
HAUKE BEHRENDT

University of Stuttgart, Institute of Philosophy
hauke.behrendt@philo.uni-stuttgart.de.

RESOLVING THE PUZZLE OF AFFIRMATIVE ACTION: DISCRIMINATION AND OPPRESSION AS GROUP-BASED INJUSTICES¹

abstract

I argue that a nuanced consideration of different types of social groups can provide a solution to the Puzzle of Affirmative Action. Opponents claim that affirmative action is inherently discriminatory because it results in preferential treatment for individuals based on a specific characteristic, comparatively disadvantaging those who do not share that trait. To address this critique, I will undertake a precise analysis of the concept of discrimination, emphasizing its nature as a form of group-based injustice. Building on this analysis, I will introduce a key distinction between oppressed social groups and egalitarian social groups. This distinction will clarify why affirmative action does not constitute discriminatory treatment in certain cases, particularly when it involves measures aimed at rectifying oppression. The proposed conceptual framework allows for a more nuanced distinction between morally acceptable and unacceptable equality measures.

keywords

social groups, discrimination, oppression, affirmative action, justice

¹ I am grateful to the participants of the Institute Colloquium at the Institute of Philosophy, University of Stuttgart, during the winter semester 2023/24, the 6th Workshop on Political Philosophy at the University of Düsseldorf, and the DGPhil Congress 2024, where I received invaluable feedback on earlier drafts of this work. In particular, I wish to thank David Schweikard, Fabian Wendt, Adis Selimi, Frank Dietrich, Benjamin Kieseewetter, and Christine Bratu for their helpful comments. I am also thankful for the feedback provided by an anonymous reviewer of this journal and by the editors, Federica Liveriero and Ingrid Salvatore, whose suggestions significantly improved the arguments presented here.

1. Introduction

Many contemporary societies grapple with a centuries-long legacy of systematic discrimination and oppression. Although both are now condemned in most societies, they persist in various aspects of life. Persistent forms of sexism and racism (among others) continue to shape the lives of countless individuals globally, despite the establishment of legal equality in many countries. While there is a widespread consensus on the need to counteract the ongoing marginalization of minorities, agreement on the implementation of such measures remains elusive. Consequently, contentious discourse has prevailed for decades, focusing on addressing these injustices through remedial equality measures based on protected characteristics – commonly known as *Affirmative Action (AA)*. However, critics argue that AA is inherently discriminatory because it results in preferential treatment for individuals based on a particular characteristic, disadvantaging those who do not share that trait (cf. Newton 1973; Eastland and Bennett 1979; Sowell 2004; Sterba 2004). In a recent case, *Students for Fair Admissions vs. Harvard*, the Supreme Court declared it unconstitutional to consider race in any form in university applications. As Chief Justice John Roberts emphasized, students should be treated based on their individual experiences, not their ethnic backgrounds.

This paper examines the persistent debate over whether AA constitutes a form of discrimination. I intend to argue in favor of AA as a crucial tool for combating systemic oppression, thereby underscoring its anti-discriminatory purpose. More fundamentally, I will address and refute the claim that AA is inherently discriminatory, a challenge I refer to as the *Puzzle of Affirmative Action (PAA)*. The puzzle is outlined through three key propositions:

- (1) Disadvantaging someone based on their social group membership constitutes wrongful discrimination.
- (2) Necessarily, AA involves favoring members of specific social groups, thereby disadvantaging non-members based on their group membership.
- (3) Consequently, AA is a special instance of morally wrong discrimination.

First of all, it's important to consider the notion of a puzzle in greater detail. I don't see AA as a puzzle because it is a difficult problem to solve, like many philosophical problems. Rather, it is because the goal of AA – the elimination of discrimination – seems to involve a form of discrimination itself. In order to solve the PAA as understood in this way, it is important to distinguish between two different challenges associated with it, each requiring a different solution.

(a) The first challenge to AA is what I call the *conceptual puzzle*. This challenge asserts that AA is inherently self-contradictory. AA aims to be anti-discriminatory, yet it appears to be discriminatory itself. This is a conceptual puzzle because AA, understood as “anti-discrimination”, appears to be an instance of discrimination, which would be conceptually contradictory.

(b) The second challenge to AA is what I call the *normative puzzle*. This version of the puzzle argues that AA may be normatively self-defeating in practice, suggesting that AA may undermine its intended goals by perpetuating or even exacerbating the very inequalities it seeks to address. It is a normative puzzle because the intended moral rightness of AA seems to be nullified and reversed by the fact that it appears to necessarily involve wrongful discrimination.

Each of these puzzles requires a different approach for its solution. Solving the normative puzzle requires justifying the apparent moral trade-offs. Specifically, it is necessary to show that, even if AA involves some form of discrimination, the moral concerns it raises are outweighed by its overall positive effects. It could be argued that AA functions as a policy instrument that contributes to the promotion of an egalitarian social model and, in light of this purpose, can be considered a form of discrimination that is justified by the egalitarian goal that underpins the policy. The empirical question then becomes whether individual policies actually achieve the desired effects to provide sufficient justification.

To resolve the conceptual puzzle, it is crucial to define discrimination in such a way that AA does not qualify as wrongful discrimination. This involves arguing that AA does not fit within the conventional understanding of “discrimination” because its purpose and effects are fundamentally different from those of actions typically regarded as discriminatory. In what follows, I will focus primarily on the “conceptual puzzle” as a potentially self-contradictory concept. To truly resolve this conceptual puzzle, it is not sufficient to argue that the benefits of AA normatively outweigh the wrongs of its discriminatory nature. A satisfactory solution must show that AA is not discriminatory at all. Simply saying: “Yes, it discriminates, but this wrong is justified” misses the point of AA. Instead, a convincing theory of discrimination must allow for AA to be consistently understood as anti-discriminatory. This does not presuppose that AA is always justified in every context, since it may raise moral concerns beyond discrimination. However, this approach would resolve the conceptual puzzle by showing that AA is not inherently self-contradictory.

I will argue that a solution to the PAA, both conceptually and normatively, can be found through a nuanced consideration of different types of social groups. Opponents claim that affirmative action is inherently discriminatory because it results in preferential treatment for individuals based on a specific characteristic, disadvantaging those who do not share that trait. To address this critique, I will conduct a precise analysis of the concept of discrimination, emphasizing its nature as a form of group-based injustice. Building on this analysis, I will introduce a key distinction between oppressed social groups and egalitarian social groups. This distinction will clarify why affirmative action does not constitute discriminatory treatment in certain cases, particularly when it involves measures aimed at rectifying oppression. The proposed conceptual framework allows for a more nuanced distinction between morally acceptable and unacceptable equality measures. Put more formally, my argument can be stated as follows:

- (1) In most cases, disadvantaging someone based on their social group membership constitutes wrongful discrimination.
- (2) The only exception is when favoring members of oppressed groups, thereby

disadvantaging non-members based on their privileged group membership – this does not constitute wrongful discrimination.

- (3) Consequently, when AA falls under this exception, it is not an instance of wrongful discrimination. All other cases that do not involve oppression count as wrongful discrimination.

However, my defense of AA is not exhaustive. I focus specifically on addressing the argument that ties AA to wrongful discrimination. Other objections, such as questions about AA’s effectiveness, will not be addressed here. The objective of this contribution is to unravel the puzzle of AA within the context of discrimination and oppression. It is valuable for a theory of discrimination to create conceptual space for the possibility that AA is not excluded from the outset for conceptual reasons. AA is a policy measure that must be argued for or against with substantive reasons. It would contradict a specific concept of discrimination if it were to preclude this substantive debate conceptually. By dissecting the nuances of these group-based injustices, I aim to illuminate the group-based nature of discrimination, oppression, and affirmative action. This inquiry not only contributes to a deeper understanding of the complex dynamics underpinning these concepts but also provides insights into the broader discussion on justice, equity, and the plurality of social groups.

2. Discrimination as Group-Based Injustice

This section aims to clarify the generic concept of discrimination, asserting that it should be understood as a manifestation of group-based injustice – a key aspect in addressing the initial premise of the PAA. Wrongful discrimination is defined as subjecting someone to a disadvantage based on their social group membership. In what follows, I will affirm and defend this claim. I will argue that the meaning of discrimination in the context of AA should align with its meaning in other cases. Specifically, I will contend that the definition of discrimination must necessarily refer to a group, or alternatively, that the concept may need to be revised to reflect this group-based understanding.

According to Lippert-Rasmussen (2014), Eidelson (2015) and others, generic discrimination occurs when an individual or entity, referred to as ‘X,’ comparatively disadvantages another individual, ‘Y,’ in a particular dimension, ‘D,’ through an action denoted as ‘phi’. Thus, discriminatory behavior involves two necessary conditions:

- i) *Relative Disadvantage*, where discrimination via phi results in Y facing a comparative disadvantage compared to (counterfactual) others, Z, within the same dimension D, and
- ii) *Apparent Differences*, where the discriminatory treatment towards Y is explained by an alleged distinguishing feature, G, between Y and Z.

In this context, “explaining a treatment” encompasses both rational and causal explanations leading to either direct or indirect discrimination (cf. Dancy 2000; Eidelson 2015). Direct discrimination arises from biased perceptions and stereotypes associated with specific social groups, resulting in *disparate treatment*. For instance, consider a hiring manager (‘X’) who discriminates against a job applicant (‘Y’) during the hiring process (‘D’) based on gender (‘G’). If X opts not to hire Y and selects (or would counterfactually select) Z instead, the treatment can be explained by the perception that Y is presumed to be a woman (‘G’) while Z is not. It is essential to note that, in addition to such rational explanations, causal explanations can also account for discrimination, specifically indirect discrimination. For example, the *disparate impact* of a recruitment process on female applicants might be caused by structural factors that discriminate against women without explicitly referring to gender.

In addition, as I now argue, discrimination against someone represents a distinct form of group-based injustice, as discussed by Edmonds (2006), Lippert-Rasmussen (2014),

Altman (2016), Simons (2016), and Moreau (2020), among others¹. This means that the distinguishing characteristic between Y and Z must not be just any characteristic to constitute discrimination; it must be a group-related characteristic. In essence, discrimination can be defined as disadvantageous treatment based on socially significant group membership. The concept of “discrimination” thus highlights a distinct category of reasons – specifically group-related ones – that must not be used to justify unequal treatment. It captures precisely those situations of comparative disadvantage where group membership (to be defined more precisely) serves as a central distinguishing feature. To reflect this within the concept of discrimination, I revise the second necessary condition by introducing a *Criterion of Group Membership*. This criterion holds that discrimination can be attributed to treatment explained by the fact that Y belongs (or is regarded as belonging) to a certain socially meaningful group (‘G’), while Z does not. ‘G’ denotes a specific social group membership, such as gender, race, or other categories. Discrimination thus manifests in various forms, including sexist discrimination, racist discrimination, and more, based on the respective social group that is discriminated against. Importantly, it is not necessary for ‘X’ to harbor negative attitudes or intentions toward the social group ‘G’ as a whole. Discrimination, in this context, is specifically directed at Y, perceived as a member of ‘G,’ and not necessarily against all individuals belonging to ‘G.’ The act of discrimination is explained by ‘G’ membership, indicating that Y is subjected to disparate treatment or impact because of her (perceived) group membership. This specification is articulated as follows: X discriminates against Y in dimension D by phi-ing iff.

- (i) Y is disadvantaged relative to actual or counterfactual others Z (*Relative Disadvantage*), and
- (ii) the disadvantageous treatment of Y is explained by the fact that Y is (perceived as) a member of a certain socially salient group G, while Z is not (*Criterion of Group Membership*).

However, one issue remains to be addressed: Why is discrimination specifically linked to social groups (‘G’) rather than to all possible distinguishing characteristics? I propose three independent but mutually reinforcing reasons for a group-based account of discrimination. Each of these reasons is a pro tanto reason in itself, but together they carry substantial weight.

1. Historically Established: The first reason for emphasizing social group membership is rooted in the recognition that various forms of discrimination are intricately linked to historical and systemic injustices against different social groups. Social groups, particularly those defined by race or gender, have often been the targets of systemic discrimination, enduring collective disadvantages over time. This focus underscores the urgency of addressing and rectifying both historical and contemporary group inequalities. Consequently, the concept of group-based discrimination serves as a crucial framework for understanding and addressing persistent injustices in our societies. Throughout the history of discrimination theory, discrimination has predominantly been defined as a group-based phenomenon. This historical perspective implies that group membership is a fundamental and necessary component of the concept of discrimination. As Altman puts it: “Any viable account of what discrimination is will regard it as consisting of actions, practices, or policies that are – in some appropriate sense – based on the (perceived) social group to which those discriminated against belong” (Altman 2016, sec. 1.1, emphasis added).² Attempting to redefine discrimination without incorporating group-based

¹ This view is contrary to Eidelson’s perspective.

² Consistent with this, see also Edmonds (2006), Lippert-Rasmussen (2013), Altman (2016), Simons (2016), and Moreau (2020)

elements risks discussing an entirely different phenomenon, thereby implicitly diverting attention from the critical issue at hand. The theoretical foundation of discrimination is essentially grounded in acknowledging the role of social groups in perpetuating or challenging systemic inequalities. Deviating from this historical understanding risks undermining the relevance and effectiveness of anti-discrimination efforts.

2. *Positive Legal Codification*: The second reason for a group-based account of discrimination lies in the established legal protection against disadvantages rooted in an individual's association with specific social groups, as enshrined in anti-discrimination laws. For example, Article 14 of the European Convention on Human Rights prohibits discrimination, stating: "The enjoyment of the rights and freedoms set forth in [the] Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth, or other status". While Article 14 provides an open-ended list of protected grounds, indicated by the phrase "any ground such as" and the inclusion of "any other status," it does not prohibit all differences in treatment but only those based on identifiable statuses. Following the logic of the explicitly listed characteristics, it is evident that "status" refers to membership in a socially significant category. The legal codification of discrimination based on group membership reflects society's acknowledgment of the need to address historical and ongoing injustices. This underscores the importance of establishing a group-based concept of discrimination. Ensuring alignment between the philosophical concept of discrimination and the positive legal provisions that significantly shape the practice of (anti-)discrimination is crucial. Any deviation from this alignment should be considered only if compelling reasons exist. Philosophical definitions of discrimination should, *prima facie*, align with the legal framework, or compelling reasons must be presented to justify any revision.

3. *Systematic Adequacy*: The third reason for limiting the criteria considered discriminatory to group membership is that it helps address the problems associated with a broader, unrestricted definition of discrimination, which recognizes any distinguishing characteristic regardless of social group membership. It is a well-established practice to test definitions through counterexamples. Extensional adequacy is characterized, among other things, by a definition's ability to capture prototypical cases and handle problematic ones. The definition must possess explanatory power, meaning it should clearly distinguish between cases that fall under the term and those that do not. Let us now consider Narveson's proposal, which defines discrimination without reference to groups, and examine this definition in light of cases that require explanation to demonstrate that a group-based definition fares better in terms of extensional adequacy. Narveson states: "A discriminates against B in relation to C by doing $x =_{\text{Def}}$ There is a property, K, such that B has K, C does not have K, A treats B worse than C by doing x, and does so because B has K" (Narveson 2002, p. 204). Clearly, this unrestricted definition is overly broad: a state that punishes only offenders, extends social welfare exclusively to the needy, or awards medals solely to deserving citizens is not discriminating against these individuals, even though it differentiates based on a distinguishing feature (cf. Alexy 2009). However, for reasons of logical consistency, these actions would fall under Narveson's unrestricted definition and would have to be regarded as wrongful discrimination. I suggest that the absence of discrimination in these cases is best explained by the lack of socially salient groups. A judge who assesses defendants solely based on guilt or innocence is not discriminating. However, it is a different matter if the judge weighs the severity of punishment or the necessary evidence differently based on the defendants' skin color. In that case, the judge is indeed discriminating. This is because guilt is a trait attributable to an

individual act, not to membership in a social group. If someone is convicted based on their gender or ethnic background, it qualifies as discrimination, as it involves considering group membership. Narveson's definition of discrimination, which describes it as unequal treatment based on any distinguishing characteristic, fails to capture and explain the difference between these two cases. A definition that employs group membership as a criterion better accounts for these and other cases, such as the distinction between meritocratic and sexist job allocation. Adverse hiring decisions made based on sexist, racist, or other biases (whether by the decision-maker or by third parties) are discriminatory because they are based on the (perceived) group membership of the affected individuals. Conversely, hiring a more qualified candidate does not amount to discrimination against less qualified competitors, as the decision is based on individual qualities—specifically, the superior qualifications of one person compared to others—and not on group membership. While this is not a conclusive reason, it strongly suggests why a group-based definition is preferable. Focusing on group-based characteristics allows us to better address the intricate and often subtle ways in which discrimination operates. It enables a more precise and systematic understanding, which is essential for developing effective anti-discrimination strategies and policies. The efficacy of defining discrimination in terms of group membership in navigating the complex cases that arise from the unrestricted definition provides a compelling reason to recognize group membership as a necessary condition for 'discrimination.' I will present additional reasons to support this point in the next section.

A group-based understanding of discrimination is deeply rooted in history, acknowledged by legal frameworks, and systematically compelling. It recognizes the central role that group membership plays in perpetuating various injustices, providing a robust foundation for addressing and combating discrimination in its various manifestations. The Criteria of Group Membership (CGM) acts as a necessary condition within the concept of discrimination, asserting that a comparative disadvantage is considered discriminatory only when the affected individual experiences it due to their (presumed) membership in a specific social group. This criterion not only serves as a necessary condition for the concept of discrimination but also plays a pivotal role in elucidating why discrimination is morally objectionable. The CGM clarifies the source of these normative claims by pointing to group-based affiliations as the root cause of discriminatory practices. In doing so, it highlights the connection between discrimination and broader moral principles that condemn unjust treatment based on attributes beyond one's control. The CGM not only clarifies the group-based origin of normative claims but also demonstrates the link to overarching moral principles, as articulated by luck egalitarians: comparative disadvantages are morally permissible only if they result from circumstances for which the affected person is responsible, traceable to their voluntary decisions or avoidable mistakes (for an overview, see Lippert-Rasmussen 2015). In essence, the CGM is a pivotal element of the concept of "discrimination," serving both as a necessary condition and as a moral principle. It sheds light on why discrimination is morally wrong by emphasizing the connection between group-based normative claims and fundamental principles of justice and equality.

However, establishing which types of group membership are pertinent to the concept of discrimination is crucial for defining its scope and applicability appropriately. I have argued so far that the concept of discrimination requires a restrictive definition, limiting it to cases where disadvantaged treatment is explicitly linked to social groups. On the other hand, I will now propose an inclusive understanding of what constitutes a relevant social group. By broadening the scope of recognized social groups, I aim to capture the full complexity of discrimination and address the diverse ways in which group membership can influence

3. The Criterion of Group Membership Reconsidered

individuals' experiences of disadvantage. This dual approach allows for a precise definition of discrimination while also acknowledging the intricate ways in which social groups impact these experiences. By reconciling these two perspectives, I seek to provide a comprehensive framework for understanding and addressing discrimination in its various manifestations. Two main approaches – the realistic (a) and the classificatory (b) approach – shed light on different aspects of group membership. I will advocate for the classificatory account for the reasons presented shortly (for the following see also Behrendt 2020).

a) Real Collectives: The realistic approach conceptualizes groups as concrete collectives emerging from social communication and interaction contexts – such as a football team or the orchestra of a city theater. Members actively engage over an extended period, nurturing the development of shared experiences, values, norms, and a sense of belonging (cf. Schweikard 2011; Schäfers 2016). As such a collective expands, however, the likelihood increases that direct contact between members occurs only sporadically, and the majority are widely dispersed. Crucial, nonetheless, is the capacity to convene the group members at a given time. For instance, a particular orchestra retains its identity as the same group even if different musicians are engaged at different times, and some may participate more or less in rehearsals. Genuine collectives stand in contrast to mere gatherings of individuals, such as the attendees of a concert or those waiting in front of a store (cf. French 1984, pp. 5-18.). Individuals standing in line at the bakery or coincidentally sitting in the same concert hall do not belong to a social group in the realistic sense. Unlike spontaneous gatherings of people with no further connecting element than being at the same time at the same place due to coincidentally overlapping intentions, genuine groups are characterized by their members being integrated into the internal organizational structure of the collective in a specific way.

This approach may be deemed too restrictive, as it does not fully encompass paradigmatic forms of discrimination, such as those based on gender or race. For sure, discrimination can frequently occur without the presence of fixed, real-world group identities. Genuine collectives often emerge in response to experiences of discrimination rather than preceding them. Moreover, it's important to note that group membership, grounded in actual shared interactions, often hinges on voluntariness or “joint commitment” (Gilbert 2006). For example, collecting contributions from members of an association based on their membership would not be considered discrimination, as association participation is typically a voluntary choice.

b) Classificatory Group Membership: Concrete, spatiotemporally locatable genuine collectives can be distinguished from abstract symbolic classifications of individuals based on shared features (cf. Goffman 1963; Haslanger 2012; Ritchie 2015; 2018). While, in principle, a distinct class of shared characteristics could be identified based on any given trait, it is crucial not to equate socially significant and purely logical categories. The latter lacks practical significance in social contexts, as Anthony Appiah rightly emphasizes: “There is a logical but not a social category of the witty, or the charming, or the greedy: people who share these properties do not constitute a social group, in the relevant sense” (Appiah 1996, 93). The classificatory approach perceives group membership based on shared attributes as socially significant. Social kinds, such as Men/Woman, Black/White, etc., in turn, can be divided into overarching social categories: “A category is a tree of classifications, or else the most general classification at the top of such a tree. [...] Race, gender, native language, nationality, type of employment, and age cohort are all what I call categories” (Hacking 1995, p. 355). These attributions can be rooted in individual characteristics like disability or in the attributes of actual groups (see above), such as religion. Social categories and their associated social kinds are not essentialistically grounded like in the case of natural kinds but owe their constitutive foundation to societal

construction mechanisms, although their historical contingency is often concealed behind essentialization.

The classificatory approach is more versatile and flexible, as it does not necessitate similarly stringent prerequisites as the realistic approach, such as a shared collective identity but can incorporate it. According to the classificatory approach, relevant group membership pertains to societally significant categories. This implies that being a part of these groups plays a pivotal role in specific social interactions and/or holds considerable importance in various social contexts (Lippert-Rasmussen 2013, p. 31). This significance can be observed in the action dimension, where group membership significantly influences interactions, or in the symbolic dimension, where group membership entails identity-forming and value-laden characteristics (Korsgaard 2009). To further illuminate this point, Sally Haslanger's distinction between "thick" and "thin" social positions is helpful, which she characterizes as follows:

Social constructionists interested in the impact of categorization on individuals are usually interested not only in the nominal classification 'widow' or 'wife' (etc.) but also in the system or matrix of practices and institutions that create 'thick' or 'robust' social positions, that is, social positions that entail a broad range of norms, expectations, obligations, entitlements, and so on. [...] the basic idea is that some social positions carry with them more demanding norms, expectations, and obligations than others; some carry more privileging entitlements and opportunities than others. 'Thin' social positions carry very little social weight. 'Thick' social positions can empower or disempower the groups standing in those positions (Haslanger 2012, p. 126).

The social world is not just imbued with meaning but is also normatively structured. Embedded within cultural symbols and systems of meaning, shared horizons of interpretation stabilize, conferring social significance to classifications and assigning social positions to those classified. Within these positions, normative behavioral expectations are consolidated, stemming from the assigned normative status encompassing rights, duties, entitlements, and more – referred to as the "deontic powers" by Searle (1995). Consequently, the careful selection of relevant types of group membership for the concept of discrimination is of great significance, as it significantly influences the scope and applicability of the concept in the analysis and mitigation of discrimination. An inclusive consideration of various aspects of group membership contributes to a comprehensive understanding of discrimination in its diverse manifestations and contexts. This approach proves instrumental in effectively addressing discrimination.

The concept of discrimination presented in the classificatory framework provides several distinct advantages over the realistic one. *Firstly*, it refines and expands the definition of antidiscrimination law, making it more inclusive and responsive to contemporary social realities. One notable advantage of this approach is its ability to recognize and address forms of discrimination that may not be explicitly covered by traditional anti-discrimination laws. For instance, consider a scenario where a hiring manager rejects a job applicant because she is a single mother. While traditional antidiscrimination laws might not explicitly list "single mothers" as a protected ground, this approach acknowledges the social significance of this category. It recognizes that single motherhood can be a defining aspect of an individual's life and that discrimination based on this characteristic is, in fact, a form of discrimination. Therefore, under this approach, such an act of rejection would also be considered discriminatory. This expanded perspective allows for a more nuanced and comprehensive understanding of discrimination, ensuring that individuals are protected

from unjust treatment based on characteristics that may not have been explicitly foreseen by legislators. By recognizing the social significance of various categories and attributes, the proposed approach fosters a more inclusive anti-discrimination framework that reflects the complex realities of contemporary society. It ensures that individuals are protected from discriminatory practices in a broader range of contexts.

Secondly, the proposed framework aligns well with a fundamental moral principle: that individuals have legitimate claims to be comparatively disadvantaged only as a result of their own responsible actions and avoidable mistakes, rather than due to inherent aspects of who they are (Dworkin 1977; Gosepath 2015). The classificatory concept of discrimination can explain why this principle holds and how discrimination is linked to it. The fundamental idea is this: Group membership represents a historically and systematically significant subset of distinguishing features that, in principle, do not provide generally acceptable reasons for disadvantage. The thesis is that comparative disadvantages are morally permissible only if they result from circumstances for which the affected person is responsible – specifically, if they can be traced back to their voluntary decisions or avoidable mistakes. This is never the case with social group membership as defined by the classificatory account, because it involves impersonal attribution practices. Given that autonomous actors are always already interconnected within social relations, individuals are fundamentally not free to act independently of a given symbolic order. Social categories, therefore, possess certain elements that appear as an unavoidable given to individuals. The social category to which someone belongs primarily depends not on the successful acquisition and affirmation by the socialized subject but on the intersubjective validity within the context of a particular cultural way of life. Because social classifications ultimately owe their validity to the normative attitudes and collective stances of their members, they present themselves to individuals as an impersonal structure with which they must always engage.

Nonetheless, it should be noted that, in some cases, people can control certain conditions that determine which social category they belong to. A clear example is the category of religion and worldview. People can belong to a particular religion from birth and remain faithful to it throughout their lives, but they can also abandon their faith or change denominations over the course of their lives. For instance, if a person converts to Buddhism for religious reasons and a military militia takes over the country where the person lives, subsequently banning Buddhists from entering public buildings, this constitutes a textbook case of discrimination, even though religious affiliation is chosen voluntarily. Thus, one might object by arguing that individuals can freely choose certain group memberships, such as religion. This appears to be a valid point. In response, let us consider the following three points.

- (1) It is crucial to differentiate between choosing membership within specific social kinds, such as one's particular religious or gender identity (e.g., Catholic or Buddhist, man or woman), and the overarching categories themselves (e.g., gender, religion). Social categories are shaped by overarching, impersonal cultural attributions over which individuals have limited control. The advantages or disadvantages associated with these categories stem from factors that can be considered purely matters of luck or misfortune – “brute (bad) luck”.
- (2) Social categories often carry profound existential significance, embodying core aspects of an individual's “true self”, as articulated by Taylor (1985). They should not be seen as objects of free choice driven by preferences or reasons but rather as manifestations of one's authentic identity, with elements that are inherently valuable and not merely instrumental for achieving other ends. Comparative disadvantages arising from these identities can be understood as a form of “constitutive bad luck”.

- (3) Crucially, individuals lack control over how others perceive and treat them based on their social kind. Discrimination can arise due to societal views and treatment of certain groups, factors beyond individuals' control. It is objectionable that individuals bear the cost of others' perceptions (Sangiovanni 2017; Moreau 2020). Given the inherent value of these categories (as indicated in (2)), discriminatory treatment can be viewed as a manifestation of "circumstantial bad luck".

In essence, this response acknowledges that while individuals may make choices within specific social categories, overarching social categories are often beyond individual control and carry inherent value. Discrimination related to these categories frequently results from societal perceptions and circumstances, constituting various forms of undeserved misfortune.

Let's now turn to premises 2 and 3 of the PAA. A perplexing aspect of discrimination emerges when members of specific groups are comparatively disadvantaged in relevant dimensions to neutralize unearned advantages. In this context, AA may appear to embody morally wrong discrimination (cf. Newton 1973, p. 310; Eastland and Bennett 1979, p. 149). For instance, affirmative action policies may prioritize hiring women over men in the public sector. According to the standard framework proposed earlier, this might seem like sexist discrimination against men, as they are being disadvantaged in one dimension (e.g., work-life) based on their gender. How can we address this while preserving the justification for AA?

One way to reconcile this paradox of affirmative action is by distinguishing between discrimination and oppression. The claim that "there is no such thing as sexism against men, racism against whites, or classism against the rich" reflects a common, though often contested, understanding of how oppression operates within societal structures. This viewpoint suggests that AA is meaningful primarily within the context of larger, entrenched power dynamics, specifically under oppressive conditions. While individuals from privileged groups, such as men, whites, or the wealthy, might experience prejudice or even individual acts of unfair treatment, these instances lack the systemic, power-driven nature that characterizes sexism, racism, or classism. This type of argument in support of AA claims that it is not discriminatory when it benefits those affected by prior discrimination, even if it results in disadvantages for others. In this view, AA is seen as a corrective mechanism aimed at remedying historical wrongs. For instance, if certain groups were systematically excluded from educational opportunities or employment, it is justifiable to offer them preferential treatment to restore fairness. This understanding emphasizes that the disadvantaged groups are, in effect, compensating for harms inflicted upon them by society over time. However, this argument runs into a significant challenge: Affirmative Action policies do not solely affect individuals who were directly involved in past discrimination. Those who benefit from AA today may not be the same individuals who directly experienced the discrimination of previous decades. Similarly, those who are disadvantaged by these policies, such as white applicants in university admissions, may have had no personal involvement in perpetuating or benefiting from historical injustices. This creates a tension between the claim that AA is not discriminatory and the reality that the direct effects of these policies may not perfectly align with the perpetrators and victims of past discrimination.

A more robust defense of AA (and the claim that it does not constitute discrimination) involves shifting the focus from prior discrimination to current structures of oppression and privilege. AA is not about punishing those responsible for historical wrongs or compensating those who directly suffered from past discrimination. Instead, it aims to address ongoing inequalities rooted in oppression, understood as structural and systemic power imbalances between social groups. This view provides a broader and more effective explanation: AA is not discriminatory when it favors those who suffer from current oppression and disadvantages

4. Resolving The Puzzle of Affirmative Action

only those who are privileged within the existing power structure. The key distinction here is between isolated instances of bias or prejudice, which can affect anyone, and the deeper, more enduring relations of power that shape societal structures. Oppression is not merely about individual acts of injustice or discrimination. It is about the social, political, and economic systems that privilege certain groups (such as men, whites, or the wealthy) while disadvantaging others (such as women, people of color, or the poor). By focusing on oppression as a dispositional power relation, rather than an isolated event like prior discrimination, we can better understand why AA policies are not simply instances of “reverse discrimination” but rather a means of addressing entrenched inequalities. Solely focusing on prior discrimination is too narrow because it treats discrimination as an event—something that happened in the past, with identifiable perpetrators and victims. This framing overlooks the ongoing nature of systemic inequalities and the intergenerational transmission of advantages and disadvantages. Oppression, by contrast, is an ongoing structure of relations that persists even if the original discriminatory acts are no longer occurring. For example, even if explicit racial discrimination in hiring or education has been formally outlawed, people of color may still face significant disadvantages due to historical legacies, implicit biases, and structural inequalities. Similarly, even if women have achieved legal equality, they continue to encounter systemic barriers to advancement in many fields due to entrenched patriarchal norms. These instances illustrate that oppression persists, regardless of whether the individuals currently benefiting from privilege or suffering from disadvantage were directly involved in the historical events that established these dynamics. Focusing on privilege helps clarify the true purpose of AA. AA does not aim to disadvantage individuals for personal wrongdoing or even for belonging to a particular group. Instead, it seeks to address the structural advantages that privileged groups enjoy, often unconsciously, due to the way society is organized. Privilege refers to unearned advantages associated with one’s social identity – advantages that may not be immediately visible to those who possess them but that significantly impact life opportunities. By framing AA in terms of oppression and privilege, we see that these policies are not about punishing individuals from dominant groups. Instead, they aim to level the playing field by reducing the unearned advantages that privileged groups (such as men, whites, or the wealthy) continue to enjoy, even if they are not personally responsible for creating those advantages. Thus, focusing on oppression as a dispositional power relation allows us to capture more cases of unjust inequality and explains why AA is not merely reversing discrimination but addressing ongoing power imbalances.

Both discrimination and oppression are subtypes of injustice, specifically concerning social groups (see, among others, Frye 1983; Young 1990; Haslanger 2004; Edmonds 2006; Lippert-Rasmussen 2013; Altman 2016). “Discrimination” represents a type of action, whereas “oppression” embodies a structural concept. In this context, “discrimination” characterizes a moral injustice enacted upon an individual in a particular situation by a specific actor, whether an individual or a collective. It denotes disadvantageous treatment rooted in socially significant group membership. On the other hand, the concept of “oppression” encapsulates an unjust state of affairs characterized by structural power imbalances, which result in systematic disadvantages for an entire group through the assignment of subordinate social positions. While discrimination in specific instances may reflect existing relations of oppression when an individual faces a disadvantage as a member of an oppressed group, it does not always require this precondition. Even in a society considered just, where no social group is oppressed – indicating the absence of structural disadvantages for any particular social group – discrimination against individuals based on their group membership can still occur. Conversely, conceptualizing oppression proves challenging without acknowledging a sequence of potentially historical acts of discriminatory disadvantage. These acts, which

initially establish and perpetuate repressive power relations, form the foundation of oppression. Consequently, a prohibition of discrimination serves to shield individuals from individual disadvantage based on their (presumed) group membership, while liberation from oppression consistently aspires to socially reposition an entire social group. To be more specific, members of a social group denoted as G, experience oppression as Gs when two necessary conditions are met (cf. Stahl 2017). Specifically:

1. *Asymmetrical Power Relations*: For oppression to manifest, a social practice (SP) must establish or perpetuate an asymmetrical power relationship (R) between members of group G and members of other groups, denoted as Z. This power imbalance should lead to a disadvantage for Gs in at least one relevant dimension (D) compared to Zs.
2. *Systematic Barriers*: The social practice (SP) involves the presence of systematic barriers that impose higher costs on Gs than on Zs for altering the structure of R or exiting SP. These barriers can be attributed to the mere fact that Gs are members of the group G.

It's essential to emphasize that oppression is a collective experience, affecting all members of the oppressed group. The term "systematic barriers" encompasses a range of factors, such as cultural stereotypes, the unequal distribution of material resources, and prevailing social norms. Furthermore, it's crucial to acknowledge that the privileged group Zs may not bear responsibility for creating or intentionally perpetuating the power imbalance (R) or the oppressive social practice (SP). In this context, privilege refers to the unearned benefits enjoyed by Zs due to existing power dynamics. This underscores the unintentional advantages that members of the privileged group receive because of societal structures, even if they did not actively contribute to their creation or perpetuation. For example, consider the case of women (G) facing sexist oppression. Women experience this oppression when a gender role model (SP) is established or perpetuates asymmetrical power relations (R) between women (Gs) and men (Zs), resulting in disadvantages for women e.g. in their professional lives (D). This may manifest in pay disparities, underrepresentation in leadership roles, or limited access to professional opportunities (Asymmetrical Power Dynamics in the Workplace). Additionally, the workplace environment includes systematic barriers, which present greater challenges for women than men in altering the power structure (R) or escaping the oppressive role model (SP). These barriers are inherently tied to the fact that women are regarded as women (Systematic Barriers). This conceptual framework for understanding oppression highlights the systematic nature of discriminatory practices and their impact on an entire group. It provides a holistic perspective on the mechanisms through which oppression operates, enabling a deeper understanding of its dynamics and the formulation of effective strategies to address it. To incorporate this distinction into the concept of AA, the framework of discrimination must encompass two distinct categories of socially significant kinds: a) egalitarian kinds and b) oppressive kinds.³

a) *Egalitarian social kinds* play a crucial role in fostering relational equality among social groups, serving as beacons of inclusivity and acknowledgment. These categories emphasize the active affirmation and recognition of group membership by individuals within these groups. Unlike hierarchical structures that perpetuate inequality, egalitarian categories strive to create a level playing field where each social group is afforded equal respect

³ It is worth noting that the distinction represents an ideal type in Max Weber's sense and rarely occurs in its pure form (Weber 2012). In reality, the spectrum of cases from oppression to egalitarian relations is more nuanced. However, ideal-type distinctions do not imply the assumption of an ideal theory in the sense of Rawls. Conceptually, this approach opens up a spectrum that does not, however, normatively presuppose ideal conditions.

and recognition. Members within egalitarian categories actively engage in affirming and recognizing the importance of their group membership, contributing to a sense of belonging and shared identity. This commitment to relational equality lays the foundation for a more just and inclusive society, where the diversity of social groups is not only acknowledged but celebrated, fostering unity and collective responsibility among its members. Through their principles of affirmation and recognition, egalitarian categories embody a vision of a society where diversity is embraced, and each social group is regarded with equal dignity and value.

b) *Oppressive social kinds*, in stark contrast to egalitarian ones, are characterized by the systematic subordination of specific social groups. Within these categories, a troubling dynamic exists where certain groups are consistently placed at a disadvantage, perpetuating an unjust hierarchy. The inherent risk associated with oppressive categories lies in the potential for objectification and alienation of their members. The imposition of social subordination not only denies individuals within these groups equal opportunities and recognition but also subjects them to the dehumanizing effects of objectification. Members of oppressed categories may find themselves marginalized, treated as objects rather than individuals with inherent worth. This risk of objectification and alienation underscores the profound moral concerns associated with oppressive categories, as they contribute to a societal framework that undermines the principles of equality and human dignity. Addressing and dismantling oppressive categories is crucial for fostering a more just and inclusive society—one that rejects the perpetuation of social subordination and values the inherent humanity and worth of every individual, regardless of their social group membership.

Both types of social categories can potentially give rise to instances of discrimination. However, in the case of oppressive categories, discrimination is exclusive to members of the oppressed groups. For instance, there is no concept of classism against the wealthy in classist societies or sexism against men in patriarchal ones, as the dynamics of discrimination are not applicable to privileged groups in these contexts. In the context of affirmative action, the distinction between egalitarian and oppressive categories helps us understand the intention behind policies that favor one group over another. Such policies are aimed at addressing the historical disadvantages faced by the oppressed group and promoting relational equality. While these policies may involve relative disadvantages for the non-oppressed group in certain dimensions, they are not discriminatory in the same sense as when members of oppressed groups are disadvantaged in broader societal contexts. This conceptual framework aids in navigating the complex terrain of affirmative action and anti-discrimination, recognizing that the intention behind these policies is rooted in the pursuit of a fairer and more equal society. By distinguishing between different types of socially significant categories, we gain a clearer understanding of the nuances of discrimination and the motivations behind policies designed to address historical injustices.

Building on the insights gained thus far, we can now develop a solution for the PAA. In cases involving social categories with egalitarian social kinds, we can maintain the standard perspective. In such instances, a disadvantage based on group membership represents morally wrong discrimination. However, the scenario changes when addressing social groups with social kinds in a relationship of oppression. In these cases, disadvantages imposed upon the privileged group do not constitute wrongful discrimination, provided that they serve the purpose of rectifying or equalizing asymmetrical power relations. Members of oppressed groups have specific normative claims for equality or liberation that outweigh the claims of privileged groups not to be disadvantaged. When members of privileged groups are disadvantaged in favor of members of oppressed groups in relevant dimensions, it does not constitute discrimination. This is true when such measures are taken to rectify or transform

asymmetrical power relations. In other words, affirmative action policies that aim to address historical injustices and power imbalances between privileged and oppressed groups are not discriminatory, as they seek to promote equality and social repositioning.

In this comprehensive exploration of discrimination, oppression, and affirmative action, I have developed a nuanced framework to navigate the complexities of these essential concepts. This framework recognizes the interplay between social categories, power dynamics, and the pursuit of justice and equality. A key insight is that, in cases of oppression, imposing disadvantages on privileged groups can be a legitimate means to rectify or alter power imbalances—essentially the core of affirmative action. These measures are not discriminatory but are aimed at promoting equality and social repositioning. This provides my solution to the conceptual puzzle. Additionally, this framework highlights the distinct roles of anti-discrimination laws and affirmative action in addressing social inequalities. While anti-discrimination laws protect individuals from unjust treatment based on group membership, affirmative action policies seek to correct historical injustices and power imbalances by promoting equal opportunities. My discussion has thus provided a valuable framework that not only addresses the complexities of discrimination and oppression but also clarifies the often-misunderstood concept of AA. This approach aligns with moral intuitions, reflects the inherent value of social categories, and advocates for merit-based treatment while recognizing the need for corrective measures to address historical injustices. It reinforces the pursuit of a more equitable and just society, where fairness and equality guide our actions and policies. However, my proposal does not make any substantive assumptions about whether affirmative action is justified all things considered. To assess this, empirical investigations are needed to first determine whether relations of oppression truly exist in the relevant context and, second, to verify whether the intended measures effectively achieve the goal of combating oppression.

Finally, I want to address a concern that might arise from my approach. The idea that AA does not represent discrimination only in cases of oppression, while being discriminatory in all other instances, might seem too restrictive. Some might want to argue that AA is defensible even in cases of social inequalities that are less systemic than oppression. Limiting the justification of AA to cases of oppression could leave too much room for the charge of wrongful discrimination, even if AA serves important policy goals regardless of whether oppression is present. It seems that my argument concedes that AA fails to justify imposing disadvantages independently of the oppressive nature of the inequalities it seeks to address. To address this concern, I would like to emphasize the distinction between normative and conceptual puzzles, as specified in the introduction. I am not claiming that equality measures are justified solely to combat oppression. Instead, I argue that AA, as spelled out in this contribution, finds its conceptual place within this framework. However, it is important to highlight that discrimination is only *pro tanto* morally wrong, meaning its wrongness in specific contexts can be outweighed by other factors (Lippert-Rasmussen 2013, Moreau 2020). AA may serve important policy goals, whether or not oppression is present. It can be viewed as a policy instrument that contributes to promoting an egalitarian social model. Given this purpose, it can be argued that affirmative action is a form of discrimination justified by the egalitarian goal underpinning the policy. This provides a further solution to the normative puzzle beyond cases of oppression. However, it is important to clarify that, in this case, it would indeed be considered discrimination. A detailed investigation of the individual case is necessary to determine whether this otherwise wrongful discrimination can be justified in the name of justice.

5. Conclusion

REFERENCES

- Alexy, R. (2009). *A Theory of Constitutional Right*. Oxford: OUP;
- Altman, A. (2016). Discrimination. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). URL = <https://plato.stanford.edu/archives/win2016/entries/discrimination/>;
- Appiah, A. (1996). Race, Culture, Identity: Misunderstood Connections. In A. Appiah and A. Gutmann (Eds.), *Color Conscious. The Political Morality of Race* (pp. 30-105). Princeton: PUP;
- Behrendt, H. (2020). Diskriminierung und das Kriterium der Gruppenzugehörigkeit. *Zeitschrift für Praktische Philosophie* 7(1), 155-190;
- Dancy, J. (2000). *Practical Reality*. Oxford: OUP;
- Dworkin, R. (1977). *Taking Rights Seriously*. Cambridge, MA.: HUP;
- Eastland, T. & Bennett W. J. (1979): *Counting By Race*. Chicago: Basic Books;
- Edmonds, D. (2006). *Caste Wars. A Philosophy of Discrimination*. London/New York: Routledge;
- Eidelson, B. (2015). *Discrimination and Disrespect*, Oxford: OUP;
- French, P. A. (1984). *Collective and Corporate Responsibility*, New York, NY: CUP;
- Frye, M. (1983). Oppression. In *The Politics of Reality: essays in feminist theory*. New York: Crossing Press;
- Gilbert, M: (2006). *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford: OUP;
- Gosepath, S. (2015). The Principles and the Presumption of Equality. In C. Fourie, F. Schuppert, & I. Wallimann-Helmer (Eds.), *Social Equality. On What It Means to Be Equals*, (pp. 167-185). Oxford: OUP;
- Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, New Jersey: Prentice-Hall;
- Hacking, I. (1995). The Looping Effect of Human Kinds. In D. Sperber et al. (Eds.), *Causal Cognition: An Interdisciplinary Approach* (pp. 351-383). Oxford: OUP;
- Haslanger, S. (2004). Oppressions: Racial and Other. In M. P. Levine and T. Pataki (Eds.), *Racism in Mind* (pp. 97-124), Ithaca: Cornell University Press;
- Haslanger, S. (2012). Social Construction: The ‘Debunking’ Project. In *Resisting Reality. Social Construction and social Critique* (pp. 113-138). Oxford: OUP;
- Korsgaard, C. (2009). *Self-Constitution: Agency, Identity, and Integrity*, Oxford: OUP;
- Lippert-Rasmussen, K. (2013): *Born free and Equal? A Philosophical Inquiry into the Nature of Discrimination*. Oxford: OUP;
- Lippert-Rasmussen, K. (2015). *Luck Egalitarianism*, London: Bloomsbury;
- Moreau, S. (2020). *Faces of Inequality: A Theory of Wrongful Discrimination*. Oxford: OUP;
- Narveson, J. (2002). *Respecting Persons in Theory and Practice*. New York: Rowman and Littlefield;
- Newton, L, (1973). Reverse Discrimination as Unjustified, *Ethics* 83, 308-312;
- Ritchie, K. (2015). The Metaphysics of Social Groups. *Philosophy Compass* 10(5), 310-321. DOI: <https://doi.org/10.1111/phc3.12213>;
- Ritchie, K. (2018). Social Creationism and Social Groups. In K. Hess, T. Isaacs & V. Ignieski (Eds.), *Collectivity: Ontology, Ethics, and Social Justice*, (pp. 13-34). Lanham, MD: Rowman and Littlefield;
- Sangiovanni, A. (2017). *Humanity without Dignity: Moral Equality, Respect, and Human Rights*. Cambridge: Harvard University Press;
- Schäfers, B. (2016). Die soziale Gruppe. In H. Korte & B. Schäfers (Eds.), *Einführung in Hauptbegriffe der Soziologie*, revised and updated edition (pp. 153-172). Wiesbaden: Springer VS. DOI: 10.1007/978-3-658-13411-2;
- Schweikard, D. (2011). *Der Mythos des Singulären: Eine Untersuchung zur Struktur kollektiven Handelns*. Paderborn: Mentis;

- Searle, J. (1995). *The Construction of Social Reality*. Free Press: New York, NY;
- Simons, K. (2016). Discrimination is a Comparative Injustice: A Reply to Hellman. *Va. L. Rev. Online*, 102, 85-100;
- Sowell, T. (2004). *Affirmative Action Around the World, An Empirical Study*, New Haven & London: YUP;
- Stahl, T. (2017). Collective Responsibility for Oppression. *Social Theory and Practice*, 43(3), 473-501;
- Sterba, J. P. (2004). Completing Thomas Sowell's Study of Affirmative Action and Then Drawing Different Conclusions. *Stanford Law Review* 57(2), 657-693;
- Taylor, C. (1985). What is human agency? In *Philosophical Papers*. Vol 1. (pp. 15-44). Cambridge: Cambridge University Press;
- Weber, M. (2012) [1904]. The "Objectivity" of Knowledge in Social Science and Social Policy. In H. Bruun & S. Whimster (Eds.), *Max Weber. Collected methodological writings*, (pp.100-138). London, New York: Routledge;
- Young, I.M. (1990). *Justice and the Politics of Difference*, Princeton: PUP.