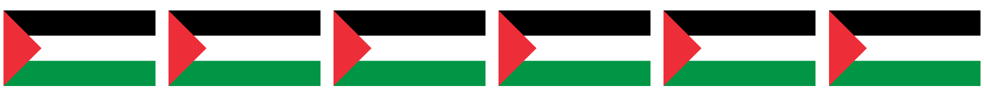
*

Algorithmic Fairness: Encouraging Exclusionary Diversity

(instead of Inclusionary Pluriversality)



K. S. Caldera
Supervisors: Dr. Sarah Carter, Dr. Stefan Buijsman, Dr. Jie Yang
EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: K. S. Caldera
Final project course: CSE3000 Research Project
Thesis committee: Dr. Sarah Cater, Dr. Stefan Buijsman, Dr. Jie Yang, Dr. Marcus Specht

*Christie, 2023.

Abstract

AI is becoming significantly more impactful in society, especially with regard to decision-making. Algorithmic fairness is the field wherein the fairness of an AI algorithm is defined, subsequently evaluated, and ideally improved. This paper uses a fairness decision tree to critique certain notions of algorithmic fairness through a postcolonial lens by applying Gayatri Spivak's theory of the subaltern alongside other postcolonial principles. A definition and criteria for a subaltern population in AI are provided, depicting that AI and algorithmic fairness rely on subaltern marginalization, silence, and faux inclusion. A theoretical case analysis is then conducted to illustrate how demographic parity, even in cases where it is the best fairness metric, does not include the subaltern. Algorithmic fairness often defines fairness through neoliberalism, assigning a "cost" to ethical considerations, wherein morality is second to profits and utility. Furthermore, a large proportion of the "justice" conducted through AI is surface level and may actually cause more harm in the long run. A proposal is made to seriously consider not using any AI in socially relevant, complex situations.

1 Introduction

As artificial intelligence is becoming increasingly important in making socially relevant decisions, the question of whether or not these machines are making fair decisions is brought up. To address this, the contemporary field of algorithmic fairness has been posited as the answer to AI's fairness issue. Within the field of algorithmic fairness, there are many definitions and metrics of fairness, as well as debate as to when certain ones are applicable in specific scenarios (Wang et al., 2022). Fairness metrics in machine learning have been rooted in neoliberal definitions of fair, defined as practically reducing the barrier of entry for those marginalized to combat systemic injustices while still trying to be accurate on the basis of "merit" (Buijsman, 2023). However, this has the adverse effect of promoting a universal, hegemonic[1], cis heterosexual, White truth while simultaneously disregarding the subaltern in a contemporary cultural imperialism (Hampton, 2021). Among these metrics is demographic parity, which is the most simple metric and is the neoliberal definition of equality. Although demographic parity appears to offer a simple solution to a complex problem, we argue it is not nearly enough to justify AI being used in a social context.

This essay will critically analyze algorithmic fairness, and more specifically, demographic parity in socially relevant situations, through a postcolonial lens. Using Spivak's "Can the Subaltern Speak?" as the basis for a framework to define a subaltern[2] population, aided by notions of pluversality to critique the hegemonic discourse, we will dissect demographic parity and highlight its extreme shortcomings and imposed ignorant liberties, which end up perpetuating inequalities instead of facilitating a "fair" machine. Pluriversality as a concept puts forth the idea that there is no one system of truth and promotes perspectives from various cultures and systems of beliefs. It rejects the Western notion of a universalizing truth (or a consensus), which in many cases does so by having a dominant group exert dominating influence over others (Vasconcelos & Martin, 2019). Postcolonialism should be of particular interest to the field of algorithmic fairness, as decision-making machines are known to perpetuate colonial inequalities (Hampton, 2021), of which postcolonialism can help us understand where and how these inequalities arise.

[1] Culturally dominant with regards to institutional power, typically comes with an imposition of its dominance on others.

[2] Marginalized position that faces epistemic oppression from the hegemonic culture.

We begin by diving briefly into details regarding demographic parity in Section 2. Following is an introduction of the subaltern as a concept, an elaboration on its manifestation in algorithmic fairness, and a criteria-based framework for highlighting *subalternity* in algorithmic fairness in Section 3. Section 4 applies the framework to theoretical aspects of situations of demographic parity to highlight the effects on the population. Section 5 is a discussion, followed by Section 6, which tentatively proposes next steps and a conclusion.

2 Demographic parity as a fairness metric

Traditional algorithmic fairness has many metrics to ensure fairness with varying prerequisites and requirements for satisfaction (Jui & Rivas, 2024). Demographic parity may be the most simple to understand, as its goal is to ensure that "the favorable outcome should be assigned to each subgroup of a sensitive class at equal rates" (Ruf & Detyniecki, 2021, p. 11), essentially ensuring the prediction is independent of any sensitive attribute (Wang et al., 2022). Demographic parity is particularly interesting to algorithmic fairness, as it equates to the lack of disparity. From a simple mathematical perspective, demographic parity is as equal as possible. To give a more tangible, toy-example, demographic parity between two nationalities, Dutch and English, would be satisfied in a situation where 100 Dutch and 50 English applicants applied for a position. Subsequently, 50 Dutch people were accepted, and 25 English people were admitted, regardless of other conditions.

When aiming for an outcome that coincides with distributive justice principles while keeping the idea of a meritocracy, conditional demographic parity has been used instead (Buijsman, 2023; Zou & Khern-am-nuai, 2023). Conditional demographic parity moves further away from a group notion of fairness to an individual idea, as it focuses less on ensuring an entire demographic's fairness and instead on trying to guarantee that a smaller, typically more privileged subset of the marginalized group is treated fairly, which allows focus to be placed on an individual's qualifications moreso than strict demographic parity. In other words, individual fairness tends to allow for more discrimination (or difference-making) within groups (Castelnovo et al., 2022; Verma & Rubin, 2018).

2.1 When is demographic parity generally acceptable as a fairness metric?

One of the main issues in algorithmic fairness is defining fairness; although there exist many mathematical notions of fairness, its universal definition is still contested. Many agree there is no silver bullet, so research is dedicated to figuring out the most appropriate fairness definition given a situation (Wang et al., 2022; Buijsman, 2023). A commonly referred to framework is Ruf and Detyniecki's heuristic decision tree: the fairness compass. The purpose of this fairness compass is to select the most appropriate algorithmic fairness metric based on simple questions. This fairness compass will be used as a basis for determining cases when Demographic Parity is the most applicable traditional fairness metric (Ruf & Detyniecki, 2021).

Based on the fairness compass, we can define four main cases where demographic parity is the most acceptable metric:

1. There is a policy in place (e.g. affirmative action) **AND** results proportional to the size of each group are desirable (e.g. given 10 women and 2 men apply to a job listing if 1 man was hired then 5 women should be hired.)

2. There is **NOT A** policy in place **AND** there are equal base rates in the measurements (e.g. the actual chances of men and women to complete secondary education **IS** the same **AND** is measured to be the same)

3. There is **NOT A** policy in place **AND** there are **NOT** equal base rates, however, there should be **AND** there is no variable to explain why

4. There is **NOT A** policy in place **AND** there are **NOT** equal base rates **AND** there is no ground truth **AND** there is no variable to explain why

In addition to Ruf and Detyniecki's technically driven fairness compass, Buijsman's philosophical addition of Rawlsian distributive justice serves as a starting point for the incorporation of more substantive fairness questions into and around the decision tree. Although Buijsman notes that "philosophical understandings of fairness... tend to be disconnected from the more technical approaches" (Buijsman, 2023, p. 2), they are important to make the decision of fairness more socially relevant, also stating "the ultimate goal is substantive equality, not formal parity" (Buijsman, 2023, p. 11). As such, Buijsman's additions include a generalization to the use of demographic parity as the primary fairness metric given a certain context, claiming it should be so whenever the context is relevant to positions in society, say a job position or even recidivism. Much like Buijsman, there are other generalizations or interpretations of Ruf and Detyniecki's fairness compass, of which this essay will be one.

3 Defining subaltern and hegemonic in the context of algorithmic fairness

Gayatri Spivak's seminal essay, "Can the Subaltern Speak?" is a significant piece in the contemporary study of postcolonialism. In the essay, she discusses the subaltern and their inability to successfully express themselves in hegemonic discourse. The subaltern, much like the hegemon, is not an identity but a state of being (similar to the colonized and colonizer). Algorithmic fairness, being dominated by the global north, leads to an excessive dependence on hegemonic knowledge. This exclusive reliance on hegemonic knowledge as a basis for discourse is in itself exclusionary and conducive to there being a subaltern group (Vasconcelos & Martin, 2019). Subaltern populations face epistemic oppression and are unable to represent themselves fairly and accurately in any mainstream arena (McEwan, 2018; Spivak, 1988; Tuck & Yang, 2014). Safiya Noble's notion of Algorithmic Oppression, extended by Hampton (2021), is a more accurate and holistic description of what is commonly referred to as algorithmic bias, wherein decision-making machines inherently cater to maintaining society's injustices, with their existence and valorization being inherently harmful to marginalized people, as the machine-made decisions are taken as the objective truth, and people are not given a real platform to express themselves (Tuck & Yang, 2014). Using Spivak's concept of the subaltern, alongside an analysis of the recipients of Hampton's extension of algorithmic oppression, aided by pluriversality, we will define a subaltern group in algorithmic fairness and develop criteria to determine who, if anyone, constitutes the subaltern in a given algorithmic fairness context.

3.1 Marginalized voices fall on deaf hegemonic ears

In defining a more democratic form of discourse, Glover (2012) critiques the hegemonic form of participatory democracy. He depicts this as a more practical account of groups being excluded, wherein their participation in a "democratic" discussion ends up silencing non-hegemonically aligned perspectives. In our context, the hegemonic form of discourse, neoliberalism, asserts that perspectives that do not have hegemonically defined societal or economic gain as the focus or justification as illegitimate (Vasconcelos & Martin, 2019). The concrete manifestation of this exclusionary, participatory democracy is made apparent when looking at fairness-based discourse. Buijsman's account of applying Rawlsian distributive justice to algorithmic fairness is an excellent example of exclusionary, participatory discourse. Buijsman applies philosophical principles to the application of fairness metrics, however, we see that their implications take the form of various neoliberal axiomatic assumptions (Mitchell et al., 2021; Buijsman, 2023). Conflating demographic parity and justice in socially relevant applications fails to consider any other perspective that is not that of neoliberal aims. This idea only allows for "participation" by the marginalized population's account of oppression. This "participation" takes the form of a lack of access, and is exclusively applied through data (Hampton, 2021; Tuck & Yang, 2014). This disregard for any alternative perspective is made obvious in the conclusion of Buijsman's (2023) article, where he concludes that the focal point of AI should be reducing the "resulting difference in absolute welfare" (Buijsman, 2023, p. 11). Equalizing absolute welfare is not enough for social (epistemic) justice, and the idea suffers from the same hegemonic form of participatory democracy described by Glover (2012). Buijsman is not alone in this shortsightedness; much of the algorithmic fairness field relies on this universal, exclusionary, and shortsighted definition of fairness and continues to build upon it (Green & Hu, 2018).

Tuck and Yang's Axiom 1 of Refusing Research 2014 depicts that any group that is subaltern **is able to express *only* their pain.** Pain narratives do not equate to knowledge; instead, they are only a form of faux-inclusion. The group can, therefore, not offer any knowledge deemed acceptable in mainstream discourse (i.e., their knowledge is only accepted if it fits the hegemonic model of knowledge). The influence of the hegemon is significantly pronounced when it comes to algorithmic fairness, as metrics are only able to recognize "well-being" as defined by the hegemonic, neoliberal notion, and the "inclusion" of the subaltern can only be seen when their pain (i.e., lack of "well-being") is used as justification for more extreme measures to "include" them (Tuck & Yang, 2014; Mitchell et al., 2021; Vasconcelos & Martin, 2019). This problematic need for algorithmic fairness to "include" the oppressed is described plainly by bell hooks: "I can talk about you better than you can speak about yourself. No need to hear your voice... I am still colonizer the speaker subject and you are now at the center of my talk" (bell hooks, 1990, p. 152). This representation is not enough correct to allow any non-hegemonic knowledge to be used and is thus both a White-echo chamber and exclusionary, regardless of what the process or result is. This form of "inclusion" paints a picture of cultural imperialism and a White-savior complex from the algorithmic fairness field.

The physical manifestation of Tuck and Yang's Axiom 1 in Refusing Research in algorithmic fairness happens in many ways, with one of the more plain ones being through the "inclusion" of marginalized people through data collection. This representation through data allows the hegemon to claim "inclusivity" and deflect claims of discrimination. However, data does not provide a voice; instead, it is only a channel through which they can express their pain. This pitfall is used to further the colonial agenda and to silence any knowledge that is not neoliberal in nature (Hampton, 2021; Tuck & Yang, 2014; Jui &

Rivas, 2024). An example of this, again, can be seen in Buijsman's addition of Rawlsian distributive justice, where he recognizes that there is indeed a more holistic notion of fairness, however, ultimately concludes the paper with a reductive, mathematical interpretation of fairness, where a parity-based measure is recommended for substantial justice. Oppressed knowledges and voices are constantly disregarded, and met with a neoliberal response that treats them as mere objects in data collection.

3.2 Criteria for identifying a subaltern in algorithmic fairness

The subaltern are the group who are not only marginalized but actively oppressed by their hegemon. Furthermore, the subaltern do not have a voice; meaning their perspectives and knowledges are excluded from discourse unless deemed acceptable by the hegemon. From this, we can develop three tentative concrete criteria that determine whether a group can be considered subaltern in the arena of algorithmic fairness:

1. **Marginalized or "Other" Group**
2. **Lack of Voice**
3. **Non-Hegemonic Perspective**

Marginalized or "Other" group Marginalization is an important aspect of being subaltern because marginalization opens the subaltern up to saviorism enacted by the hegemon. White saviorism entrenches entire marginalized groups within societies and is justified by the same people perpetrating it; the hegemon gets to act as judge, jury, and executioner. The hegemon is able to disguise their downstream influence on marginalized populations as "neutral research." This research forgoes the acknowledgment of alternative perspectives. Instead, it only uses its subjects as objects for pain: a pain in which the same group discussing it feels the compulsion to fix, while the discussion never adequately reaches those actually experiencing pain (Tuck & Yang, 2014). As reported by CNBC, a Google contractor paid homeless people in a majority Black city for their headshots to be part of a facial recognition dataset (Elias, 2019). This form of objectification of the marginalized is degrading and highlights how far removed the field of fairness is from their "other" subjects: so much so that the exploitation of Black people's bodies is justified to further "inclusive" data collection.

The neoliberal, technosolutionist[3] nature of algorithmic fairness sees this societal pain and decides that this pain can 1. be solved by **only** rational, mathematical approaches and 2. can **only** be solved by the "civilized" hegemon and their rational, mathematical approaches (Mitchell et al., 2021) (Green & Hu, 2018). The physical manifestation of this closed-feedback loop of action leads to a cis-gendered, heterosexual, able-bodied, White-dominated space of algorithmic fairness, where the technical elite impose solutions they deem fair on the rest of society as if it was their imperative to do so (Hampton, 2021).

Lack of voice Central to Spivak's idea of the subaltern is that subaltern groups are marginalized, but not all marginalized groups are subaltern; the difference lies in whether the groups can speak. In a 1992 interview, Spivak noted that being oppressed does not necessarily equate to being subaltern and that the overuse of the word is problematic and

[3] An ideology wherein non-technical problems are solved with technical solutions.

dangerous (de Kock, 1992). Accordingly, it is paramount to the meaning and power of the idea to ensure that any one or group referred to as "subaltern" **cannot** participate in the hegemonic discourse. Spivak herself notes some examples, such as "discriminated-against minorit(ies) on the university campus" and "the working class" as being certainly oppressed, however not deserving the distinction of "subaltern". *Subalternity* denotes those who are unable to represent themselves when it counts i.e. in their relevant, mainstream arena. The aforementioned oppressed examples are able to speak and represent themselves in the language of institutional power and have, at least somewhat, conformed to the standards of the hegemon (i.e., been victims of cultural imperialism). The marginalized university student and the working class citizen both by being part of and contributing to a hegemonic institution allows them to be heard, or at the very least, the ability to speak. Although oppression is extremely nuanced, we can see clear differences when observing their struggles: the non-subaltern, oppressed are looking for a greater piece of the pie, whereas the subaltern are in another room, making the pie (de Kock, 1992).

Non-hegemonic perspective The principles of pluriversality argue that multiple truths can exist simultaneously and multiple forms of existences can be true (Vasconcelos & Martin, 2019). Having a perspective that is different from that of the hegemon is an important aspect of being subaltern in algorithmic fairness. Diversity is posited in many contexts, especially that of algorithmic fairness, as having different attributes to the same, universally applicable categories. In order for this notion of "diversity" to hold, a reductionist view of humanity, where membership to a group is boiled down to certain criteria, is required. In algorithmic fairness, the criteria for groups are made by one dominant group, the technocratic[4] elites. This directly goes against the pluriversal truth that is the condition of human uniqueness, in fact, in many arenas, it is possible to gerrymander one's way around the definition of groups, and therefore definition of **inclusion** of groups, allowing practically untrue or useless "diversity" to take place, essentially ethics washing through diversity (Todd, 2011; Hampton, 2021). This exact idea appears in many places, including in "Can the Subaltern Speak?", where Spivak depicts three dominant groups and one subaltern class, wherein only the higher class of elites are "foreign" (or hegemonic) elites. The three lowest classes are all considered "indigenous" when the hegemon deems that fact to be useful. However, a functional difference is made, again, when the hegemon decides it to be useful. A modern-day example is that of the diversity hire. The diversity hire is deemed Black or queer or female and paraded as such for statistics and optics, but expected to mask their unique traits to serve their purpose in the workplace, silencing all that makes them Black or queer or female. This makes the diversity hire both hypervisible, in that their difference is paraded and brought to the forefront, and invisible, in that their difference is reduced to Whiteness when the hegemon just needs another employee (Reddy, 1998).

The current manifestation of the ideal of "diversity" in algorithmic fairness leaves those not at the table at an arguably even worse off position, as their "inclusion" brings with it perspective baggage, where the hegemon uses the "inclusion" of a member of a cultural group as the institution's key to fixing the problem of fairness and this person, under a universalized perspective, becomes a tokenism of inclusion, without leading to any real perspective shift (Hampton, 2021). Non-hegemonic groups are only allowed within institutions because they can be useful to the hegemon; the relationship is comparable to the indigenous groups in "Can the Subaltern Speak?", where they are a sort of translation layer between the out-of-touch elite and everyone else. Where the Indigenous elite would provide useful work

[4] Governed by technical specialists.

upstream to the colonizing class, the diversity hire sends subservient silence through their representation down to their supposed group. The hegemon can get away with this because their perspectives or immediate interests align, which is to maintain the status quo for both, as they are both benefiting from the institution. Originating from a subaltern group, made to conform to the standards of the hegemon, and therefore sharing a perspective, makes one inherently unfit to represent the subaltern (let alone be called subaltern) even from which they may be normatively deemed to have membership in (Piu, 2023).

4 Examining the subaltern despite demographic parity

For this section, we will analyze the four cases defined in Section 2.1, where Ruf and Detyniecki's fairness compass deems demographic parity to be the best metric. For each case, we will assume there is a subaltern population prior to the application of demographic parity. We will use the three aforementioned subaltern in algorithmic fairness criteria to determine whether or not the satisfaction of demographic parity exacerbates or alleviates specific aspects of *subalternity*.

4.1 Case one

There is a policy in place ***AND*** *results proportional to the size of each group is desirable*

This first case shows the scenario where there is a brute-force policy in place to mandate the numerical representation of groups to be equal, which will undoubtedly increase diversity of some sort from a numerical perspective.

Marginalized status It is difficult to generalize whether or not a group's marginalized status will be affected by policy, as it is a deeper social issue of which its impacts can be mixed. On the surface, individual level, easier access to jobs or university could lead to easier social mobility.

Voice Bringing subaltern populations to the current hegemonically-dominated table of algorithmic fairness is meaningless for the voice of the subaltern. However, it can be argued that increasing the presence of a population will increase the relevance of their knowledge. Representation in itself is not enough to allow the subaltern to have a voice. Instead, the framing of the discourse must change. The implication that demographic parity provides a relevant voice is harmful.

Perspective Representatives for subaltern populations will be used as examples (i.e., tokens of subaltern participation in hegemonic institutions) (Hampton, 2021), and as such, the individuals present and benefiting from the institutions in place (e.g. applications for tech jobs), will no longer share a perspective with the subaltern. This makes them no longer subaltern. This is problematic as it is a form of new-age cultural imperialism. Furthermore, the brute-force, individual-level inclusion is directly harmful to the greater subaltern population, as the benefiting, subaltern-adjacent group is used as a token: used to show that the subaltern's perspective is (falsely) included in the current discourse. They are not subaltern. They are victims of cultural imperialism who were previously subaltern.

4.2 Case two

There is ***NOT A*** *policy to encourage a goal in place* ***AND*** *there are equal base rates*

This second case displays the example where the measured and actual positive rates, prior to the application of the algorithm in question, between groups are the same and should be maintained after the application of the algorithm.

Marginalized status The precondition of equal base cases implies that groups are on a similar footing in this **specific**, measurable aspect. Although marginalized groups are certainly still marginalized, the fact of equal base rates implies that they do not suffer in this particular aspect. From a surface level utilitarian perspective, this does not appear harmful, however, we cannot conclude that it helps. In fact, Mitchtell et. al (2021) describes a pitfall with mathematical fairness models: "while mathematical definitions in the algorithmic fairness literature discussed below may be able to certify the fair allocation of decisions across a population, they have nothing to say about whether any of the available actions are acceptable in the first place." Fundamentally, the issue with using demographic parity in this case, is that only the positive result rates are considered, essentially being agnostic (or ignorant) to the complex social reality that comes before and after the decision-making machine. The contexts in which these decision-making machines are used do not offer such luxuries, and instead, such careless assumptions can lead to further marginalization. For example, in the UK, in 2022, men outnumbered women 4.4 to 1 in Computer Science applications and 4.3 to 1 in first-year Computer Science students, which implies equal base rates in the application procedure, assuming there is not already a policy in place (bcs.org, 2022; Erika, 2024). Placing a demographic parity compliant, decision-making algorithm in this spot ignores very real systemic issues that make the women who do apply just as strong applicants as the men and allows the cycle of inequality to continue. Essentially, the "fair" machine is deluding us into believing there is equality.

Voice Given the idea of this case is that, since the rates are the same, they are allowed to remain the same, the problem of representation remains: the hegemon will continue to dominate discourse as mathematical notions of fairness are not analogues to real world, social dynamics, of which in this case is the ability to participate in a version of non-exclusionary discourse (Green & Hu, 2018).

Perspective Having truly equal base rates and maintaining them, with respect to the criteria perspective in itself is not a significant concern, however, this does not mean that the subaltern are not still vulnerable to downstream social effects from the hegemon. In fact, being in such a hegemonically dominated environment, such as the EU's solution for intercultural education (which essentially posits European standards as the universal goal), is prone to leading to forceful conformation to the hegemonic perspective (Todd, 2011). Equal base cases, i.e., numerical notions of fairness, in this case, cannot represent the issue, let alone reduce these effects (Green & Hu, 2018; Mitchell et al., 2021).

4.3 Case Three

There is ***NOT A*** *policy in place* ***AND*** *there are* ***NOT*** *equal base rates, however, there should be* ***AND*** *there is no variable to explain why*

This third case displays a situation where there exists a disparity between groups in the measurements, and the numbers themselves cannot explain it, thus a systemic issue is assumed.

Marginalized status If used to gauge fairness in a socially relevant situation, the satisfaction of demographic parity in any machine is not enough to meaningfully reduce marginalization of a group. This case of the application of demographic parity attempts to correct the wrong-doings of history and society by simply evening numbers, even though it has accepted that there is a deeper-rooted systemic issue. For those particular members of a subaltern group, this may increase access to resources and capital, however, mathematical parity still does not equate to a social ideal and cannot fix the issues that lead to these inequalities on a group level (Green & Hu, 2018).

Voice The rationale for voice is extremely similar to Case One (Section 4.1), where the voice of the subaltern population in itself is not directly hindered but not helped either.

Perspective Algorithmic fairness is incomplete and can therefore not explain many inequalities. The data in this case deems that there is a measured disparity a priori. This leads to a neoliberal default of attempting to promote "prosperity," which is fundamentally flawed and problematic. In fact, this case, in particular, admits there is a larger systemic issue that plagues the application and, therefore, the machine. Demographic parity attempts to solve systemic issues by totally ignoring the nuances and social dynamics in which the machine will be used. In other words, this case exists because scholars recognize that there is deep-rooted inequality and attempt to fix it by ensuring that numbers match up.

Western philosophies and perspectives create the idea of "self", as well as the binaries of "us and them", where the "civilized" hegemon is able to portray it**self** as a capable savior (McEwan, 2018, p.183). In algorithmic fairness, this binary narrative pushes the idea that being more like the hegemon (i.e., cultural imperialism) is "prosperity." The modern idea that (western) development is the solution for "saving" the subaltern in itself is rooted in racist ideals, flawed logic, and quite literally false narratives. Attempting to simply equalize numbers as a method of fairness is akin to promoting social "elevation" through developing (i.e., "civilizing") the subaltern. This is conducive to a "new imperialism," where the hegemon spreads their influence to marginalized populations under the guise of prosperity (McEwan, 2018, p.184). These ideals manifest themselves in cases such as this: where there is no rationale other than a lack of explanation and an ignorance of anything outside the hegemonic, universalizing truth; thus, the de facto standard is to enforce universalizing Whiteness, packaged as some buzzword such as "prosperity", "justice", or "access", upon subaltern populations.

4.4 Case Four

*There is **NOT A** policy in place **AND** there are **NOT** equal base rates **AND** there is no ground truth **AND** there is no variable to explain why*

For the purposes and intents of this essay, this case suffers identical pitfalls to case three. Although this case does not explicitly state that base rates should be the same, given that this essay only concerns socially relevant algorithmic fairness, base rates should be as close to even as possible to promote equal opportunity. Additionally, a lack of a ground truth does not change much, as the correctness of an algorithm is not relevant in this discussion.

5 Discussion and future work

This analysis of demographic parity has highlighted many fundamental flaws in the field of algorithmic fairness. We will briefly discuss them to supplement the findings of the case analysis and stimulate future analyses. This is by no means an exhaustive discussion.

5.1 Superficial approach to fairness

Algorithmic fairness's approach to fairness is simply that of numbers. Buijsman (2023) introduces varying perspectives to aid algorithmic fairness, but for many (including Buijsman), the basis is still superficial; it relies on some sort of savior complex by the technocracy that is the field of algorithmic fairness. This manifests itself most often through the (re)distribution of hegemonically defined social and economic status. This idea, rooted in neoliberalism, ignores significant amounts of social context, especially that of epistemic oppression, and reduces justice to possessions and status. Reparations do not make the result inclusive, and are not even a "step in the right direction." Currently, they serve to silence marginalized people, while painting them as helpless (Tuck & Yang, 2014).

Much of the discourse in algorithmic fairness revolves around trying to maximize some superficial numerical aspect for some underprivileged group, and nothing more. Little consideration is taken for the consequences. Engineers and "ethicists" make and facilitate decision-making machines and allow them to be deployed in society, while assuming false social-analogous mathematics translate into real world results (Hampton, 2021). The assumptions simply do not translate (Green & Hu, 2018).

5.2 Price on lives?

A common theme in the current fairness discourse is its affinity to frame the issue of fairness as a "cost", or some sort of numerical criteria that can simply be filled (or often minimized) (Kozodoi et al., 2022). This framing turns it into a balancing act, essentially putting a price on fairness and, therefore, a price on equality, and finally, a price on lives. The field cheapens fairness by boiling it down to numbers, which often are connected to a neoliberal definition of prosperity (e.g., money from a mortgage or job access), of which we then (falsely) equate to well-being. It is evident that the hegemonic culture is not concerned with what is truly fair, but rather, what is fair enough to silence people and justify the use of AI that cheapens the decision-making process[5].

Hypervisibility The issue of hypervisibility begins prior to algorithmic fairness, and is exacerbated many fold by its rudimentary assumptions. The framing of the discourse is presented with a default, where the hegemony is "defined as absence, as purity" (Reddy, 1998, p. 61). This places a burden on anything that is not the hegemony. Anything queer, anything Black, and anything not White is treated as the "other"; it becomes something to neutralize, to make pure, to **civilize.** With this, we highlight how the current discourse

[5]Vanhaute, 2018, Discusses the idea of cheapening things for capitalism's sake.

related to traditional algorithmic fairness is focused on how to "include" those "others", which leads to extreme scrutiny of the "other", as they are seen as abnormal (Hampton, 2021; Reddy, 1998). These current fairness metrics attempt to extract a "normal" or "objective" aspect, which requires special attention to be placed on what it means to be non-hegemonic, which the field justifies using numbers. Whiteness, straightness, affluence, and manhood (not masculinity), are the the norm (i.e. the hegemony). These aspects are what it means to be "normal" in the context of algorithmic fairness, and special care and observation is only taken when looking at not "normal".

The analysis on the not normal, i.e."other", through the hegemonic algorithmic fairness discourse as a sort of alien being, leads to a representation through the eyes of "normal", and the differences the hegemon chooses too see (Tuck & Yang, 2014). They are not viewed objectively, but put side by side with "normal" people, and their differences highlighted. Think of the term "demographic **parity**": parity to whom? Parity, not to each other, but all to a White, "normal". Everything is framed with respect to the White, where the differences, i.e., non-Whiteness, are put on blast, and the judgment occurs on parts that do not fit the mold.

Another common theme in this essay is that of the idea of voice. We'd like to stress how important it is for a group to have a real voice: not to naively include the population, but to respect their knowledge and its production. We stress this point to show how the field essentially says "I know you what's best for you" as if "non-[W]hiteness" were some disease that leads to a lack of knowledge. The entire field mimics society and perpetuates its oppressive notions: "because [W]hite heterosexuality is treated as the norm, identical to humanity" (Reddy, 1998, p. 57). In this context, *subalternity* is a condition **imposed** by the exact hegemony trying to fix it. The solution begins with the know-it-all engineers and technocrats learning when to not impose and listen instead of insisting that numbers will lead us to "prosperity."

6 Conclusion

As depicted, the justification behind demographic parity is rooted in neoliberal assumptions that serve to exclude subaltern knowledges, values, and perspectives. As such, the mathematical metric is not currently not fit to deem what is "fair" to govern society or make any socially relevant decisions. This sentiment extends far beyond just demographic parity and likely even the entire field of algorithmic fairness.

We tentatively propose that a decision point is added to Ruf and Detyniecki's fairness compass prior to the use of Demographic parity where the following question is asked: "In this context of algorithmic fairness, does there exist a subaltern population?" If the answer is yes, and such a population does exist, then we suggest AI not be used in this context. AI and its technocratic considerations cannot include multiple perspectives, and the use of an acceptable demographic parity satisfying algorithm would, as we have reasoned before, pretend to include their perspectives and objectify them. This then means the solution to the problem is not technical in nature. Instead, it is a social issue, most likely deep-rooted and systemic, and as such, engineers cannot solve it with numbers; instead, sociologists and planners are better equipped to solve the problem.

We additionally propose that prior to the use of AI (i.e., the first decision point of the compass), the following question needs to be asked: "Does the situation in which the machine will be deployed face and/or include aspects (e.g., systemic issues, possible varying definitions of 'good') that numbers cannot holistically represent?" If the answer is yes, then consider not using AI. Although human decision-makers are indeed prone to the same type of errors that decision-making machines are prone to making, non-machine, alternative forms of decision-making can be more transparent and capture historical and societal nuances that machines simply cannot(Mitchell et al., 2021). Furthermore, given the current universalizing, hegemonic framing of AI, any discourse involving “inclusion” and its subsequent “optimization” is, by definition, entirely unable to include multiple co-existent truths (Glover, 2012; Vasconcelos & Martin, 2019). As such, any of these machines are entirely unfit to govern a society that includes multiple, co-existent truths.

We urge that more critiques, specifically with other parts of the compass, and additions are made to the compass, as it serves as a solid foundation for the use of socially relevant AI that does not stray too far from hegemonic discourse (otherwise it would be written off as activism).

Through a post-colonial lens, it is very difficult to create a truly pluriversal and non-exclusionary version of algorithmic fairness in this context, however, this would be valuable future work, and a solid basis in which to make algorithmic fairness truly fair. As of now, we, as a society, need to ask ourselves if we should make these oppressive, inequality-perpetuating, universalizing, hegemonic machines at all.

References

bcs.org. (2022, December 1). *Women choosing computing degrees in record numbers*. Retrieved June 11, 2024, from https://www.bcs.org/articles-opinion-and-research/women-choosing-computing-degrees-in-record-numbers/

bell hooks. (1990). *Yearning: Race, gender, and cultural politics*. Routledge.

Buijsman, S. (2023). Navigating fairness measures and trade-ofs. *Springer Link*, 11. https://doi.org/10.1007/s43681-023-00318-0

Castelnovo, A., Crupi, R., Greco, G., Regoli, D., Penco, I. G., & Cosentini, A. C. (2022). A clarification of the nuances in the fairness metrics landscape [Publisher: Nature Publishing Group]. *Scientific Reports*, *12*(1), 4209. https://doi.org/10.1038/s41598-022-07939-1

Christie, R. (2023, May 25). Zona duru mi protesta. https://www.behance.net/riangelochristie/projects

de Kock, L. (1992). Interview with gayatri chakravorty spivak: New nation writers conference in south africa [Archived 2011-07-06 at the Wayback Machine]. *ARIEL: A Review of International English Literature*, *23*(3), 29–47.

Elias, J. (2019). Google contractor reportedly tricked homeless people into face scans. *CNBC*. https://www.cnbc.com/2019/10/03/google-contractor-reportedly-tricked-homeless-people-into-face-scans.html

Erika. (2024, February 27). *Women in UK are applying for computing degrees in record numbers* [CEPIS]. Retrieved June 11, 2024, from https://cepis.org/women-are-applying-for-computing-degrees-in-record-numbers/

Glover, R. W. (2012). Games without frontiers? democratic engagement, agonistic pluralism and the question of exclusion [Publisher: SAGE Publications Ltd]. *Philosophy & Social Criticism*, *38*(1), 81–104. https://doi.org/10.1177/0191453711421605

Green, B., & Hu, L. (2018). The myth in the methodology: Towards a recontextualization of fairness in machine learning.

Hampton, L. M. (2021). Black feminist musings on algorithmic oppression. https://doi.org/10.1145/3442188.3445929

Jui, T. D., & Rivas, P. (2024). Fairness issues, current approaches, and challenges in machine learning models. *International Journal of Machine Learning and Cybernetics*. https://doi.org/10.1007/s13042-023-02083-2

Kozodoi, N., Jacob, J., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. *European Journal of Operational Research*, *297*(3), 1083–1094. https://doi.org/https://doi.org/10.1016/j.ejor.2021.06.023

McEwan, C. (2018). *Postcolonialism, decoloniality and development* (2nd ed.). Routledge. Retrieved April 26, 2024, from https://www.perlego.com/book/1574508/postcolonialism-decoloniality-and-development-pdf

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2021). Algorithmic fairness: Choices, assumptions, and definitions [Publisher: Annual Reviews]. *Annual Review of Statistics and Its Application*, *8*, 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Piu, P. (2023). The journey of subalternity in gayatri spivak's work: Its sociological relevance [Publisher: SAGE Publications Ltd]. *The Sociological Review*, *71*(6), 1258–1276. https://doi.org/10.1177/00380261231194495

Reddy, M. T. (1998). Invisibility/hypervisibility: The paradox of normative whiteness [Publisher: Penn State University Press]. *Transformations: The Journal of Inclusive

Scholarship and Pedagogy, *9*(2), 55–64. Retrieved June 7, 2024, from https://www.jstor.org/stable/43587107

Ruf, B., & Detyniecki, M. (2021, September 30). Towards the right kind of fairness in AI. https://doi.org/10.48550/arXiv.2102.08453

Spivak, G. C. (1988). Can the subaltern speak? Macmillan.

Todd, S. (2011). Educating beyond cultural diversity: Redrawing the boundaries of a democratic plurality. *Studies in Philosophy and Education*, *30*(2), 101–111. https://doi.org/10.1007/s11217-010-9215-6

Tuck, E., & Yang, K. (2014). R-words: Refusing research. https://doi.org/10.4135/9781544329611.n12

Vanhaute, E. (2018). A history of the world in seven cheap things: A guide to capitalism, nature and the future of the planet. https://doi.org/10.1080/03066150.2018.1543005

Vasconcelos, A. B., & Martin, F. (2019). Plurality, plurilogicality and pluriversality: A literature review. a review for the creativity and emergent educational-futures network (CEEN), graduate school of education, university of exeter .

Verma, S., & Rubin, J. (2018). Fairness definitions explained. *Proceedings of the International Workshop on Software Fairness*, 1–7. https://doi.org/10.1145/3194770.3194776

Wang, X., Zhang, Y., & Zhu, R. (2022). A brief review on algorithmic fairness. *Management System Engineering*, *1*(1), 7. https://doi.org/10.1007/s44176-022-00006-z

Zou, L., & Khern-am-nuai, W. (2023). AI and housing discrimination: The case of mortgage applications. *AI and Ethics*, *3*(4), 1271–1281. https://doi.org/10.1007/s43681-022-00234-9