’s past performance on the standardized test.

A formal-plus conception of this procedure would posit that the standardized test overestimates the abilities of group B and underestimates the abilities of group A, and so the test is not a suitable measurement of applicants’ *future* academic performance. Alternatively, it would select different thresholds for each group, shown in the figure, to correct for the test’s error. The threshold for group A does admit some people from this group, and this is exactly Fishkin’s idea of “compensatory bonus points” for the group whose abilities the test underestimates. This is a forward-facing view of *fair contests* that distributes outcomes based on accurately estimating people’s future performance.

Next, a luck egalitarian approach following Roemer would look at people’s positions within the score distributions of their type (the 80th percentile for group A and group B are shown in Figure 2). The luck egalitarian would posit that people who are at the same percentile in the score distribution of their type have expended the same degree of effort in the past, and hence should receive similar outcomes from the procedure. This is a backward-facing view of substantive EO: one that designs fair contests by correcting for unequal life chances in the past, and adjusts people’s qualification score to only reflect their morally relevant effort.

In order to differentiate Rawls’s approach to this problem from the luck egalitarian one, let us further look at two individuals, Alice (from type A) and Bob (from type B), who sit at the same percentile of the score distribution of their type, and hence are equally “talented” according to Rawlsian view. We saw that the luck egalitarian would give them both the same outcome (i.e., a positive admission decision) because these individuals have demonstrated the same degree of effort in the past. By contrast, the Rawlsian would make a further consideration: How likely are Alice and Bob to succeed in this desirable position, respectively? Even if they both receive the positive outcome, Alice will probably have to work much harder than Bob to actually do well in the program. The effects of circumstance are such that the absolute amount of effort

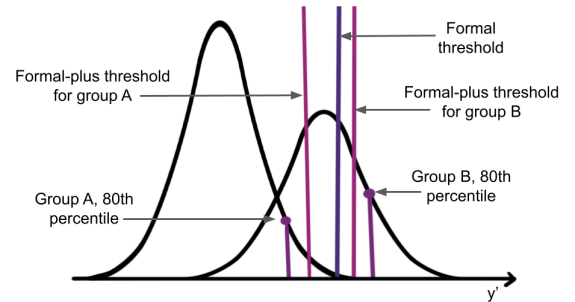


Figure 2: Distribution of test scores for groups A and B in the college admission example, discussed in Section 9.1

they have to spend to match the same level of achievement is higher for Alice than for Bob.

The luck egalitarian view does not correct for the continued future effects of circumstance, while the Rawlsian approach tries to, and makes positive admissions decisions for both Alice and Bob, but also distributes additional resources (such as tutoring and scholarships) for Alice to make up for the lack of resources and developmental opportunities they had access to, leading up to this competition. This is why we classify Rawls’s FEO as a forward-facing view of a *fair contest*: it wants to set up equally talented people to have equal prospects of winning the next contest.

Note that the actual outcomes distributed by these different procedures (motivated by different EO doctrines) may, in fact, coincide in certain contexts, based on the relative effects of morally irrelevant features. What distinguishes these procedures are their normative approaches and value judgements, and not the outcomes that are finally distributed.

9.2 Illustrative Example: COMPAS

We now demonstrate how the taxonomy of EO doctrines in Table 2 enables us to debate in normative terms, based on our value judgements. Consider the infamous example of COMPAS, as exposed in an investigation by ProPublica [4]. We would like to clarify that we chose COMPAS as the example with which to illustrate our framework not because we believe it to be representative of the kinds of contexts for which algorithms *should* be designed, and *could* benefit from ethical and moral grounding. On the contrary, we unequivocally believe that an algorithm such as COMPAS *should not* be used. Yet, without grounding different statistical measures of fairness in the moral desiderata they encode we have no way to reconcile disagreements about the implications of ProPublica’s findings. We view our EO-framework as a necessary first step towards being able to debate in values, and to audit algorithmic systems more holistically.

COMPAS is an algorithm that predicts the risk of violent recidivism among people awaiting trial. This can be posed as the problem of distributing access to resources such as counseling or other positive interventions, based on individuals’ propensity to re-offend, and translates to a problem to which we can apply EO doctrines. Northpointe argued that COMPAS did not exhibit racial discrimination because the risk scores that it produced were equally well calibrated for both black and white defendants [12]. Connecting this argument to the moral desiderata of formal EO, we see that Northpointe’s value judgement was that the algorithm “accurately” rewarded (and punished) people for their *past* actions (i.e., past criminal behavior). ProPublica, on the other hand, demonstrated that the error rates of COMPAS skewed along racial lines — the algorithm systematically overestimated the risk of black individuals and systematically underpredicted the risk of white individuals — and argued that this was evidence of racial discrimination [4]. ProPublica’s critique of COMPAS can be seen through the lens of formal-plus EO: the use of COMPAS to determine parole sentences is not formal-plus EO compliant because test performance skews along the lines of race — a morally arbitrary and irrelevant feature. ProPublica appears to be making the value judgement that COMPAS ought to accurately estimate the *future* prospects of crime of

both black and white individuals. Given that the purpose of COMPAS was, in fact, to *predict recidivism* (i.e., criminal re-offense in the *future*), viewing ProPublica’s audit through the lens of formal-plus EO makes their results even more compelling.

10 RELATED WORK

Equality of opportunity doctrines have been quite influential in fair-ML. Heidari et al. [20] were the first to formalize these ideas in fair-ML using economic models of EO. We were inspired by their work, but, in constructing this taxonomy using EO doctrines from political philosophy, find critical mistakes in their framework. Most importantly, their framework takes a reductive view of substantive EO doctrines, as independent fairness criteria at discrete decision points. This misrepresents the nature of substantive EO, which is not limited to the first domain of EO and takes a more holistic view of *fair contests* in regards to people’s *fair life chances*, as we explain in Sections 1 and 6.

With this clarification, we refute Heidari et al. [20]’s mappings of substantive EO doctrines with statistical measures, including their assertion that statistical parity, equalized odds, and accuracy map to Rawls’s FEO. We agree with their characterization of Roemer’s EO, and their application of it to predictive contexts. An important point here is that formal EO and Rawls’s FEO are doctrines from political philosophy, and not economics, while Roemer’s EO is an economic doctrine. Hence, Heidari et al. [20]’s choice to use economic models of EO might be the cause of disagreement between our framework (using EO doctrines from political philosophy) and theirs (using economic models of EO). Specifically, Heidari et al. [20] characterize methods that are consistent with luck egalitarian EO as taking a relative view of effort, and they characterize methods that are consistent with Rawls’s FEO as taking an absolute view of effort. The latter is a misconception, in that even economic interpretations of Rawls’s FEO do not characterize it as a doctrine that takes an absolute view of effort [26].

Further, following Arneson [6], libertarianism has been introduced as a possible version of EO in fair-ML [20]. However, the libertarian view focuses on a narrow notion of procedural fairness: It would object to a procedure that allows illegal or unfair means of gaining access to opportunities. While libertarianism (as a limited notion of procedural fairness) may be interpreted as a fairness-preserving position from a legal standpoint, it does not satisfy EO’s characteristic commitment to eliminating irrelevant and arbitrary barriers to achievement. The libertarian principle of self-ownership asserts that people are entitled to the full benefit of their natural personal endowments [26], and so the disparity in people’s access to desirable positions arises simply from them exercising free will. According to this view, nothing is morally arbitrary or irrelevant, and so nothing needs to be corrected for. Hence, we reject the characterization of libertarianism as a form of EO by Arneson [6], and its adoption in fair-ML by Heidari et al. [20].

There are several contemporary works that attempt to clarify the normative foundations of algorithmic fairness [5, 7, 11, 16, 17, 21]. To the best of our knowledge, we are the first to arrange fairness desiderata under the two EO principles of *fair contests* and *fair life chances*, to introduce Fishkin [15]’s doctrine of formal-plus EO as a criterion on balance of error rates between groups, and,

most importantly, to introduce a temporal dimension to the moral desiderata of EO doctrines (forward-facing vs. backward-facing).

There has also been some interest to design fairness-enhancing interventions that go beyond a single-point view of fairness [8, 19, 22, 24, 28, 29], and we hope that our framework will help ground such work in strong normative foundations.

11 DISCUSSION

It is widely accepted that *fairness* is not a statistical concept, but rather a philosophical and moral one. Yet, current approaches in fair-ML do not explicitly state the value judgements they make. In this work, we attempt to fix this deficit by making connections between influential results in algorithmic fairness, and the normative considerations of EO doctrines. We construct a taxonomy, summarized in Table 2, that introduces a temporal dimension to influential EO doctrines (forward vs backward-facing).

In making these connections between EO doctrines and algorithmic fairness approaches, we identified limitations in current approaches, specifically, a narrow focus on designing *fair contests* at discrete decision points, without broader considerations of *fair life chances* and the overall opportunity-sets available to people. In order to move beyond formal approaches, we propose modern interpretations and plausible procedures for two substantive EO doctrines in Section 7.

It is also widely accepted that *fairness* is inherently context-specific, yet, to the best of our knowledge, an understanding of the suitability of different fairness conceptions in different contexts still lacks. We take an important step in this direction: Equal opportunity doctrines, in contrast to equality of outcome doctrines, allow us to connect the nature of the opportunity with fairness desiderata. Our taxonomy makes explicit the value judgements that go into different conceptions of a *fair contest*, underscoring that different conceptions are suitable for different contexts: Do we seek to reward past performance (formal) or accurately estimate future performance (formal-plus)? Do we care about removing the effects of past inequality (luck egalitarian), or to preemptively correct for inequality that, if left unchecked, will compound in the future (Rawls's)? Thinking in these terms can guide decision-makers in selecting a suitable fairness-related intervention that aligns with their value judgements for what is suitable for their specific context.

Lastly, we would like to re-iterate that, While the EO principles are a helpful frame within which to reason about our justice-related goals, EO doctrines offer an incomplete normative palette for thinking about discrimination. An important limitation of EO doctrines is that they are only applicable to contexts where desirable outcomes are distributed on the basis of some relevant qualification.

12 CONCLUSION

In this work we showed that extant approaches to algorithmic fairness have mainly been limited to formal conceptions of *fair contests* at discrete decision points. Through the lens of EO doctrines, we provided a moral interpretation of the impossibility results as the incompatibility between two different conceptions of a *fair contest*—a forward-facing view vs. backward-facing one—when people do not have *fair life chances*. We used this result to motivate the need for substantive conceptions of algorithmic fairness, which look

more holistically at the opportunity sets that people have available to them over the course of a lifetime, and outlined two plausible procedures to do this. We hope that our work will foster similar approaches from law and social sciences in grounding current and future research in fair-ML in strong normative foundations.

ACKNOWLEDGMENTS

This research was supported in part by National Science Foundation awards No. 1934464, 1922658, and 1916505.

REFERENCES

- [1] Amanda Agan and Sonja Starr. 2018. Ban the box, criminal records, and racial discrimination: A field experiment. *The Quarterly Journal of Economics* 133, 1 (2018), 191–235.
- [2] Elizabeth Anderson. 2010. *The Imperative of Integration*. Princeton University Press.
- [3] Elizabeth S. Anderson. 1999. What Is the Point of Equality? *Ethics* 109, 2 (1999), 287–337. <https://doi.org/10.1086/233897>
- [4] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. *ProPublica* (2016).
- [5] Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2021. Translation tutorial: Fairness and Friends. In *2021 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery.
- [6] Richard J. Arneson. 2018. Four Conceptions of Equal Opportunity. *Wiley-Blackwell: Economic Journal* (2018). <https://doi.org/10.1111/econj.12531>
- [7] Reuben Binns. 2018. Fairness in Machine Learning: Lessons from Political Philosophy. In *Conference on Fairness, Accountability and Transparency*, FAT 2018, 23–24 February 2018, New York, NY, USA (Proceedings of Machine Learning Research, Vol. 81), Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 149–159. <http://proceedings.mlr.press/v81/binns18a.html>
- [8] Avrim Blum, Kevin Stangl, and Ali Vakilian. 2022. Multi Stage Screening: Enforcing Fairness and Maximizing Efficiency in a Pre-Existing Pipeline. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 1178–1193. <https://doi.org/10.1145/3531146.3533178>
- [9] Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (2017), 153–163. <https://doi.org/10.1089/big.2016.0047>
- [10] Alexandra Chouldechova and Aaron Roth. 2020. A snapshot of the frontiers of fairness in machine learning. *Commun. ACM* 63, 5 (2020), 82–89. <https://doi.org/10.1145/3376898>
- [11] Kathleen Creel and Deborah Hellman. 2021. The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision Making Systems. In *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency*, Virtual Event / Toronto, Canada, March 3–10, 2021, Madeleine Clare Elish, William Isaac, and Richard S. Zemel (Eds.). ACM, 816. <https://doi.org/10.1145/3442188.3445942>
- [12] William Dieterich, Christina Mendoza, and Tim Brennan. 2016. COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity. *Northpointe Inc. Research Department* (2016).
- [13] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. 2012. Fairness through awareness. In *Innovations in Theoretical Computer Science 2012*, Cambridge, MA, USA, January 8–10, 2012, Shafi Goldwasser (Ed.). ACM, 214–226. <https://doi.org/10.1145/2090236.2090255>
- [14] Ronald Dworkin. 1981. What is Equality? Part 1: Equality of Welfare. *Philosophy and Public Affairs* 10, 3 (1981), 185–246. <http://www.jstor.org/stable/2264894>
- [15] Joseph Fishkin. 2014. *Bottlenecks: A New Theory of Equal Opportunity*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199812141.001.0001>
- [16] Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2016. On the (im)possibility of fairness. *CoRR abs/1609.07236* (2016). [arXiv:1609.07236](http://arxiv.org/abs/1609.07236)
- [17] Ben Green. 2021. Escaping the “Impossibility of Fairness”: From Formal to Substantive Algorithmic Fairness. *arXiv preprint arXiv:2107.04642* (2021).
- [18] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain*, Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.). 3315–3323. <https://proceedings.neurips.cc/paper/2016/hash/9d2682367c3935defcb1f9e247a97c0d-Abstract.html>
- [19] Hoda Heidari and Jon Kleinberg. 2021. Allocating Opportunities in a Dynamic Model of Intergenerational Mobility. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) (FAccT

- '21). Association for Computing Machinery, New York, NY, USA, 15–25. <https://doi.org/10.1145/3442188.3445867>
- [20] Hoda Heidari, Michele Loi, Krishna P. Gummadi, and Andreas Krause. 2019. A Moral Framework for Understanding Fair ML through Economic Models of Equality of Opportunity. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, FAT* 2019, Atlanta, GA, USA, January 29-31, 2019*. ACM, 181–190. <https://doi.org/10.1145/3287560.3287584>
- [21] Corinna Hertweck, Christoph Heitz, and Michele Loi. 2021. On the Moral Justification of Statistical Parity. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) (*FAccT '21*). Association for Computing Machinery, New York, NY, USA, 747–757. <https://doi.org/10.1145/3442188.3445936>
- [22] Corinna Hertweck and Tim Rüz. 2022. Gradual (In)Compatibility of Fairness Criteria. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 11926–11934. <https://ojs.aaai.org/index.php/AAAI/article/view/21450>
- [23] Christopher Jencks. 1988. Whom Must We Treat Equally for Educational Opportunity to be Equal? *Ethics* 98, 3 (1988), 518–533. <http://www.jstor.org/stable/2380965>
- [24] Sampath Kannan, Aaron Roth, and Juba Ziani. 2019. Downstream Effects of Affirmative Action. In *2019 ACM Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) (*FAT* '19*). Association for Computing Machinery, New York, NY, USA, 240–248. <https://doi.org/10.1145/3287560.3287578>
- [25] Jon M. Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores. In *8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA (LIPIcs, Vol. 67)*, Christos H. Papadimitriou (Ed.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 43:1–43:23. <https://doi.org/10.4230/LIPIcs.ITCS.2017.43>
- [26] Arnaud Lefranc, Nicolas Pistolesi, and Alain Trannoy. 2009. Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France. *Journal of Public Economics* 93, 11 (2009), 1189–1207. <https://doi.org/10.1016/j.jpubeco.2009.07.008>
- [27] Zachary C. Lipton, Julian J. McAuley, and Alexandra Chouldechova. 2018. Does mitigating ML's impact disparity require treatment disparity?. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (Eds.), 8136–8146. <https://proceedings.neurips.cc/paper/2018/hash/8e0384779e58ce2af40eb365b318cc32-Abstract.html>
- [28] David Liu, Zohair Shafi, William Fleisher, Tina Eliassi-Rad, and Scott Alfeld. 2021. RAWLSNET: Altering Bayesian Networks to Encode Rawlsian Fair Equality of Opportunity. In *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021*, Marion Fourcade, Benjamin Kuipers, Seth Lazar, and Deirdre K. Mulligan (Eds.). ACM, 745–755. <https://doi.org/10.1145/3461702.3462618>
- [29] Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. 2018. Delayed Impact of Fair Machine Learning. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, July 10-15, 2018 (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer G. Dy and Andreas Krause (Eds.). PMLR, 3156–3164. <http://proceedings.mlr.press/v80/liu18c.html>
- [30] John Rawls. 1971. *A Theory of Justice*. Harvard University Press. <http://www.jstor.org/stable/j.ctvjf9z6v>
- [31] John E. Roemer. 2002. Equality of opportunity: A progress report. *Social Choice and Welfare* 19, 2 (2002), 455–471. <http://www.jstor.org/stable/41106460>
- [32] Bernard Williams. 1973. *The idea of equality*. Cambridge University Press, 230–249. <https://doi.org/10.1017/CBO9780511621253.016>
- [33] H. Peyton Young. 1994. *Equity: In Theory and Practice*. Princeton University Press. <https://books.google.co.in/books?id=XVK5AAAAIAAJ>