

# IMPLICIT BIAS, COLORBLINDNESS AND INSTITUTIONAL RACISM

Megan Mitchell

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill  
2014

Approved by:

Bernard Boxill

Thomas E. Hill, Jr.

Robert Paul Wolff

Jan Boxill

Ryan Preston-Roedder

©2014  
Megan Mitchell  
ALL RIGHTS RESERVED

## ABSTRACT

MEGAN MITCHELL: Implicit Bias, Colorblindness and Institutional Racism  
(Under the direction of Bernard Boxill)

This dissertation concerns the nature and extent of institutional racism. It consists of three articles, each of which draws upon the history of and current conditions facing black Americans to establish more general conclusions about institutional racism.

In the first article, I observe that striking inequalities in education, income, incarceration rates and employment have persisted between black and white Americans despite a decline in explicit anti-black racism among whites and a rise in legislation intended to prevent many of the most virulent forms of discrimination. I ask, “Are these pervasive inequalities the result of institutional racism?”

I notice that for some social scientists (I call them “structuralists”) the answer to this question is trivial. They hold that racism is a system wherein advantages are divided along racial lines. Consequently, institutions that create or support racial inequalities are, simply in virtue of that fact, racist.

I maintain that the question is meaningful and the answer important. Against the structuralists, I defend the view that *institutions* are only racist insofar as they perpetuate the racism of *agents*. So, if current racial inequalities are the result of institutional racism, then it must be the case that the institutions that cause them were created or are currently sustained

by racist individuals, who, consciously or unconsciously, express their racist beliefs and attitudes through the policies, practices, and organizational structures they adopt or maintain.

However, if part of the motivation behind presenting a theory of institutional racism is, as I think it should be, to aid in its eradication, one might worry that this more restrictive theory will fail to yield any targets for political action. The increasingly covert and unconscious nature of much agential racism could make it very difficult to prove that any specific institutional actions are racist. I assuage this worry by demonstrating in the second article that the pervasive phenomenon of implicit racial bias is an instance of institutional racism. In the third and final article, I argue that colorblind policies and rational racial profiling, when perpetuated against black Americans by the state, are racist.

To Christian, for providing an example worth following

## TABLE OF CONTENTS

<b>1. TWO CONCEPTIONS OF RACISM.....</b>	<b>1</b>
I. INTRODUCTION.....	1
2. THE TWO CONCEPTIONS OF RACISM.....	6
2.1 <i>The Agential Conception of Racism</i> .....	7
2.2 <i>The Structuralist Conception of Racism: Three Interpretations</i> .....	13
3. EXTENDING THE AGENTIAL ACCOUNT .....	26
4. THE MESSY AND COMPLICATED TASK OF CARVING UP RESPONSIBILITY .....	34
<b>2. IMPLICIT RACIAL BIAS AND INSTITUTIONAL RACISM .....</b>	<b>36</b>
1. INTRODUCTION .....	36
2. IMPLICIT RACIAL BIAS: WHAT IS IT AND WHAT DOES IT DO? .....	37
3. ARE IMPLICIT BIASES RACIST? THE STATE OF THE DEBATE .....	46
3.1 <i>What Is the Bias Itself? Implicit Biases as Mental States</i> .....	52
3.2 <i>Are Agents at Fault for their implicit Biases?</i> .....	54
4. IMPLICIT BIAS AS INSTITUTIONAL RACISM .....	61
4.1 <i>The Agential Account of Implicit Bias as Institutional Racism</i> .....	66
5. CONCLUSION .....	76
<b>3. THE SEARCH FOR FORMAL EQUALITY OF OPPORTUNITY: A CASE FOR THE PREFERENTIAL TREATMENT OF BLACK AMERICANS ....</b>	<b>78</b>
1. INTRODUCTION .....	78

1.1. <i>The Backward-Looking Approach</i> .....	80
1.2 <i>The Forward-Looking Approach</i> .....	83
2. PREFERENTIAL TREATMENT FOR THE VULNERABLE .....	86
2.1 <i>Formal Equality of Opportunity and the Problem of Distinguishing Vulnerability</i> . 90	
3. BLACKS AS A VULNERABLE POPULATION? .....	94
3.1 <i>A New Theory of Black Vulnerability</i> .....	97
4. PROTECTING THE VULNERABLE FROM HARM THROUGH PREFERENTIAL TREATMENT ..	101
5. WHAT’S WRONG WITH RATIONAL RACIAL PROFILING? .....	108
5.1 <i>The Contingent Moral Wrongness of Statistical Discrimination</i> .....	111
6. CONCLUSION .....	115
<b>REFERENCES</b> .....	<b>117</b>

# **1. Two Conceptions of Racism**

## **I. Introduction**

A great deal of confusion surrounds the correct content and extension of the terms “racism” and “racist.” When, exactly, are we warranted in calling someone a racist? Can only people be racist? What about symbols? Can a non-racist person have racist thoughts? Behaviors? What if someone says something racist— does that make *her* racist? What would it take to make a statement racist? And when, exactly, is an institution racist?

All these complexities could leave one wondering whether it is pointless to engage in the question at all. Why care if someone or something is racist? Perhaps we could all agree that some act is disrespectful or harmful, without having to further show that it, or the person who committed it, is racist. What does calling it racism add to our understanding of these people or institutions and their thoughts and actions? Maybe it is best to just avoid the term altogether, especially since labeling a person or an action racist often only serves to inflame tensions and shut down dialogue.

I think the question of whether or not some agent or action is racist, as opposed to disrespectful, harmful, or even vicious, is important (Mitchell). Finding out that an action one has taken or a belief one holds is racist should, and often does, fill one with a greater level of



dread than the discovery that one has been disrespectful or vicious. This is due in part to racism's sordid history; many of humanity's most despicable acts and widespread atrocities have been justified by racist attitudes and ideologies. Racism has motivated actions that reach far beyond normal, everyday injustices. It has been used to justify and motivate treating human beings as if they were not human.

Consequently, discovering that one is a racist or has acted in a racist manner is to find one has fallen in with a particularly terrible crowd. Though, like other moral wrongs, racism admits of degrees, recognizing its history gives us special reason to avoid it; Not only do we do not wish to be associated with those horrific people and their actions, but also, by endorsing racist thoughts and actions we open ourselves up to the possibility that we, too, could commit them.

Unfortunately, the seriousness of racism as a moral wrong tends to make it difficult to talk about at all, much less reach agreement on its content and extension. When discussing racism, there is a tendency to go towards one of two extremes—either everything is racist or nothing is. Both of these extremes lead to absurdities and both should be resisted. We risk trivializing the term if we use it too liberally, but it may well be true that more things are racist than we currently suspect. And, if that is the case, then people could end up expressing racism, and performing racist actions, while sincerely believing themselves to be non-racist. Due to the dangers associated with racism, we ought to be especially concerned to recognize when we are acting on it. This should spur us to get clear on its conditions.

Accounts of racism can be divided into two main camps. Most theories in the philosophical literature either assume or explicitly argue for and endorse what I will, for the

purposes of this paper, term the ‘agential’ conception of racism.<sup>1</sup> They use ‘racism’ to describe a quality that is found, in its most fundamental form, in individuals. Within this approach, debates are waged over whether the essential component of racist is attitudes, beliefs or behaviors, and exactly which attitudes, beliefs and behaviors are racist.<sup>2</sup>

Among sociologists and anti-racism activists, one often encounters a very different conception of racism, which is seemingly at odds with (or at least an addition to) the agential account. On this ‘structuralist’ account, the term ‘racism’ denotes a system in which advantages are divided along racial lines. Individuals and discreet institutions are racist to the extent that they participate in that structure. Those who participate by simply ‘going with the flow,’ and so reap the benefits of their privileged position in that society without actively endorsing the racial division, are ‘passive’ racists. Those who subscribe to positive attitudes towards or endorse the belief that the hierarchical relationship between racial groups by ought to obtain are ‘active’ racists.<sup>3</sup>

These two frameworks offer us very different starting points. The agentialist begins with individuals and explains all instances of racism in terms of their beliefs and attitudes. The structuralists begin with the structure as a whole. Though individuals occupy roles in the structure and some of them are likely racist, none of them need *be* racist (or *ever have been*

---

<sup>1</sup> Glasgow (2009) also uses the term ‘agential’ to describe one species of accounts of racism but in a somewhat different fashion. For Glasgow, to call something ‘agential racism’ is to mark it off from ‘institutional racism.’ I resist this usage because, as I will argue, there is an understanding of institutional racism such that it is also agential.

<sup>2</sup> For example, see Garcia (1996), Shelby (2002), Blum (2002), Glasgow (2009).

<sup>3</sup> While structural accounts of this sort have sometimes been referred to as ‘institutional’ accounts (for example, by Pierce, 2013) I think we ought to resist this move. First, I will argue that agential racism can also be institutional. Individuals’ attitudes and beliefs can move into and through institutions, both public and private, as they are drawn on by those creating policies, structuring organizations, dictating internal practices, and assessing costs. Second, on a structuralist picture discreet institutions are racist in the same way that individuals are. They either passively participate by upholding the system, dividing privileges along racial lines, or actively participate by expressing racist attitudes and beliefs.

racist) for the structure to be racially unjust. The structure can be racist even if *no agents hold or have ever held* racist attitudes or beliefs. Thus, we might think of them as bottom up and top down approaches.<sup>4</sup>

These conceptions also have different extensions; they diverge with respect to the phenomena they pick out as racist. In fact, the structuralist claims that the agentialist cannot pick out all instances of racism. There are certain uniquely structural racial injustices that the agentialist approach does not capture. Recognizing a central usage of “racism” as a synonym for “racial injustice” could leave us wondering whether the structuralist is right. By concentrating on the actions of individuals, perhaps the agential account does miss the bigger picture. If justice is not primarily a matter of the attitudes and beliefs of individual agents but, fundamentally, a quality that pertains to our institutions and the principles by which they are regulated, is it not possible that institutions could be racist or racially unjust even if no individual agent involved, in the past or presently, could rightly be termed ‘racist?’ In other words, could practices that are not properly thought of as the racist actions of any agent, institutional or otherwise, be racially unjust? This is the question that accounts of structural racism urge us to consider. And the question is important— If some (or most) of the problems that plague racially marginalized groups are unjust in this way, should we not revise our usage of the term ‘racist’ to cover, or even privilege, those harms? By casting our net too narrowly and refusing to invoke the moral reprobation that comes with calling a particular injustice ‘racist,’ we might allow for serious racial injustices to continue unchecked.

---

<sup>4</sup> Pierce (2013) refers to these as micro- and macro-analyses of racism. He writes that microanalyses or “individualistic approaches tend to favor moral, ethical or psychological methods, and [macro-analyses or] structuralist approaches tend to favor socio-political and economic methods” (p 1).

Building on this structuralist worry about the limited nature and scope of agential racism, my aim is to discover if there is anything that the structuralist can say about racial injustice that cannot be captured on the agential account. If there is, then there is at least one (albeit defeasible) reason to prefer the structuralist account to the agential one. Of course, the import of this additional structuralist insight would have to be weighed against the various advantages of the agential account before we could decide on the all-things-considered, best account of racism. However, if the structuralist account adds nothing to the content of the concept of racism, it would stand at a considerable disadvantage when stacked against the more familiar and intuitive agential conception.

In adjudicating this dispute, my first task is to describe what racial injustice amounts to on each of the two accounts. For the agential account, this requires showing how that theory might be extended to talk about institutions and institutional arrangements. On the structuralist account, where injustice is deducible from divisions in institutional power, this means showing how certain states of affairs, because of the unintentional but nevertheless non-accidental practices which give rise to them, can be said to be *racially* unjust rather than unjust for other reasons.

Using these accounts, I argue that the structuralist claim to possess some deeper insight about the nature of racial injustice, which cannot be captured by the agentialist, is false; So long as the agentialist agrees with the structuralist about the content of the concept of injustice and its scope, then she is able to express nearly everything that the structuralist can about racial injustice. The only difference is that the agentialist will not agree with the structuralist that a relation is unjust just because it non-accidentally divides power along racial lines even if it arises in such a way that no one could have reasonably predicted its

appearance. While the agentialist can easily argue that justice demands such power divisions be undone and the victims compensated (and that a failure to do so would be racist), the mere instantiation of such relationships, on the agentialist view, would not be a racial injustice.

I argue that this consequence of the agential view is, in fact, a point in its favor, as it leaves room for the possibility of accidents (human and non-human) that, although lamentable, are nevertheless, not unjust. This is pragmatically important because the appropriate response to instances of truly accidental harm is likely to be quite different from that of cases of intentional or negligent wrongdoing. Moreover, the agential view is a more flexible theory when it comes to describing institutional racism, compatible with a wide range of views about the nature and scope of justice.

## **2. The Two Conceptions of Racism**

Naturally, there are some points of agreement between structuralist and agential accounts of racism. Both, for example, acknowledge that there is some special social category called ‘race’ (biological fiction though it may be) into which people are often sorted and which can be related, in some important way, to ill treatment.<sup>5</sup> However, on the major questions of exactly who and what is racist and, most importantly, *why* it is racist, there appears to be significant and seemingly irresolvable disagreement between the two views.

---

<sup>5</sup> Hoyt (2012) writes that both conceptions are ways of pointing out or rejecting the “false and pernicious idea that people can and should be sorted into subgroups based on arbitrary phenotypic markers, assumed to have attributes that correspond with the markers, valued on the basis of the possession or lack of possession of the markers and treated differentially according to the valuations” (p. 226). This characterization of the commonality is somewhat too broad. It does not go any way towards specifying the specific sorts of phenotypical markers that we generally think of as “racial markers” and which, for the most part, are distinct from, for example, “sex markers.” There seems to be agreement among most proponents of these two views that race-based mistreatment can be picked out and separated from other forms of wrongdoing motivated by attitudes towards and beliefs about people with certain phenotypical features.

The following section will spell out exactly what that disagreement consists in, by exploring what each has to say about racial injustice.

I begin with the agential account because I suspect that, due in part to the fact that it takes its cues from ordinary language analysis, it is the more familiar and intuitive of the two. Whether or not we should trust those intuitions is, of course, a matter of some debate (though not one I will broach here).<sup>6</sup> In what follows, I provide a brief sketch of the agential account and demonstrate how, on this view, institutions can come to be racist. Assuming that justice is a matter of the arrangement of our basic institutions, then unless ‘racial injustice’ is not actually a synonym for ‘racism,’ by giving an account of agential institutional racism, I will have articulated a theory of agential racial injustice. It is this theory, the structuralist will argue, that is insufficient to capture all of the dimensions of racial injustice.

## **2.1 The Agential Conception of Racism**

As I indicated briefly in the introduction, the agential account of racism is, in its most basic form, concerned with the qualities possessed by individuals. Racism is a matter of an individual holding certain attitudes and/or beliefs.<sup>7</sup> I will not attempt to adjudicate the dispute over precisely which property (or properties) is (are) essential. The crucial point is that

---

<sup>6</sup> Tatum (1997) for example, in her discussion and critique of the agential account recounts the following about when she teaches the structuralist conception of racism: “Someone in the group is usually quick to point out that this is not the definition you will find in most dictionaries. I reply, “Who wrote the dictionary?”” (p. 9).

<sup>7</sup> I will leave to the side views on which racism is, most fundamentally, a matter of individual behavior since they introduce unnecessary confusion and, I think, have been adequately refuted. (Philips, 1984 is the most influential proponent of the behavioral view, but neither he nor any other theorists have defended it in recent years). However, I want to leave open the possibility of hybrid and disjunctive views such that either behaviors, or beliefs, or attitudes (or a combination any two) are necessary for racism.

whatever that property turns out to be, it must be such that it is evaluable according to the dictates of morality or rationality, or some other scale appropriately applied to agents.<sup>8</sup>

The agential theory can be extended to describe and evaluate institutions, and thus, racial injustice via what Joshua Glasgow (2009) has termed the ‘genetic thesis.’ According to this thesis, from racist agents, racist attitudes and/or beliefs can spread upwards and through the various institutions in which those agents participate and find expression in the structure and policies of those institutions. Take, for example, an instance of institutional racism that respects the genetic thesis:<sup>9</sup>

Imagine a town in which there are significant economic and social disparities between two racial groups (say, the ‘pinks’ and the ‘greens’) because the local government passed a series of laws that explicitly prevented one group (the pinks) from buying property in a particular area of town, where otherwise, the pinks would have liked to live in. The lawmakers imposed this restriction because, as greens, they disliked the pinks, found their skin color distasteful and took it as a sign of an underlying bad character. So, to avoid the pinks, green lawmakers bought land in one area of town along with other greens and passed legislation to prevent the pinks from living there. At some later date, the greens discovered that there was a large amount of oil under their land. They were paid huge sums of money by the local government for the right to extract this oil, making them all fabulously wealthy.

---

<sup>8</sup> On some of these views, it is important to calling the relevant attitude or belief ‘racist’ that the agent who possesses it stand in such a position that she could reasonably evaluate or revise it. For example, Garcia (1996) argues that a person who holds negative attitudes towards a particular racial group because she was raised in a racist culture and have never been exposed to any opportunities to revise those attitudes is ignorant, but not a racist. Shelby (2002) and Appiah (1990) both of whom defend the doxastic model, agree that racism cannot simply be a matter of holding false beliefs but that those beliefs must also be irrational. However, there are other agential accounts on which racism is a form of, or closely parallels, mental illness. On such a view, we might appropriately evaluate the content of the beliefs and attitudes (or compulsions and impulses) that make up ‘racism’ according to rationality or morality, without requiring that the agent herself be in a position to do so.

<sup>9</sup> This example is modified slightly from Glasgow’s original for reasons I will make clear in the next section.

The pinks didn't benefit at all from the discovery of the oil and though they enjoyed a reasonable standard of living via other means, they were not anywhere near as wealthy as the greens.

This, I take it, is a familiar example of institutional racism, in line with the way in which we normally understand why the legally sanctioned enslavement of black Americans and the Jim Crow laws in the American South were racist. The racist attitudes/beliefs of green individuals were inserted into and expressed by the institutional mechanisms of the town government. The genetic thesis explains why any direct consequences of those institutional expressions are racist. Just as individual behaviors motivated by racist attitudes and beliefs are racist, institutional behaviors that are so motivated are also racist. The discriminatory laws instituted by the greens on the basis of their racist attitudes towards or beliefs about the pinks prevented the pinks from owning land that they otherwise might own. Even if we assume that had the pinks, and not the greens, owned all the oil rich land then the government still would have paid the same amount to extract the oil, (in other words, the amount paid and its recipients was not an action dictated by any racist beliefs or attitudes), the current wealth disparity is a consequence of a initial racist restriction on property ownership. Thus, the disparity is not merely a racial one, but a racist one.

One of the essential aspects of the genetic thesis, which is captured by the above example, is that because institutions can outlast their creators, institutions with policies created or sustained by racists are capable of expressing racist beliefs and attitudes and carrying out racist acts even if no members of that institution are currently racist.<sup>10</sup> Once

---

<sup>10</sup> Glasgow (2009) argues that such cases, in which an institution perpetuates racism that is not held by any of its current members, are instances of "pure institution racism" and cannot be adequately accounted for on a reductive agent-based account. A reductive agent based account is "a time-slice thesis, which holds that for every institution at every moment it is racist, its racism is based in the racism of some agent at that moment" (p.



racist attitudes and beliefs have been introduced into an institution's policies or organizational structure, the institution can remain racist even if those individuals who created the racist policy are long since retired or dead.<sup>11</sup>

Agreeing to the genetic thesis, however, does not yet introduce the concept of institutions as anything over and above the joint actions of individuals. The genetic thesis could, it seems, equally apply to ordinary, non-institutional relationships. For example, suppose I ask you to do me a favor and, because we are friends and you have a general disposition to fulfill my requests, you do. As it turns out, the favor I ask you to perform is

---

74). By contrast, "the genetic thesis is a historical thesis, which holds that for every institution, at every moment it is racist, its racism at least partly originated in the racism of some agent at some point up to the moment at which the institution is racist" (p. 74). Glasgow's claim that an institution's racism must stem "at least partly" (not wholly?) from the racism of some agent is a puzzling one, and has led others, including Pierce (2013), to misunderstand Glasgow's view about the nature of institutional racism. On what basis might we condemn other parts of the institution (those that did not originate in a racist agent) as racist? Glasgow rejects the structuralist view that resources divided along racial lines can, in themselves, be racist. However, he writes, "even if the genetic thesis is correct, that is, even if rightly calling institutions 'unjust' or 'racist' is contingent on their having come about or being sustained in such-and-such a manner, we still have reason to deny that the injustice or racism themselves consist in the beliefs, attitudes, or behaviors of the individuals who once generated or sustained such institutions. Most directly, those who are otherwise engaged in a dispute over whether racism is subject to the genetic constraint can come to an independent agreement on what cases to recognize as cases of racism: they can agree that racism is not necessarily found in the agents themselves. And, most fundamentally, it is plausible that what makes the present state of affairs in Real Estate objectionable is not that long-dead generations had racist attitudes that led to it but that it itself is racist. We don't look at a society like that in Real Estate and object that the ancestors who set things in motion were racist but that it is, at present, perpetuating unjust racial inequality. Thus it seems that institutional racism can be present at t without any agents who interact with the institution at t being responsible for that racism" (p. 76). I think the most plausible interpretation of this statement by Glasgow (2009) and his earlier claim that the past attitudes or beliefs are only partly responsible for the racism, is that the institutional reaction to past injustice is itself racist—it disrespects the victims of the earlier racial injustice and compounds their harm. Institutions, then, must be apt for moral condemnation, as Glasgow argues that "racism" is always a term of moral disapproval. Exactly how we should think about the moral character of institutions is not a question that Glasgow explores in much detail. Are institutional reactions the expressions of present attitudes of disrespect or disregard on the part of individuals or some attitude that is not reducible to the individuals who make up the institutions? He seems to think that laws can express attitudes in a manner that is not reducible to the attitudes of the lawmakers, writing, "prochoice advocates hardly seem to betray conceptual confusion when they assert that antiabortion laws fail to respect a woman's right to choose..." (p. 83) and "Antimiscegenation laws failed to respect rights of intimacy and love and the people whose rights were at stake" (p. 83). He also claims that societies can be disrespectful but admits that this way of talking could be shorthand for saying something about the authority figures in a society or some number of institutions taken together.

<sup>11</sup> The agential account, more so than the structural one, preserves certain features of the initial formulation of the term, as introduced by Carmichael (Ture) & Hamilton (1967), who argued that that institutions are racist when their policies and actions are *intended*, often through subtle and covert means, to subordinate one racial group to another.

harmful to blacks and I have requested it because of my anti-black racist attitudes. Even if you are unaware that the action you perform is harmful and motivated by my bad attitudes towards black people, it still seems right to describe the action as racist (though we ought not describe you as such, on the agential account, because you lack the relevant beliefs and attitudes). The same would be true if I asked you to do me a favor but died before you carried it out.<sup>12</sup> Thus, it seems that there may be instances in which the genetic thesis could be applied fruitfully to describe why some individual actions are racist, although the individual who performs them lacks racist attitudes or beliefs. Nevertheless, it does not follow that the actions he performs are properly described as institutional.<sup>13</sup>

A full account of agential, institutional racism requires not only an account of the pipeline by which institutions can come to express racism (a fully developed version of the genetic thesis that includes the conditions on which it is applicable), but also an account of what makes an instance of the genetic thesis *institutional*. This raises a host of complicated questions about the nature of institutions themselves—What structure, aims and policies are properly thought of as institutional? Must those aims be explicit and those policies codified? Can violations of those policies, if they are widely tolerated, be thought of as institutional

---

<sup>12</sup> One question that could arise in these cases is whether or not the action is properly thought of as belonging to the person who requested it or the person who carried it out. If the action belongs to the former, then it might not be a true case of the genetic thesis, since the racism has not transferred from one entity (capable of performing actions) to another.

<sup>13</sup> In other words, in order to differentiate between these cases of action at a distance by individuals and what we would want to call “institutional” racism, we would need some additional criteria that mark off institutions and institutional relationships from relationships among individuals. That they are candidates for applications of the genetic thesis is not enough. Instead, a satisfactory account must provide criteria for distinguishing certain joint actions from other, interpersonal, non-institutional joint actions. Possibilities include functional or teleological accounts of institutions, as well as group agency accounts. Spelling out exactly what features are necessary and sufficient for institutional, as opposed to individual or interpersonal action, is a question I leave to the side, though it is an important component of any fully fleshed out agential account of institutional racism.

actions? Must institutions possess agency in order for their actions to express racism?<sup>14</sup> Or is it sufficient that an institution be the mouthpiece of a particular individual in order to be thought of as expressing racism?

I will not address these questions here. So long as some form of joint action, or action through others is possible, then agential racism is possible that is not purely individual. We can define institutions and institutional actions according to some criteria such that it does not require that an institution *do* or *be* anything beyond an entity that performs particular actions (perhaps those aimed at a certain set of goals). For my purposes, the only claim the reader need accept is that if anything counts as an institution then surely a state does. I take the inclusion of the state as an institution to be a desideratum of the correct account of institutions. So, we can call the actions of a state ‘institutional actions’ even if that means nothing more than an individual expressing his will through the coordinated behavior of others.

Finally, before turning to the structuralist view, I want to briefly address how an agential conception of racism might account for the phenomenon of systemic racism. We certainly want the ability to acknowledge on the agential view that our institutions, like individuals, might interact with one another and that there may be certain situations in which talking about “institutional racism” is insufficient to describe the widespread racist attitudes

---

<sup>14</sup> This discussion of the conditions of institutionalization of racism raises another possibility for when institutions can be racist on the agential account that does not depend on the genetic thesis. It is possible that some institutions are not merely aggregates of individuals acting in conjunction with one another, but actually possess agency of their own that is not reducible to the agency of those involved. See List & Pettit (2011). Thus, it is possible that a group of individuals who are non-racist might come together to form an institution, or group agent, and though, as individuals, they harbor no racist attitudes or beliefs, they together create a racist group agent. I don’t argue here that group agency of this sort is, in fact, possible or spell out the conditions under which it might obtain. Instead, I just wish to highlight that if group agency is possible, then this seems like a distinct sort of institutional racism that would be agential, but that racism would not be captured by the genetic thesis.

and beliefs in a particular society and their pervasive expression in institutional policies. For example, it seems right to say that, at least prior to the Civil Rights Act, anti-black racism in the U.S. was not merely institutional but systemic, given the large number of institutions involved in the perpetuation of racist attitudes and beliefs. On the agential account then, perhaps a system is racist when the majority of the institutions of that make up that system are racist. This is a description of an aggregate of agential actions and should not be thought of as picking out some entity 'x' that is a system and to which we can ascribe moral properties. Consequently, if institutions are agents, then 'systemic racism' is a way of speaking about an aggregate of institutional actions, calling each of them racist and merely noticing that 'there appear to be a lot of racist institutions around here'. If institutions are not agents then it is another way of assigning the term 'racist' to an aggregate of individual agents who interact in (independently identifiable) institutions and who have managed to enact individual racism in those institutions.<sup>15</sup>

## **2.2 The Structuralist Conception of Racism: Three Interpretations**

Having articulated the agential view, including its individual and institutional expressions (and, potentially, its extension to systems), I turn now to the structuralist approach. My aim here is to present the most compelling version of this view, which takes systems or structures as the proper object of the term 'racist,' and the racism of institutions and individuals to be, at least in some instances, derivative from those racist systems of which they are members. To accomplish this, I first want to discuss and dispense with two

---

<sup>15</sup> One interesting question that arises in describing individuals and institutions as racist is whether or not we can call someone or something racist when it has no responsibility, morally or epistemically, to change. This worry arises particularly in cases where the aggregate of individual actions, each of which is fairly innocuous on its own and creates, we might think, little or no responsibility to change, can have devastating consequences when a large number of people perform it.

common views that, though often understood as structural accounts of racism, collapse into an agential account on further analysis.

*Structural Account #1:  $P+P=R$*

At times, a view quite similar to the one I will articulate here is billed as a structuralist-style alternative to agential accounts of racism. This is the “ $P+P=R$ ” (power + prejudice = racism) theory. This formulation was first penned by Pat Bidol (1970) and has since been adopted by countless antiracist activists.<sup>16</sup> It is useful in introducing people to the differential impact of prejudice when it is accompanied by power. Insisting that racism is always accompanied by power makes abundantly clear that the harms that a prejudiced group can cause when they control institutions are, in general, far more severe and far-reaching than those a prejudiced minority are capable of enacting. This, in turn, helps make sense of why we often judge the prejudiced attitudes of a white person more severely than those of a black person.<sup>17</sup> However, it places too much emphasis on personal prejudice (individual attitudes

---

<sup>16</sup>Bidol (1970)’s view is summed up in a manual by the National Education Association (1973). The manual states, “When the racial prejudice of the majority group in a society is reinforced by the culture and institutions of that society, then the prejudice becomes racism...In the United States at present, only whites can be racists, since whites dominate and control the institutions that create and enforce American cultural norms and values” (p. 12) It continues, “In our society it would be very difficult for any individual to be racist all by himself. If everyone just had his own individual prejudices, which were never reinforced by the society, racism would not exist. But when those individual racial prejudices are reinforced by the culture, we have institutional racism. Thus, all white individuals in our society are racists. Even if a white American is totally free from all conscious racial prejudices, he remains a racist, for he receives benefits distributed by a white racist society through its institutions. Our institutional and cultural processes are arranged as to automatically benefit whites, just because they are white” (p. 13) The manual goes on to further distinguish between three types or levels of racism, individual, institutional and cultural. Institutional racism exists when “the social systems of the culture [are] rigged in some way to distribute the benefits differentially and preferentially to whites” (p. 14). Cultural racism, on the other hand, seems to be largely a matter of having white bodies as the assumed norm for all fictional characters, religious figures, etc. (p. 15).

<sup>17</sup> Blum (2002) rejects the  $P+P=R$  account and argues, “Power does play a role in the moral seriousness of instances of racism, but not in its existence” (p. 39). He continues, “...I contend that the racial identity of the victim (and, to a lesser extent, the perpetrator) is indeed morally pertinent. Hatred of blacks against whites is not a primary form of racism; hatred of whites against blacks is. Some forms of racism are central and paradigmatic, others secondary. The former have defined for us what racism is. They are more directly tied to the rationale...for the intense moral opprobrium carried by the term ‘racism.’ That rationale involved oppression, hatred, and discrimination against people of color, and most especially blacks and Native

and beliefs) to be considered truly structural. Though structures play a role in this account by giving context to individual prejudice, they are not the primary explanatory features.

Individuals co-opt or create structures and use those structures to enact and enforce racist policies. Whites who benefit from those structures, though they may not be consciously prejudiced, are considered racist not simply because they receive advantages from them, but because they are complicit in institutions that continue to express racist attitudes, even if all those who possessed the relevant attitudes are dead.

However, it is useful to note that many of the same criticisms that are often leveled against the structuralist view are also engaged to critique the  $P+P=R$  formula. In particular, some critique the view for failing to respect the ordinary usage of the term, ‘racism,’ and, consequently, ostracizing whites who feel as though they are unfairly targeted by this conception. The power condition is generally taken to mean that blacks, who lack the relevant power, cannot be racist. A white and a black individual might say the very same thing, on this account, and the former would be racist while the latter would be merely ‘prejudiced.’ Thus, apart from the well-known difficulty in making precise the “power” condition such that all and only whites possess power sufficient to make their prejudice racist, the  $P+P=R$  account is unapologetically revisionist. In general, a black person who expresses negative attitudes towards whites because they are white is thought to be racist. But, accepting that this and all other structuralist accounts are revisionist, I will make no appeals to ordinary language in my critique. Though an account that respects our ordinary

---

Americans, by whites, not the reverse. Everything else being equal, greater moral opprobrium rightly attaches to racism by whites against people of color than the reverse. This is the most important moral asymmetry in racism”(p. 43-44). This strikes me as a more plausible justification for our divergent judgments in such cases but it is an agential, not a structural, account.

language usage of the term might be preferable in some small way, it is far more important that our account be capable of capturing everything we agree, upon reflection, is racist.

*Structural Account #2: Racism without Prejudice*

A second structuralist approach, similar in some ways to the first, was first articulated by David Wellman in his influential work, *Portraits of White Racism* (1977, revised 1993). Since then, social theorists have adopted and rearticulated it and it has emerged as one of the dominant, if not the dominant account, of racism in the sociological literature.<sup>18</sup> In fact, P+P+R is likely a less sophisticated species of this view. As such, Wellman's structuralism is vulnerable to many, though not all, the same critiques as the aforementioned approach. However, it is useful to examine it separately as it has several advantages over its more simplistic formation.

According to Wellman (1993), racism is the "culturally sanctioned beliefs, which, regardless of intentions involved, defend the advantages whites have because of the subordinated position of racial minorities," (xi). Tatum (1997) expounding on Wellman's view writes that, "racism is a system of advantage based on race" (p. 127). The system, she argues, is made up of cultural messages, institutional policies and practices, beliefs and actions of individuals. Racism is present whenever it is better to be white in American society, which given the existence of pervasive and persistent racial disparities, it nearly always is.

---

<sup>18</sup> See, for example, Pinderhughes (1989); Tatum (1997); Adams, et al. (1997). In his review of Elizabeth Anderson's "The Imperative of Integration," (2010) sociologist Michael Emerson (2011) chastises Anderson for wading into the "somewhat tired debate of what and who is racist" arguing that her use of the term "seems to ignore the extensive social science work devoted to understanding and defining the term—such as the misuse of collective power that results in diminished opportunities for some racial groups (p. 318)."

What is essential to Wellman's theory is that race plays an explanatory role in the position that individuals occupy within American society, but it does so for reasons other than prejudicial beliefs about various racial groups. Blacks are, by and large, deprived of goods and services that white Americans enjoy and that denial is defended in terms that make no reference to racial superiority or inferiority. Unlike past justifications for the white privilege, a white person who defends her advantages does so without prejudicial intent. Instead, writes Wellman (1993), "The terms in which middle-class professionals defend traditional institutional arrangements are, strictly speaking, not examples of racial prejudice. They are neither overtly racial nor, given these people's interests, misinterpretations of 'facts.' However, while the sentiments may not be prejudiced, they justify arrangements that in effect, if not in intent, maintain the status quo and thereby keep blacks in subordinate positions" (p. 33).

So, without supposing prejudicial intent, how exactly, does one's race work to explain an individual's position in American society? Take a familiar example: Disparities in housing between whites and blacks are, in some cases, created and sustained by the mortgage industry practice of awarding home loans on the basis of one's credit. If blacks tend to have lower credit scores than whites, then they will struggle to secure home loans. A white person who receives a loan when a black person does not could cite her higher credit score and argue that, in general, the practice of awarding home loans in this way is economically efficacious. She need not express any bad attitudes towards blacks to defend her privileged position. The structure affords her that privilege, on the basis of her membership in the white racial group, and she acts rationally and without malice when she



defends it. The system she participates in is racist, in that it advantages her because she is white, even if she has no racist intent when she secures her loan or defends her privilege.

Though this is one possible interpretation of the structuralist view (and quite likely the one that Wellman was attempting to articulate), it is relatively uninteresting. It trades heavily on the notion of conscious prejudicial intent and, unlike a truly structural approach, agents are required for racism. Most Americans, one would hope, are sufficiently aware of some of the devastating consequences of past anti-black racism that any defense they might launch of policies that maintain racial disparities would be an exercise in willful ignorance. Whites occupy an advantaged social position vis a vis blacks because of centuries of uncompensated harms perpetuated against blacks. Even postulating that the advantages whites enjoy are not the result of discrimination against blacks, the current condition of the black population is most assuredly due, at least in part, to past racial injustice. So though the mortgage industry's practice might be perfectly sound on its face, whites who defend it and the advantages they secure from it are disregarding the rights of black citizens. They are failing to acknowledge that flagrant violations of blacks' rights led to their lower credit scores and reduced options now.

Imagine I have good reason to suspect that your pen was stolen before class and you were denied or unfairly discriminated against in all of your attempts to procure a new pen. I then insist that my having a pen for class entitles me to certain privileges, like the ability to take a test. You, I assert, ought not be permitted to take that test because you failed to bring a pen to class. There is little question, in this scenario, that I have acted badly. Even if I insist I have nothing against you, in particular (I am just defending the classroom pen rules, which are a generally fair way of solving pen-related problems), I act callously towards you when I

refuse to acknowledge the hardships that you have endured and how they affected your pen-less status. I may not have stolen your pen, or limited your ability to procure another, but I can certainly be accused of failing to care, or care enough, about you. The same could be true of Wellman's modern racist. Though she expresses egalitarian views, she refuses to acknowledge how past injustices have rendered what would otherwise be egalitarian or colorblind policies, racist.

If this is the correct interpretation, then structural racism simply collapses into an agential account, though the agents may not have explicit anti-black attitudes or beliefs. Whites are racist in that they exercise a willful disregard for blacks when they defend the advantages they enjoy because of past racism. They need not express racist beliefs because defending current colorblind practices, given the lack of sufficient compensation for past racial injustice, is sufficient to maintain white privileges. But, it is also a way of disrespecting blacks, so that defending those practices without acknowledging the suffering of black people is still racist.<sup>19</sup>

As I said, this is a relatively uninteresting way of understanding Wellman (1993) and Tatum (1997)'s structuralist account, even if it turns out to be the correct interpretation. A second reading is more sophisticated, and it might be what the whites Wellman describes would claim, were they pressed to think more deeply about the source of their advantages.

---

<sup>19</sup> It is possible that some white Americans who defend their privilege in securing a home loan, and the mortgage companies who use this metric, are unaware that whites' higher (and blacks' lower) credit scores are a result of past racial injustice. Those people, one might argue, are truly without prejudicial intent when they defend their advantages. They articulate a "what's mine is mine" defense of the advantages they possess because they do not even recognize them as advantages. They think that what they have, they earned and what blacks lack, they deserve. Quite likely, those who are not sufficiently aware of the history of anti-black racism in the U.S. to draw inferences about the fairness of the credit industry are morally culpable for that ignorance. However, that some may not be does not erase agential wrongdoing from the picture entirely. That atrocities are swept under the rug does not erase responsibility. Thus, the state might still be at fault for failing to make information about the wrongdoings committed against blacks, and their lingering effects, more widely available.

On this interpretation of the structuralist account, the past injustices against blacks are recognized, their lingering effects are acknowledged, but white advantages are defended on the grounds that no white person now, nor any state or private industry, is responsible for those injustices. Nor, it is claimed, could they bear that responsibility without incurring an unacceptable level of harm. If this is the case, then the mortgage industry could perpetuate racial disparities through their credit score based loan policy without harboring any racist attitudes. The agents involved may even sincerely regret that the practice produces bad outcomes for blacks, but they realize that, if they were to revise their practices and offer loans to blacks with lower credit scores, many companies would go out of business. They cannot be held responsible for rectifying injustices that they did not commit, especially if they would suffer devastating consequences were they to attempt to.<sup>20</sup>

---

<sup>20</sup> It is these considerations that led me to change the initial intuition pump I offered from Glasgow (2009)'s initial formulation. In the original case, Glasgow argues that, intuitively, a mortgage industry is racist if it offers mortgages to those with better credit ratings in a society that, several hundred years in the past, supported a system of state sponsored slavery for which reparations were never made. Many of us (though certainly not all) would probably be inclined to accuse the state of racism for any "race-neutral" policies they enacted, given its past actions. It is not at all clear that the mortgage industry (or rather, the particular institutions that make up that industry) could be so accused unless they too were part of the original discrimination. Clearly, the mortgage industry is exacerbating racial inequalities caused by racism and we might object to their practice on this basis. However, were we to criticize the companies for failing to consider this preexisting disparity when crafting their policies, we would still have to present an argument for why this failure makes the company policy racist. In other words, how is such an action indicative of the agent who created it having beliefs and/or attitudes that are racist? My worry is that without such an argument, a failure to draw a distinction between institutions which are actually racist and those that are conducting "business as usual" in a society with a racist past could lead to the absurd conclusion that racial injustice in a society's past makes all current institutions unjust which exacerbate racial disparities. For example, a Reconstruction Era black family's practice of passing on their traditions and accumulated knowledge could, given that no members of the family received any formal schooling, exacerbate racial disparities by making their children less well informed than white children whose families also pass on their traditions and knowledge. But we would not want to say that the black family was a racist institution. To do so would deny the very aspect of racism to which the genetic thesis draws our attention; namely, there must be a special sort of causal relationship between a racial disparity and an agent's racism, and that racism must be taken up by (or inserted into) an institution in order for that institution to be considered racist. Thus, the agentalist must be prepared to provide an account of institutional uptake (when an institution is a possible purveyor of racism), what it takes for an institution to express racism (institutional responsibility) and what our collective responsibility is to regulate against institution racism.

Similarly, whites who receive the loans need not cite racial superiority as grounding their claim to money. They need only refer to their higher credit score and the fact that they cannot be held responsible for injustices that they did not commit. They would also deny that they have any responsibility to aid whites who are poor because of past injustices. Furthermore, though they may regret that the system is set up to favor them, they would be much worse off were they not to take advantage of the home loans that their credit score affords them.

I will revisit this interpretation in the following section. I think, in fact, it is the correct diagnosis of how it is that there can appear to be racism without racist agents. When structuralists claim that an agential account cannot make sense of some racial disparity, what they actually mean is that a narrow individualist picture of moral responsibility cannot. It may well be true that neither the mortgage industry nor any individual who secures a loan is at fault for current racial disparities in housing. But it does not follow that the racism is not agential. First, agents were responsible for perpetuating injustices against blacks and one consequence of those injustices is that blacks today have lower credit scores. Though the mortgage industry may not be at fault for using credit scores, the current state of black homeownership (including, for example, the homes blacks own, do not own, and the geographic locations of those homes) express past racism. They are the product of racist attitudes and bear those attitudes in their continued existence. So, for example, the existence of the almost exclusively black North Lawndale neighborhood of Chicago, with a homicide rate triple the national average and poverty rate twice that of the rest of Chicago, is an expression of racism (Coates, 2014). It was created by the discriminatory practice of ‘redlining’ and so is a direct consequence of racist attitudes.

Second, if we think, that justice is not only about our current actions, but also a matter of taking into account how past agents, acting wrongly, may have affected current conditions, then it is not the case that the practice today is devoid of agential racism. Perhaps, for example, the state is responsible for allowing past injustices to continue to affect blacks' ability to secure home loans. Alternatively, it may be that we are all, as a collective, responsible (in a forward-looking sense) for creating a more just system that will take those past injustices into account. That no single agent, be it a corporation or an individual, is responsible does not rule out the possibility that as a collective, we may be.

So, why, if a sophisticated theory of collective responsibility (either backward or forward-looking) for past injustice would explain such cases, are structuralists instead committed to developing an account that does not appeal to agents and agential wrongdoing, individual or collective? Perhaps the move towards structuralism is purely pragmatic. The structuralist recognizes that there are a huge number of potential problems involved in tracing current racial inequalities back to their source in the attitudes and beliefs of agents and so recommends revising the term. That is the best way, they might think, of addressing the pressing issues of the American black population.

They may be right. Given that the culpable agents are likely long since dead or dissolved, the counterfactuals needed to prove culpability are likely to be extremely difficult to evaluate. Furthermore, even if that issue were resolvable, there is little chance that the subsequent task of convincing present Americans to atone for past harms would succeed. White Americans would continue to defend their privilege despite the fact that it is the consequence of brutal racism against blacks. They might continue to insist that because they were not responsible for that racism, they are under no obligation to give up the many

advantages that they currently enjoy. Collective responsibility theories are controversial and especially unlikely to be accepted by those who occupy a position of privilege. Thus, frustrated with the practical impossibilities of this project, the structuralist might decide that the more politically efficacious move in the fight to end racism is to redefine the term itself to attack its consequences—namely, those advantages that whites have because they are white. Seeing the obvious injustice in a society where black people consistently end up on the bottom, structuralists might have become frustrated in the attempt to show that deciding home loans purely on the basis of credit scores is, in fact, racist and so move to an account that avoids the complicated task of assigning motivations to persons long dead.

But a pragmatic decision is not the only potential explanation of the structuralist move. Though it might accurately describe some structuralists' motivations, it assumes that the agential account is basically the right way to understand how racial disparities come about. As such, it is not capable of truly capturing the structuralists' basic contention that at least some racism can only be explained in terms of systems. However, there is a third, and final, structural account that offers a purely (though not exclusively) structural account of racism. I turn now to Sally Haslanger (2004)'s account of racial oppression.

### *Structural Account #3: Unintentional, Non-Accidental Racial Harms*

According to Haslanger (2004), structural racism (or, as she prefers to call it, 'racial oppression') is "crucially concerned with power" (p. 98).<sup>21</sup> In instances of structural racism,

---

<sup>21</sup> Though Haslanger (2004) declines to label her account as one of 'racism,' preferring to refer to the phenomenon as 'racial oppression,' the position she articulates is a clear account of racial injustice. As she writes, "In cases of structural oppression, the focus is on our collective arrangements—our institutions, policies, and practices..." (Haslanger, 2004, p. 100-101). The examples she gives range from Jim Crow laws to cultural norms. As I am assuming that anything that is racially unjust is also racist, I will consider her position an account of racism. And, since I am not relying on an ordinary language critique of her view, she has nothing to fear from this assumption. Thus, for the remainder of this paper, I will treat Haslanger's formulation as a reliable and sophisticated articulation of a structuralist account of racism.

the harm to individuals or groups is a result of unjust arrangements of institutions that misallocate power and so doing, fail to accord human beings their proper moral status. For example, regardless of the benevolence of the dictator or her epistemic abilities, one might claim, as Haslanger (2004) does, that an absolute dictatorship is unjust in itself. She writes, Tyranny is wrong not because (or just because) tyrants are immoral people intentionally causing harm to others but because a tyrannical governmental structure is unjust. Theorists will vary on what exactly constitutes its injustice, but key considerations include such matters as the fact that tyranny is not a structure in which individuals count as moral equals. (p. 100) In other words, given some understanding of human beings' moral status, this just isn't the sort of relation that one person should stand in to another. To the extent that the dictator uses his position to cause harm to his citizens, he is guilty of abuse of power but the structural wrong exists regardless of how he chooses to act (Haslanger, 2004, p. 100).

Some structural wrongs are clearly the product of the intentional actions of agents. Jim Crow laws, for example, are unjust institutional arrangements that are the expression of agents' racist attitudes and beliefs. But some, Haslanger (2004) argues, are brought about by social and cultural practices for which no agent, either group or individual, is responsible. They are unorganized, uncoordinated, collective endeavors that include everything from transportation and administration to our rules of etiquette and linguistics (Haslanger, 2004, p. 105). Nevertheless, they result in power relations that are unjust in themselves. As Haslanger (2004) writes, "...the oppressive structures in question may be intentionally created or not...those responsible for the structure may even be acting benevolently and with the best information available" (p. 103). As such, there appear to be a number, perhaps a quite large

number, of practices that bear illegitimate power relationships without any agent having behaved wrongly in creating or sustaining the practices.<sup>22</sup>

Some of those practices, Haslanger (2004) argues, non-accidentally (but unintentionally) divide that power along racial lines (p. 114). Haslanger (2004) proposes that a non-accidental correlation can be established counterfactually. For a group to be oppressed as a group, being a member of that group must be causally relevant to the oppression suffered (p. 114). Take causal relevance to be a matter of close possible worlds; members of some group G are oppressed as Gs if, in the closest possible world where the current members of group G are non-Gs, they would not be harmed by the unjust policy. Thus, black men are oppressed as black men by racial profiling if, in the closest possible world in which they are white men, racial profiling would not harm them. If we are tempted to refer to those practices as racially unjust, it seems we must admit that there are racial injustices that arise without racist agents.

Consider the case of social norms: In many places in the Reconstruction era American South, blacks were expected to address whites by their titles and last names, while whites addressed blacks by their first names. This expectation, an instance of an illegitimate power relation between whites and blacks, might have been expressed to a greater or lesser degree by any particular individual (and as such, we may hold them more or less blameworthy for its maintenance) but need not have been the intentional creation any particular racist individual. It is, if not historically accurate, at least conceivable that this norm arose not because some racist white wanted to assert her power over blacks but

---

<sup>22</sup> As Haslanger (2004) puts it, “If power resides in the relationships created by practices, and no individual agent is responsible for a particular practice, then there is an important sense in which the distribution of power may be unjust and yet the injustice not be properly explicated in terms of an agent’s wrongdoing” (p. 104).



because, while enslaved, many blacks had no last name. During slavery, even an anti-racist white person who wanted to demonstrate respect in addressing an enslaved black person would have had no choice but to refer to her by her first name. For the enslaved black person to show the same respect to a white individual, however, would have involved calling her by her title and last name. Once firmly established, this practice might have continued after emancipation, when freed blacks acquired surnames. Though the whites who perpetuated it, particularly those who actively enforced it, might have been blameworthy for maintaining it, they were not responsible for creating the practice with its corresponding power relation.

### **3. Extending the Agential Account**

If Haslanger (2004) is correct and racial injustices, in the form of illegitimate and non-accidental power differentials, can arise without racist agents, then the structuralist has a distinct advantage over the agentialist. Agentialists can explain why the power relationship instantiated by something like Jim Crow laws is racist but seems at a loss to explain how a power relations stemming from uncoordinated, collective activities could be racist. With respect to the former, they could argue that Jim Crow laws were the product and expression of objectionable racist attitudes and beliefs of U.S. lawmakers and those laws, once made, could continue to express those attitudes and beliefs even if no one in power currently agreed with them. But, with respect to the latter, what could they say? Agentialism, as the name suggest, requires agents. More specifically, it requires that those agents be the bearers of racism and express that racism through the practices they engage in and the power relationships to which those practices give rise. However, in this case the racism is in a practice that is not properly thought of as the action of any agent or agents.

This would seem to be a strong point in favor of the structuralist account. After all, I mentioned at the outset that if there is some aspect of racism that the agentialist simply cannot capture or capture as well as the structuralist, but the structuralist can capture all of the phenomena that the agentialist can, then we have reason to embrace the structuralist account.<sup>23</sup> But, if the agentialist can also say everything that can be said by the structuralist, then we have no content-based reason to prefer the structuralist account, and we will have to move to other methods for evaluating the two. We might consider pragmatic reasons (which Haslanger (2004) argues point in favor of a structural approach but which I will argue favor an agential one) or how we use the term in ordinary language. Thus, a form of racism that arises without agents would seem to be exactly what the structuralist account needs to gain an advantage over the more familiar agential account.

Unfortunately for the structuralist, I think we should not be too quick to discount the agential view. For the structuralist to assert that certain power relations are unjust she must make two claims: First, she must claim something about the nature of justice, namely, there are some relationships that are inconsistent with the moral status of human beings and so, oppressive and wrong in themselves. Second, she must claim something about the scope of justice, namely that it can be applied to the uncoordinated and collective activities of human beings. I will address these claims in turn, arguing first, that oppressive relationships can only arise as a result of intentional human activity. Second, I argue that even if we were to ignore that argument, extending the scope of justice beyond the reach of intentional human activity would make it impossible for us to preserve a distinction between injustice and unfortunate accidents.

---

<sup>23</sup> Or, as Haslanger (2004) puts it, “a theoretical reason to reject the [agential] approach is that it cannot account for some forms of injustice for which no [agent] is responsible” (p. 105).

Glasgow (2009) attempts to resist the structuralist conclusion on intuitive grounds by presenting the case of a natural accident that gives rise to power relation between two parties. One party is, through no one's intention or negligent action and unbeknownst to all parties involved, completely at the mercy of the other. Glasgow (2009) writes,

Accident. At time  $t$ , a society is arranged such that one person has almost all of the wealth and power, while the other million people live in dire poverty with no power. This distribution fails to maximize utility or real freedom, was not arrived at through free transfers, can be reasonably rejected by just about everyone, and fails to conform to principles that would be chosen from a fair initial position. (In short, the distribution fails to satisfy any remotely plausible theory of distributive justice.) It arose when a freak accident killed off everyone except those trapped in a mine plus its one bunker-protected media baron, who alone can open the mine. The media baron has no way of knowing that the accident spared but trapped those in the mine, so he does not know that he has all the power (p. 75).

Given that, on the structuralist account, we are concerned primarily or crucially with power distributions, and in this scenario, the power lies completely in the hands of one individual, the situation described is potentially unjust. To determine with certainty whether or not it is actually unjust, we need to discover if it causes the trapped individuals unjustified harm—are they disabled or disempowered by the institutional arrangements? Is the power arrangement inconsistent with respect to the moral status of the million? Glasgow argues that even if we answer those questions affirmatively, it is hard to see this as a case of injustice rather than as an unfortunate accident. However, he admits that since 'racism' is a term that picks out not only individual moral failings but also social injustices, if we assent to the structuralist notion that certain relationships are unjust in themselves, then the distribution of power described might be morally condemnable (in the sense of unjust) even if no agent is condemnable (in the sense of morally culpable). The million are oppressed by the one.

That conclusion, however, is mistaken. Haslanger (2004) claims that just as dictatorship is wrong in itself or inherently oppressive so, too, are the illegitimate power

relations created between races by the unintentional, uncollected actions of individuals. But what, exactly, is an illegitimate power relationship? Presumably, not all inequalities in power are illegitimate. A doctor has more power than a patient on the operating table but that relationship does not seem oppressive in itself. Illegitimate power relations are at odds with our status as moral agents or violate our human dignity. What sort of power relation might, apart from any harm they cause or are likely to cause, violate our human dignity? Building on Haslanger (2004)'s example of the absolute dictator, one plausible candidate is the relationship of domination.

Recent literature in the republican tradition has done much to explicate the concept of domination. As Philip Pettit (1997) understands it, domination occurs when one party has “the capacity to interfere on an arbitrary basis in some of the other's choices” (p. 52). Apart from any harm that it causes, domination is wrong in itself because it is a violation of our status as free agents. Thus, an absolute dictator is, by definition, oppressive of his subjects regardless of how he acts. He always has the capacity to arbitrarily interfere with their choices.

We need not agree with Pettit's particular version of domination but we can use it to illustrate an important point about oppressive power relations—namely, for some relationship to be a power relationship, illegitimate or not, it requires more than inequality, even when those inequalities are extreme. Just as illegitimate power relationships require a specific sort of inequality (one that violates our human dignity), power, too, must be of a certain sort to separate it from mere potential. Pettit (1997) calls it a ‘capacity,’ but we could also term it a ‘readiness.’ He writes,

...the capacity to interfere must be an actual capacity, as we might call it—a capacity that is more or less ready to be exercised—not a capacity that is yet to be fully

developed: not anything like the virtual capacity of the musically gifted person who has yet to try out the piano. Consider a collection of people who, if they were to constitute themselves as a coherent agent, would have a ready capacity to interfere with someone. Or consider the agent, personal or corporate, who would have such a capacity, did they only recognize the presence of the potential victim, or the availability of causal modes of contact. In such cases there is only a virtual capacity to interfere, not an actual capacity, and I shall not say that there is domination. There is virtual domination, we might say, but not actual domination. Virtual domination may be something for republicans to guard against, of course, because of the future dangers it represents. But it does not yet constitute the central evil to which they are opposed (p. 55-56).

Unlike a dictator, the media baron in the bunker lacks the capacity to interfere with the trapped million because he is totally unaware that he possesses that power. That this capacity be *actual*, and not merely *potential*, is an essential aspect of illegitimate power relations, even if they are understood outside of Pettit (1997)'s framework of domination. Without this distinction, oppression would result whenever any members of society could bring about an illegitimate power relationship. But this result would be absurd. I could, perhaps, if I crafted a careful plan, spent years accumulating resources, and got very lucky, come to dominate some other person or persons, but I do not thereby oppress them.

Consider the following case: Suppose that the CDC were to develop a new vaccine that protects the population against a deadly disease but has the completely unintended and unexpected side effect of making everyone inoculated slightly more racist against blacks. The government mandates the vaccine without first testing it for this effect because no vaccine had ever been known to have such an effect. Even if the agentalist and the structuralist agree that this situation created a harmful power relation that was non-accidentally divided along racial lines, the agentalist could not call the relationship itself racist. Like the media baron, the white population may now have a virtual or potential capacity for domination but, if they do not recognize blacks as potential victims and take no

steps to use this unfortunate event to their advantage, they do not dominate. However, in the interest of guarding against actual domination, an agentialist could still insist that the government intervene to restore racial equality, compensate the victims and check all future vaccines for this side-effect and that a failure to do so would amount to racism.

Returning to Haslanger, we can now ask ourselves if uncoordinated actions of individuals are capable of producing the capacity for domination all on their own. Can they give rise to structural wrongs in the form of illegitimate power relations that are wrong in themselves? I think the answer is, absolutely. As a result of many activities of other agents, oppressed groups find themselves in a position such that they are vulnerable to the arbitrary will of another. To use Marilyn Frye (1983)'s famous illustration, they are caged in, unable to move in any direction, trapped by bars that may have each been created without the realization that the others already existed. And so, they are dominated. Many people, not just one, have the capacity to interfere with them on an arbitrary basis. However, unlike the million who are trapped by the mine, and the vaccine that causes racism against black people, this domination is not the result of some freak accident. It is the product of intentional human actions. That these actions were uncoordinated is not to say that they had to have been. And if we think that relationships of domination ought to be guarded against, then when domination arises in this way it is the product of a great deal of neglect.

Noticing that our uncoordinated activities, including the social and cultural norms we create and employ, have created illegitimate power relations between racial groups, the agentialist can declare that these activities, and their coordination, are within the scope of justice. We are required, collectively or as a state, to pay attention to the relationships between our actions and guard against their domination. That there might be some action that

none of us is individually morally responsible for performing but nevertheless, is required of us collectively is an intuitively plausible claim. For example, some have held that we have a collective moral responsibility to enter into political society and regulate our behavior according to the laws of that society. (We may also have a collective moral responsibility to dissolve that political society or refuse to be bound by some of its laws.) Others have noted our collective responsibility to feed the hungry or protect the defenseless. If Haslanger (2004) is correct, and cultural norms and other uncoordinated human activities give rise to domination, then it is our collective responsibility to arrive at some scheme for engaging in that regulative activity, and failing to do so is an expression of racist disregard. However, the agentialist does not have to arrive at this conclusion. Given a different view about the nature of justice, the agentialist might argue that social norms fall outside its scope.

To argue that power relations can arise for which no individual can be held responsible but that are nonetheless racially unjust is not, then, the same as claiming that racial justice extends beyond the scope of moral responsibility, understood collectively. Though it may well be true that no individual or collection of individuals intended to produce the total result, they had a collective moral responsibility to prevent it. Presumably, what makes it the case that the power relation instantiated by the media baron and the harmful vaccine is unfortunate, but not unjust, is that some natural disasters (the unpredictable ones) that produce harmful power relations fall outside the scope of justice.

And so, unlike the structuralist, the agentialist must admit that there are some human activities that, though they may create potentially harmful power relations, are not unjust. These are the true accidents of human interaction, like the vaccine case, which could not have been predicted and which no one was prepared to regulate against. But perhaps, we

should not only be concerned with power differentials that are created by intentional, and sometimes neglectful, human interaction. Maybe injustice is present before any intentional actions take place. Might the scope of justice extend beyond the scope of any agent's (or agents') moral responsibility?

If Haslanger (2004)'s view rests on the claim that no agent is responsible (including collective agents), she is committed to the claim that it is the power relation, independent of any actual agential capacity, that is unjust. But, noticing this, it is hard to see how the structuralist can avoid defending the counter-intuitive position that unpreventable natural disasters can independently create injustice. What reason do we have for limiting the scope of justice to the sphere of human activity if not for the fact that we often take human activity, uncoordinated and collective though it may be, to be the sort of thing that can (whether or not it ought) be regulated? By eliding the difference between the content of our claims about justice and the scope of those claims, the structuralist implausibly asserts that justice applies to power relations whenever and however they arise, or is forced to assert a division between the sphere of human activity and the rest of the natural world as a matter of brute fact.

Far from being a limitation, the agentialist's inability to declare that inequalities created by human accident or natural disaster are racist is actually an advantage of the theory. Even if we could assert that some phenomenon is racist without recognizing a connection to the attitudes and beliefs of agents, we ought not. To do so would not only trivialize racism and stretch the application of our concepts of justice and injustice beyond recognition but could also blind us to the dangers of ignorance and the need for vigilance in the face of the unknown. That human beings can suffer in ways that we did not cause and cannot fix is not a new idea. They may be born sick or mentally disabled or lose themselves to dementia or



mental illness. But, to argue that the mere instantiation of such circumstances is an injustice is to refuse to engage in the messy and complicated task of carving up responsibility, both individual and collective.

#### **4. The Messy and Complicated Task of Carving Up Responsibility**

As I mentioned briefly above, the strong account of collective responsibility that would allow the agentialist to capture (almost all) the racial injustices described by the structuralist is not a necessary feature of the agentialist picture. Just as the agential account is compatible with a number of different views about what, exactly, the wrong of racism consists in and what institutions can be thought to instantiate that wrong, it is also compatible with a number of different views about the scope of our collective responsibility. For instance, the agentialist can set the scope of justice more narrowly, arguing that certain sorts of human activities, though they might produce harmful power relations, cannot be properly understood as racist because they are too diffuse and thus, incapable of expressing anything that could be thought to be connected to some agents' racist attitude or belief. They might, for example, argue that social norms cannot be properly thought of as the bearers of racism because they are incapable of the necessary uptake of individual attitudes or beliefs. Laws, meanwhile, might be a prime example of institutional entities that can express the attitudes and beliefs of their creators. Alternatively, (or additionally), an agentialist might argue that there are many instances of non-accidental racial harms (such as the structuralist describes) that lie outside the scope of individual or collective responsibility. Some agentialists may even deny there is a realm of collective responsibility that is more than the aggregate of individual responsibility, and may even construe that individual responsibility in such a way as to rule out any obligation not to perpetuate the injustices of past agents.

These questions of collective responsibility are complicated and unlikely to be easily resolved. Structuralists attempt to side step these worries by arguing that certain power relations are, themselves, unjust but in doing so they commit themselves to the claim that the scope of justice is implausibly large. Agentialists might try to side step the issue as well, by offering an account that neglects to tell us which of our institutions (from laws, to social norms, to transportation systems) can be the bearers of racism (in other words, which are capable of expressing agential racism), what it takes for those institutions to express racism (is it enough that they perpetuate past racial injustices or must they go further to be thought of as racist?) and what our responsibility is to guard against these expressions of racism. The strength of the agential account is its flexibility. It is capable of capturing a number of accounts of the wrong of individual and institutional racism. But it is flexible precisely because it leaves so many questions unanswered and any theorist hoping to offer a comprehensive account of institutional racism must be prepared to fill in those gaps.

In the meantime, however, the agential structure gives us room to explore the possible contours of racial injustice in our actual world. By filling in those gaps in different ways we can begin to discover points of common agreement between different plausible theories of agential racism, institutional uptake, and collective responsibility. Those common points provide initial targets for anti-racist political action. And, ultimately, the aim of a theory of racism, institution or individual, is not simply to gain conceptual clarity but to aid us in alleviating the suffering of many real victims of racial injustice.

## 2. Implicit Racial Bias and Institutional Racism

### 1. Introduction

The term ‘aversive racism’ has been around since the 1970s to describe the cause of subtle signs of biased or avoidant behavior towards blacks by whites who profess to be non-racist.<sup>24</sup> However, over the last 20 years psychologists have gathered an impressive amount of evidence not only about the ubiquity of these racially-biased attitudes but also an apparent mismatch between them and individuals’ consciously-held beliefs and attitudes. This phenomenon is known as *implicit racial bias*.<sup>25</sup>

In this paper, I present ample evidence that white subjects who harbor these implicit biases exhibit behaviors indicative of racial bias against blacks. But, this literature also suggests it is possible they do so without being aware either that they have these attitudes (they believe themselves to be non-racist) or that they are acting upon them. In fact, these

---

<sup>24</sup> Joel Kovel (1970) was the first to draw a distinction between aversive and dominative (conscious) racism. Though the distinction he described has been modified, the term ‘aversive’ has survived as a way of marking out a type of racism that does not include conscious or explicit race-based antipathy. Exactly what aversive racism does include, however, is a subject of much debate, some of which will be addressed in this paper.

<sup>25</sup> For the purposes of this paper, I will use “implicit bias” interchangeably with “implicit racial bias” and I will concentrate my discussion on anti-black implicit racial biases. But implicit biases are not confined to the realm of race. Researchers have also used the implicit association test to discover implicit biases in any number of domains, including other social groups (genders, and sexual orientations).

attitudes may be no more than associative tendencies caused by a mere awareness of common racist stereotypes and attitudes. I argue that this interpretation of the biases and individual agents' relationships to them presents problems in accounting for implicit biases on any backward-looking theory of moral responsibility. That, in turn, makes it difficult to say that implicit biases are racist, if racism is not simply a matter of differential consequences but must stem from the bad attitudes (including hatred, inferiorization, or disregard) of agents.

However, I argue switching to an institutional account of racism makes it possible to explain how the discriminatory behaviors that these biases lead to are racist, even if the biases themselves and the agents who harbor them are not. I outline how an institutional account of this sort might go, distinguish it from other common conceptions of institutional racism, and end by briefly addressing some questions that arise for the moral responsibility of the relevant institutions.

## **2. Implicit Racial Bias: What Is It and What Does It Do?**

Perhaps the most famous tool used to measure this phenomenon is the Implicit Association Test (IAT). In an early experiment using the IAT by Greenwald, et al. (1998), white college students were asked to link stereotypically white and black names with pleasant or unpleasant words. They were told to move as quickly as possible and a computer recorded the speed with which they were able to perform each task. The students moved significantly slower when asked to associate black names with pleasant words and white names with unpleasant words than the reverse. The variance in response time was taken to reveal people's implicit biases towards the two races because, as previous research on

associative tendencies had shown, the stronger the association between the two concepts, the more quickly the subjects would respond.

Participants in the experiment were also asked to fill out confidential questionnaires that measured their explicit racial biases towards the two racial groups. The study found that the presence of implicit biased behavior was much greater than self-reported explicit bias. For the vast majority of the participants, their explicit racial bias against blacks measured at zero, finding them neutral for racial bias or indicating a preference for blacks over whites. However, all except one participant was found to exhibit anti-black prejudice on the IAT (Greenwald, 1998, 1475).

In the decade and a half since this study, IAT tests have been widely used to confirm the existence of and further explore implicit racial biases. They have discovered (among many others) implicit associations between Blacks and hostility, laziness, unintelligence, and athleticism. Other tests that require subjects to make split-second decisions have also uncovered implicit biases. A study by Payne (2006) found that subjects are more likely to mistake a tool for a gun when primed with a black face. In addition to these more straightforward response-time measures, a variety of other tests designed to measure subjects' automatic associations between two objects have consistently found whites hold negative implicit bias towards blacks.<sup>26</sup> Studies using subliminal messaging (to prevent test subjects from self-censoring) have also revealed whites' implicit biases against blacks.<sup>27</sup>

---

<sup>26</sup> See, for example, Blair (2001), Dovidio, et al. (2001)

<sup>27</sup> For example, an early study by Dovidio, et al. (1997) presented subjects with a subliminal sketch of a black or white person (flashing it on a computer screen so quickly that they were not consciously aware of it) and then measured how long it took them to decide whether or not a particular characteristic ("good," "bad") could be used to describe a person generally. Shorter response times were taken to reveal greater association between the subliminal sketch and the characteristic. White subjects consistently tested as having greater association between the subliminal black face and negative characteristics, and the white face and positive characteristics, despite the fact that they were neither aware of the face nor that the test was measuring their racial attitudes.

While the results of the IAT are certainly startling, they likely would not be as troubling (and would not have inspired such a great deal of psychological research and philosophical analysis) if implicit biases only affected participants' behavior in interactions with computers. However, since the IAT was first performed, a plethora of studies have confirmed a correlation between an individual's score on an IAT and her social behavior, including how she judges and decides.<sup>28</sup> For example, in experimental settings, whites with implicit biases against blacks are more likely to talk and smile less and make more speaking errors when interacting with a black, rather than a white, interlocutor. They make less eye contact, sit farther away, and blink more often— all microbehaviors associated with unfriendliness (McConnell and Leibold, 2001). In other words, agents with implicit bias perform small-scale racial discrimination. Because agents are not attending to these behaviors— how often do any of us think about the number of times we blink our eyes while engaging in conversation?— they are often the sites of subtle biases that the agents are unaware they are displaying and may not be aware they have.

However, to say that these subtle nonverbal behaviors are performed unconsciously is not to imply that they are without effect. Without realizing it, we all pick up on the subtle nonverbal behavior of those around us. We may not notice what it is about someone that gives us the impression that she is uncomfortable around us or dislikes us but, nevertheless, we get the message.<sup>29</sup> And the same is true for blacks interpersonal interactions with whites

---

<sup>28</sup> A meta-analysis of 122 IAT reports by Greenwald, et al. (2009) found that the automatic associate process recorded by the test was a reliable measure of social behavior (more so than self-reported attitudes). And, important for my purposes here, "For 32 samples with criterion measures involving Black—White interracial behavior, predictive validity of IAT measures significantly exceeded that of self-report measures" (p17). This research is not without exception. Gehring, et al. (2003), for example, hesitate to claim that scores on the IAT are necessarily linked to racially biased behavior or unconscious racial prejudice in the individual.

<sup>29</sup> Word, et al. (1974) found that, under experimental conditions mimicking job interviews by white interviewees for both white and black applicants, the black applicants received less immediacy (increase in

who have implicit biases. When asked to report on their experience of such an interaction, one study found that white subjects generally assess it positively (Dovidio, et al., 2002). They reflect on their conscious attitudes and deliberate actions to judge the experience. However, black participants generally judge that the whites they interacted with hold conscious racial prejudices. They pick up on whites' unconscious nonverbal behaviors and judge the interaction by those.<sup>30</sup>

The mixed messages that whites send in these interactions by verbally expressing non-racist views or friendliness but giving off nonverbal cues that they dislike or distrust their interlocutors can create anxiety for blacks.<sup>31</sup> And that, in turn, could create a looping effect such that blacks, in interactions with whites, will send back nonverbal signals of unease and discomfort that put whites on guard or cause them to judge blacks as hostile and prejudiced. More worrisome is the fact that whites' negative microbehaviors not only affect how blacks' perceive them but also affect blacks' performances. One study found that when whites with implicit bias were teamed up with blacks, their team solved problems less efficiently than interracial teams with either non-implicitly biased whites or explicitly biased

---

physical proximity or visibility), more speech errors and a shorter interview time. In a second experiment, white applicants were exposed to these same conditions (they were given a shorter interview time, less physical proximity and exposed to more speech errors) and they performed worse in the interview than whites in the first experiment.

<sup>30</sup> One possible explanation for whites' seemingly unfriendly and racist behavior in interactions with blacks is that they are nervous or anxious about appearing racist and so, display microbehaviors indicative of that discomfort or anxiety. Garcia (2011) levels a similar objection against Faucher and Machery's (2009) contention that his theory of volitional theory of racism is inadequate because it fails to account for implicit bias. He wonders why, exactly, we should think behaviors that indicate discomfort are racist unless they stem from an objectionable cognitive or non-cognitive attitude. If this hypothesis is correct then implicit biases, despite the well-established predicative validity of IATs, are not causing these behaviors. Rather, implicit biases are inert as mental states and the behaviors are a response to unrelated discomfort or anxiety. However, given the predictive validity of the IAT for these behaviors, I find this hypothesis implausible. If implicit biases are not active, in some way or another, in this process then it odd that they are co-occur with these microbehaviors such that they can be used to predict the presence of the behaviors. Nevertheless, I remain open to the possibility that implicit biases are highly correlated with, but not caused by, implicit biases.

<sup>31</sup> See Hyers and Swim (1998) and Shelton (2000).

whites (Dovidio, et al, 2002). Moreover, apart from the effects of the implicit biases themselves, the combination of overt friendliness with subtle indicators of unfriendliness can also cause harm. Dovidio and Gartner (2000) write,

These different perspective and experiences of white and blacks in interracial interaction, which happen inadvertently and occur daily, can have summative effects over time and help to contribute to the climate of misperception and distrust that characterizes contemporary race relations in the United States...The mixed messages that aversive racists often convey can create fundamental miscommunication in interracial interaction and produce divergent impressions among interactants that can undermine their ability to interact efficiently in task-oriented situations as well as effectively in social situations. (p24)<sup>32</sup>

This perceptual divide between blacks and whites with respect to whites' attitudes and behavior towards blacks appears to have persisted, with a recent Gallup poll (2013) showing that while 67% of whites are satisfied with the way blacks are treated in society, only 47% of blacks are.

There are other indications that implicit biases may be playing a role, not only in our interactions with computers or inside the lab, but also in our daily lives. These include a well-known experiment by Bertrand and Mullainathan (2003) that demonstrated that when employers are presented with identical resumes topped with either white or black names, they are overwhelmingly more likely to call the white candidate in for an interview than the black one. Other experiments have uncovered similar results. An IAT study by Green, et al. (2007) of 287 physicians at four academic medical centers in major metropolitan areas found that

---

<sup>32</sup> One could argue that blacks' mistrust of whites is the initial instigator of these microbehavioral loops. In other words, whites' negative microbehaviors are response to the negative microbehaviors of blacks. If that were true, however, one would expect that white participants would issue a negative report when asked their impressions of an interracial interaction but the opposite is generally true; Whites report feeling positive about the interaction. But perhaps whites' microbehavioral responses to black microbehaviors happen at an entirely unconscious level. Given whites' societal dominance, they could less aware of microbehaviors than blacks because knowing and managing blacks' opinions of them has fewer consequences for how they live their lives. However, if this were the case then individuals possess two independent but highly correlated unconscious processes (implicit biases and unconscious microbehaviors) rather than one interrelated process.



though the doctors registered no explicit bias against black patients, their IAT results showed an anti-black bias. The stronger that bias, the greater the likelihood that the doctors would treat white but not black patients with thrombolysis. Another study published by the consulting firm Nextions (2014) found that supervising lawyers not only rate the same memo draft lower when they are told that it was written by a black rather than a white associate, but they actually find more of the intentional errors planted by researchers when they believe that the author was black.<sup>33</sup>

These real world cases do not prove definitively that implicit biases cause discrimination. In the study of attorneys, for example, it is possible that those individuals held explicit racist attitudes that caused them to rate the black associates lower than the white ones. After all, they were given neither an IAT nor an explicit attitudes' test. The same is true of the resume test. It is possible, though statistically unlikely, that researchers who discover these subtle biases in real world settings are encountering all and only the actions of conscious racists and not those with implicit bias. Furthermore, most studies of implicit bias (both generally and cited here) use U.S. white college students as test subjects and may ask them to perform tasks that are beyond their level of competence. A study that finds that these college-aged test subjects discriminate against black employees when asked to perform a hypothetical managerial task does not prove that experienced managers who test as implicitly biased will do the same. The study of physicians mentioned above would seem to tell against this hypothesis because that population was given both the IAT and an explicit attitude test

---

<sup>33</sup> Furthermore, the qualitative comments the supervising lawyers provided for the white associate were more positive than those for the black associate. A similar experiment by Moss-Racusin, et al. (forthcoming) focused on gender bias among faculty in science departments across the U.S. Researchers found that when presented with an identical resume for a laboratory manager position, topped by either a male or a female name, both male and female faculty members rated the male applicant more competent and hireable.

(testing as anti-black on the former and egalitarian on the latter). They were also asked to adjudicate a case that they would be likely to encounter in their workplace. But, again, those cases were hypotheticals. Perhaps, when dealing with actual patients, only those individuals with explicit racial bias would discriminate. It is possible that those with implicit biases perform behaviors in a hypothetical experimental lab that are indicative of racial discrimination (facial expressions of unfriendliness, for example) but only those with explicit racist beliefs and attitudes act on these behaviors in real life.<sup>34</sup>

This conclusion strikes me as implausible given the accumulation of evidence that a biased rating on the IAT (combined with either the same biased or a neutral rating on explicit measures) is predictive of agents' real-world behavior.<sup>35</sup> But, prediction does not prove causation. There is an alternate explanation for these real-world results that is far more plausible than the hypothesis that they influence agents' behavior in the laboratory but not outside; Perhaps most people hold conscious anti-racist beliefs and attitudes but are hesitant to self-report due to a general shift away from the social acceptability of explicit racism. It would not be hard for those individuals to fake explicit non-racist attitudes on self-report

---

<sup>34</sup> Garcia (2011), for example, expresses skepticism over the claim that those who test as biased on the IAT will discriminate in real-world settings. He writes, "What reason is there to think the spontaneous, quick, unthinking, unconscious, conceptual associations evidenced in these association or identification tests will also manifest themselves in a tendency (statistical tendency, e.g., or a psychological inclination) to racially discriminatory (and racist) actions when subjects act with time, deliberation, attention and care in such self-consciously careful, thoughtful, even painstaking endeavors as deliberating about hiring or promoting employees, admitting candidates for education or training programs, grading tests taken in school or at work, leasing to or relocating tenants, processing loan applications, siting public facilities, convicting or paroling defendants, voting, and so on? That connection cannot be inferred, but requires independent empirical support. Without it, the tests results may have little to do with the more serious and consequential spheres of racial discrimination" (p. 259). Garcia is right to point out that implicit biases are far less likely to influence behavior when individuals are given time to deliberate. However, as the studies of physicians, attorneys, and NBA referees (as well as the weapons bias test) indicate, I think he underestimates the frequency and potential impact of split-second decisions.

<sup>35</sup> For example, Greenwald has collected a list of 59 independent studies demonstrating the predictive validity of the IAT on real-world behavior including voting and political attitudes, mental health, medical (including the study of physicians' racial biases cited above), employment, education, forensics and relationships.

tests. If this is true, then implicit biases are not actually implicit at all. The test is simply a better way of measuring an agent's explicit racist attitudes and beliefs. Any discriminatory behavior they cause is actually the result of agents' conscious racist beliefs.

It is likely true that some of the participants in these studies hold conscious racist attitudes or beliefs and are lying on their self-reports. However, the fact that reporting such attitudes has become less socially acceptable is one indication that the prevalence of conscious racist beliefs has decreased. At least some whites agree that racist beliefs and attitudes are morally wrong and yet large percentages of whites continue to harbor implicit racial biases. As a study by Bobo (2001) of racist attitudes in the U.S. found, "The single clearest trend in studies of racial attitudes has involved a steady and sweeping movement towards general endorsement of the principles of racial equality and integration" (p. 269). Furthermore, subtle biases of this sort can show up even in situations where the parties are thought to be strongly incentivized not to behave in a biased manner; A study by Price and Wolfers (2010) of black and white NBA referees found that despite extensive training to perform their duties impartially, referees showed a slight but statistically significant own-race bias when calling fouls.

However, I do not want to dismiss the possibility that implicit bias tests measure conscious attitudes that individuals are reluctant to admit. I will revisit this claim, along with the psychological literature that supports it, in the next section. Instead, supported by the empirical evidence, I want to assume simply that the results of implicit bias tests are predictive of behavior that should be of moral concern. Whether or not these behaviors are caused by implicit bias, let's agree that the behaviors are real, show up in real-life situations, and that their effects are harmful. Although the discovery of racial discrimination in NBA

games might not inspire the same ire as discrimination in the medical field, both should leave us troubled. Even in individuals who are highly trained, heavily scrutinized and incentivized to make non-biased decisions, these subtle, discriminatory behaviors appear. These biases are particularly likely to manifest in situations that demand rapid decisions. And though the biases may be subtle, their consequences need not be. A snap-second judgment can have devastating effects. I mentioned earlier a study that showed subjects were more likely to mistake a tool for a gun when primed with a black face. Might they also be more likely to mistake a wallet for a gun, as police apparently did in 1999 when they shot 23-year-old Guinean immigrant Amadou Diallo 19 times? That implicit biases are harbored by and appear to affect the behavior of highly trained experts in sports and medical fields should make us wary that others, including the police, might be equally susceptible.

Of course, it is not news to any of us that, despite trying our best to act well, we can do devastating and wrongful harms to others. But, the literature on implicit bias drives home exactly how common and how easy this might be. Even more disturbingly, it suggests that many of us could be (perhaps unintentionally) harming a particular set of people, a racial group, because they are members of that group. And this realization naturally directs our attention to the question of whether or not our implicit biases, and we ourselves are racist. It is this question to which I turn now.

In the next section, I briefly review the current debate over the nature of implicit biases, their moral status, and individuals' moral responsibility for them. In doing so, I address three related questions: (1) Are implicit biases racist? (2) Are the behaviors they cause racist? and (3) Are agents with implicit bias racist? I argue that while many interpretations of implicit biases and their cause can be accounted for on some backward-

looking, individualist picture of moral responsibility (and so, are racist), one plausible interpretation of what implicit biases are and how agents come to have and act upon them, cannot. I show that if that interpretation is correct, we would be forced to admit that implicit biases, the troubling and discriminatory behaviors they cause, and the agents who harbor them are not racist.

### 3. Are Implicit Biases Racist? The State of the Debate

Despite the fact that the phenomenon of implicit bias is both well established and widespread, it is unclear exactly what these biases amount to and what their connection is to theories of racism; more specifically, are implicit biases, the behaviors they cause, and/or the agents who harbor them *racist*? This set of questions has received some attention in the philosophical literature. However, because implicit biases are not confined to the realm of race but occur in response to any number of social groups, the broader question of whether or not the biases are immoral (and the agents who harbor them are blameworthy for those biases and the behaviors they cause) has been asked with a great deal more frequency.<sup>36</sup> Continuing developments in the psychological literature, as well as uncertainty as to how to interpret those results has led to a host of divergent individualist moral analyses. While philosophers

---

<sup>36</sup> Implicit biases have received a fair amount of attention in philosophical literature lately (even spawning the Implicit Bias & Philosophy International Research Project and a forthcoming two volume collection due to the interesting questions they raise in epistemology and ethics. However, with a few notable exceptions, including Faucher and Machery (2009) Garcia (2011), Kelly and Roedder (2008), very little work has been done to bridge the divide between literature on implicit bias and accounts of racism. In the psychological literature, it has sometimes been assumed that the term appropriately applies and implicit biases have been labeled as a type of ‘aversive racism.’ See, for example, Dovidio and Gaertner (2004). Less still has been written on the relationship between implicit bias and institutional racism but several interesting accounts have linked implicit bias to justice, more generally. For example, Miranda Fricker (2007) has argued that implicit biases are a form of “epistemic injustice,” while Sally Haslanger (2014) contends that the U.S. education system perpetuates the implicit ideology of racialization that regulates blacks to inferior positions *vis a vis* whites. (On Haslanger’s view, racialization just is the creation of a racial hierarchy.) Given the tight connection between ‘institutional racism’ and ‘racial injustice’ perhaps these could be considered theories of implicit bias as institutional racism though they are different from the view I will explore here.

are generally united in their concern for the harms that implicit biases may cause, and generally committed to the view that such harms should be rectified, they are divided on the moral status of both the biases themselves and the individual agents who harbor them.

However, given substantial unanimity among philosophers that the harms caused by implicit bias are of moral concern, attempting to discover whether or not the biases themselves are immoral, much less racist, might seem like a pointless exercise. If we agree some behavior ought to be rectified, why care if it is *racist*? Why think it deserves some special label that sets it apart from both the harms it causes and from the many other ways in which it is possible to be vicious towards other human beings?

In this question, there are two significant critiques of engaging in the project of showing that implicit biases are racist. The first is a fairly familiar criticism of backwards-looking arguments for moral responsibility; why should we care about establishing fault if we all agree that we have a responsibility to rectify the harm? We could instead concentrate on fixing the problem and cease the unproductive activity of finding someone at fault for it. Though I lack the space to respond to this argument in full, I do want to flag a suspicion that while blame can sometimes be counter-productive it can also serve to motivate political action. Finger pointing, when it is done tactfully, can take problems as vast and seemingly insurmountable as implicit racial bias and begin to break them down into more manageable demands for change.

The second critique is more specific to my task. Why should we care if implicit biases are *racist*, in particular, as opposed to vicious or hateful in some other way? There are a great many bad attitudes that one can adopt towards other people that fail to accord them their proper moral status and that failure should always trouble us, regardless of whether or

not it is based on race or some other, morally arbitrary feature. The question of whether these biases are racist does not and should not decide whether or not we should care about them. Nevertheless, I think the answer is important. In the last several centuries, some of humanity's most deplorable crimes have been committed in the name of or rationalized away by racist doctrines and attitudes. The genocide of the indigenous people of the Americas, the trans-Atlantic slave trade, the Holocaust, and the slaughter of Chinese at Nanjing all had racist ideologies at their center. Racism, it seems, is a particularly awful way of being bad. The justificatory work such attitudes and ideologies have performed represents the worst in humanity. Recognizing this, racism is a moral wrong we should take particular care to avoid.

Some philosophers shy away from the term, preferring to reserve it for particularly vicious or deliberate expressions of bad attitudes.<sup>37</sup> I do not want to restrict it in this way. Like most moral wrongs, there are gradations of racism; we need not think ourselves a Nazi because we discover we hold a racist belief, or have acted on a racist attitude. But, finding ourselves in the ranks, if only in thoughts and attitudes, of those who committed such unimaginable atrocities should give us pause. Even the less virulent forms of racism, though less worrisome in themselves, might be easier to inflame into the truly bad kind. Racism, we might think, is a particularly dangerous form of immorality. So, though one need not be awful to be racist, among types of badness, it is an awful one to be. Much more so than 'lazy,' or 'rude,' the label 'racist' is a serious one, and the discovery that our behavior is racist should be taken seriously. And so, it is of particular importance whether or not some moral wrong is racist.

---

<sup>37</sup> See, for example, Anderson (2010), Garcia (1996).

That both the label and the wrong it denotes are serious, however, is not to say that its proper application is well understood. Even less clear is when and under what circumstances a particular harm counts as “institutional racism.” Due to this country’s history of racist oppression, an institutional policy that has a differential impact on racial groups, disadvantaging a group that was previously discriminated against, is often assumed to be racist. And, given that history, differential impact is often an important, though not a sufficient, condition for racism. In actuality, institutional racism is much more complex. It requires examining the attitudes and intentions of institutions and/or the individuals who make them up, their organizational structures, and the scope of their moral responsibility, as well as their behaviors. Discussing how a particular phenomenon like implicit bias is an instance of institutional racism is one way of increasing clarity about institutional racism more generally and raising (if not always answering) some of the tough questions about how and when the concept appropriately applies.

My goal in this section is to begin that process, by connecting recent empirical and philosophical literature on implicit biases to theories of racism. I show how individual theories of morality and moral responsibility might account for implicit bias and how their ability to do so depends on what, exactly, those biases are and how we interact with them as moral agents. To facilitate this process, I will assume that racism is always immoral. Agents, their attitudes, and their actions can all be racist, though when actions are racist it is because they stem from racist attitudes. This assumption is useful because it allows for a fairly straightforward connection between the concept of racism and various theories on moral



status of implicit bias and biased agents. Moreover, it is also an assumption that is well supported by the current literature on racism, and upheld by most of the dominant theories.<sup>38</sup>

In order to ensure that every possible (non-consequentialist) way in which implicit biases and biased agents could be racist is fully explored, and no plausible interpretation is ignored due to controversial, pre-theoretical commitments about the nature of racism, I will adopt a hybrid view of racism on which both cognitive and affective attitudes can be racist.<sup>39</sup> Racist cognitive attitudes consist in inferiorizing beliefs (beliefs that they are inferior) about a member or members of a racial group because she(they) is(are) member(s) of that group. Racist affective attitudes are expressions of antipathy, disrespect, or disregard for a member or members of a racial group because she(they) is(are) member(s) of that group.<sup>40</sup> Either is sufficient for racism, but they need not be full descriptions of the wrong that racism often is. In many cases, treating people in a racist manner includes treating them as if they were not human. Though I will not explore it here, I take it that there are more and less inferiorizing cognitive attitudes that one could have and degrees of affective attitudes one could hold such that not every racist belief or affective attitude is as bad as every other, but all racist attitudes are immoral.

---

<sup>38</sup> Nevertheless, this assumption is neither universal nor entirely uncontroversial. While “racism” is generally a term of moral condemnation in our ordinary language, Charles Mills (2002) and Tommie Shelby (2002) argue that whether or not racism is immoral cannot be stipulated ahead of offering an account of the concept. They claim that to do so unfairly stacks the deck against doxastic account of racism (on which racism is primarily a matter of beliefs). Here, I assume that, while not all racial generalizations need be racist, I follow Blum (2002) in asserting that certain inferiorizing beliefs are (even if they do not stem from or co-occur with attitudes of hatred or disregard). This helps to make sense of cases of racial paternalism, which might present problems for an account, like Garcia’s (1996) that focuses exclusively on the connection between racism and certain negative affective attitudes. So, though I assume that racism is immoral, I make room for the possibility of racist beliefs unconnected to vicious attitudes of hatred or disregard.

<sup>39</sup> Adopting a more restrictive view on which, for example, only beliefs can be racist could unnecessarily truncate the discussion if implicit biases turn out to be associations or affective attitudes.

<sup>40</sup> I take “disregard” to include a broad spectrum of objectionable affective attitudes including culpable ignorance (turning a blind eye towards one’s actions and beliefs) and insensitivity.

That being said, I want to leave room for the possibility that beliefs and attitudes that are characteristically racist can fail to be so if they stem from genuine ignorance. Beliefs that seem racist may, in fact, be naïve. Attitudes that appear hateful may turn out to stem from fear or confusion. However, ignorance is not a blanket excuse. Inferiorizing beliefs offend against what we all know to be human beings' proper moral status, and so are racist even when the people who hold them do not do so out of hatred. Paternalistic attitudes towards a racial group are racist.<sup>41</sup> Vicious attitudes or expressions of disregard are racist if those who hold them could have done better. Yet, a behavior is not racist just because it offends a racial group. It is racist only when it is a product of inferiorizing beliefs or the affective attitudes described above.

A few additional caveats before I begin: First, I assume that if implicit biases (whatever they turn out to be) are racist then the unconscious discriminatory behaviors (as described in the preceding section) that are their direct consequence are racist. However, even if implicit biases are racist, that does not entail that the agents who harbor them are racist. Instead, I will explore a number of possible excusing conditions under which an agent might have, and act on, racist implicit bias without being properly understood as a racist.<sup>42</sup> Second, if implicit biases are racist, and if the agents who harbor them are at fault for those biases, then, I assume, those agents are properly referred to as "racists." In other words, I

---

<sup>41</sup> Blum (2002).

<sup>42</sup> This view is different from one in which beliefs and attitudes can be racist irrespective of their relationship to the individual. For example, Anderson (2010) distinguishes between racism which she reserves for "judgments of serious vice" and racial stigmatization which, she argues is a "broader concept" that does not require a connection between the stigmatizing behavior and some objectionable attitude or belief on the part of the individual who performs it. For example, the actions of a hotel guest who hands his keys to a black businessman, believing him to be a member of the hotel staff expresses racial stigma (given stereotypes and power relations that keep blacks in a subordinate role) even if he was completely unaware of such stereotypes and just made a mistake. She writes, "Conduct expresses stigma if it insults or disadvantages the stigmatized group in ways that fit the stigmatized ideas..." (p 48-49).

apply the term “racist” to any individual who is properly understood as morally responsible (at fault) for her racist attitudes, implicit or otherwise.<sup>43</sup> Finally, I assume that an agent could be racist even if her implicit biases are not racist, if the biases stem from her racist character. I make these assumptions because, in presenting these individualist accounts of the immorality (or not) of implicit bias and implicitly biased individuals, my goal is not to rule them out. Instead, I hope to show that although there are many possible phenomena for which these individualist pictures can account, there is at least one plausible, minimalist interpretation of the nature of implicit bias that the individualist picture cannot accommodate. I argue that if this minimalist interpretation is correct, implicit biases are racist, albeit on an institutional, rather than an individualist, framework.

### **3.1 What Is the Bias Itself? Implicit Biases as Mental States**

Implicit biases are often referred to as unconscious racist attitudes; Individuals who harbor implicit biases are typically unaware of them.<sup>44</sup> However, to say that an implicit bias is unconscious is not to say that it (and the person who holds it) is not racist. Most, if not all, plausible theories of racism make room for the possibility that one can harbor unconscious beliefs and attitudes that are properly termed ‘racist.’ In terms of assessing their racism, unconscious attitudes are subject to the same evaluation as consciously held ones: are they

---

<sup>43</sup> This marks a departure from theorists like Blum (2002) who wants to reserve the term “racist” for individuals with particularly vicious beliefs or attitudes, or whose character is marked, in some significant way, by their racism. My aim here is not to deny that may be a better use of the term. I only intend to use it to mark out when a person is at fault for their racist attitudes and beliefs.

<sup>44</sup> Here, I mean to indicate that implicit biases are unconscious insofar as the agents who harbor them are unaware of the existence of the attitude. They would lack what Gawronski, et al. (2006) have termed “content awareness.” They may also lack “source awareness” or “impact awareness” (they may be unaware where their attitude comes from or the effect it has on their behavior, respectively) but those features, I would argue, are less central to assessing a) whether or not the attitude itself is racist and b) whether or not the agent who harbors it is racist. However, they are not completely unimportant. I will revisit some potential implications of impact awareness for moral responsibility later in this section.

vicious, disrespectful, disregarding or inferiorizing to members a racial group because they are members of that group? If so, then they are racist.

For example, imagine a landlord who consistently fails to rent rooms to a particular racial group. When queried about his decision, in each particular case he cites his suspicion that the tenant would have been ‘too loud’ as his reason for refusing to rent the property. This phenomenon is robust over a variety of circumstances including those in which it was clear that some tenant of another racial group whom he did rent to was, in fact, more likely to be loud (a large, white family, for example, vs. a single, black female.)

After gathering data over a period of time, we bring it to the landlord’s attention. He is shocked to see the evidence that he has acted in a discriminatory fashion. But, when he reflects upon his behavior and pays attention to his thoughts when he is around black people, he begins to notice that he does expect them to be louder than white people. He avoids sitting near them on buses or in other public spaces and is reluctant to rent apartments to blacks, even when he has good reason to believe the individual in question will be just as quiet, if not quieter, than a white renter. He admits that, though he was previously unaware of it, he harbored unconscious negative attitude about blacks because he believed that black people are louder than whites and it caused him to discriminate against black applicants. He had an unconscious racist attitude. Perhaps implicit biases are unconscious attitudes like that, and so, are racist.

However, it could be that implicit biases are not attitudes, conscious or unconscious, but mere associative tendencies. In their paper reviewing the philosophical implications of implicit bias, Kelly and Roedder (2008) formulate this minimalist construction in a particularly helpful and intuitive manner: they write that on this view, “...an implicit attitude

is simply a tendency to associate one concept with another, in the way that, for instance, the concept *salt* might prime the concept *pepper*” (p. 528). So an agent’s high rating on an IAT might be, from a cognitive perspective, *nothing more than* a strong tendency to associate Blacks with unpleasant rather than pleasant words.

Unlike unconscious beliefs or affective attitudes, perhaps associative tendencies are not properly subject to moral evaluation. Garcia (2011) argues that without an account of how the individual associates the two concepts, it is impossible to know whether or not the association is racist. An association between, say, blacks and guns could be caused by the knowledge that blacks are more likely to be the victims of violent crimes rather than the belief that they are more violent than whites or more likely to perpetrate crimes (Garcia, 2011).<sup>45</sup> However, the point is not merely that we do not know why the individual with implicit bias is associating the two concepts and so, cannot know for sure if they are racist. Rather, the more interesting result is that if an individual’s implicit biases are merely associative tendencies (and the individuals possess no cognitive or affective attitudes that account for the associations), and those associative tendencies are too minimal as mental states to be apt for moral evaluation, then there is nothing in the individual that can properly be called “racist.” Implicit biases, on this view, might be harmful but they are not robust enough, as mental states, to be racist.

### **3.2 Are Agents at Fault for their implicit Biases?**

Even if it turns out that implicit biases are unconscious racist attitudes, it may be that agents with implicit biases are not at fault for them. And so they are not, as I have stipulated

---

<sup>45</sup> As Garcia (2011) writes of implicit bias, “How, it needs to be explained, is the implicit bias that consists in mere association of concepts tied to prejudice and thus to real psychological states (either affective, cognitive or volitional)...?” (p. 256).

that the term will be used, racist. On my account, a racist agent must, in some sense, endorse such attitudes and beliefs.<sup>46</sup> The attitudes must be, on some plausible interpretation, *hers*. To determine if they are hers, we might ask: Could individuals with implicit biases have had different attitudes? This question can be posed in (at least) three different ways: First, are agents who have implicit biases aware of them or could they be rightly criticized for failing to be aware of them? Second, is it within an agent's power to change such biases? Third, if an agent becomes aware of and finds she cannot change her implicit biases, does she embrace them? Or does she disavow them and the actions they cause her to perform, wishing sincerely that she could act differently?<sup>47</sup>

Let's begin with a discussion of whether or not agents are aware, or should be aware of their biases, and so, at fault for them. In the case I described above, the landlord was unaware that he had a racist attitude towards blacks. But is that enough mitigate his responsibility? Is the landlord not a racist, because he was unaware of his racist attitude? Not necessarily. One could argue that had the landlord been paying attention to his actions, he would have noticed that he harbored a racist attitude. It was not buried deep within his subconscious, but was available to him if he vigilantly searched his psyche. And, raised in a racist culture, where racist attitudes are quite easy to pick up, maybe the landlord was morally obligated to regularly subject himself to such self-examination. So, that his racist attitudes were unconscious does not necessarily absolve him of the charge of racism.

---

<sup>46</sup> Even if implicit biases are mere associative mental states an individual could still be at fault for them, depending on how she came to possess them and how she reacts to discovering she possesses them. I will return to this point later in this section.

<sup>47</sup> In other words, is she responsible in the sense that she is accountable for them or are they merely attributable to her? Robin Zheng (forthcoming) argues that we ought to employ the latter account of responsibility when talking about implicit bias, given the difficulty of knowing if excusing conditions obtain when individuals act on implicit bias.

Early research on implicit biases, which found a marked divergence between white's self-reported explicit attitudes and their implicit attitudes, hypothesized that the gap was due to the fact that the subjects were unaware of their implicit attitude and so, could not report it;<sup>48</sup> The whites tested appeared to genuinely believe that they were non-racist. However, more recent tests have called this conclusion into question, showing that the divergence between white subjects' explicit attitudes and their implicit biases is lessened when whites believe that the IAT is capable of accurately measuring their racial attitudes (Gawronski, et al., 2006). When whites are not incentivized to respond in a socially desirable way, their self-reported explicit attitudes towards blacks are more negative, and thus, more closely match the results of the implicit bias tests (Nier, 2005). Decreasing the time that subjects have for deliberation when completing self-reports also increases correlation (Gawronski, et al., 2006). These data, along with other studies examining the effects of motivational and deliberative factors on subjects' self-reports, suggests that whites are actually conscious of their so-called 'implicit' attitudes. If this is true then implicit bias tests are just a means of showing that white subjects are lying about what they consciously feel and believe about blacks. They hold conscious anti-black attitudes.

Regardless of whether or not someone with implicit bias is aware of her biases, she may still not be responsible for them if it is not in her power to change them. As Garcia (1996) might ask, in assessing whether or not an agent is racist, 'Are the negative attitudes *volitional*?' Perhaps an agent harbors conscious racist attitudes, which are revealed by implicit bias tests, but sincerely desires to rid herself of them. She is convinced by the

---

<sup>48</sup> See, for example, Banaji (2001), Cunningham et al., (2004), Jost et al. (2002), Phelps et al., (2000) and Rudman et al., (1999).

evidence that her racist beliefs are false and bad attitudes unwarranted but she cannot stop thinking and feeling them.

Of course, on certain pictures of how our cognitive attitudes are formed, this will seem absurd. Surely, if she were truly convinced that her beliefs were false, then they would not be her beliefs. The beliefs she had would have changed. (Although, importantly, this need not be the case for noncognitive attitudes, which seem capable of sticking around well after we have decided they are unwarranted. The same is true of our perceptual experience. If the evidence that a stick remains straight when dipped in water is convincing, then although our perceptual experience of the stick may never change and our feelings about the stick might take a very long time to change, our beliefs about the stick and its form are altered. We no longer behave as though the stick becomes bent each time we dip it in water.)

However, early studies suggested that this was not the case for implicit biases. Researchers discovered that even after subjects were made aware of their implicit biases, they were unable to revise them despite the fact that they appeared to be convinced that the attitudes revealed were false or immoral. One possible explanation for this result is that subjects were unable to truly hold the opposing view. Irrespective of their professed attitudes and beliefs, the subjects were actually unconvinced. Still, numerous studies have shown that implicit biases, are unaffected by what one might think of as these traditional modes of belief and attitude revision. One possible explanation is that implicit biases are automatic associations and not full-fledged attitudes. Attitudes, cognitive or noncognitive are our responsibility because we can control them. They are revisable. Automatic associations, however, lie outside our control and if implicit biases are those it may be that individuals are not at fault for them.



And yet, recent studies have shown that implicit biases (or the behavior they give rise to) may be at least partially under an agent's conscious control. However, strategies for altering biases or mitigating their effects are not intuitive.<sup>49</sup> They cannot be overcome simply by drawing one's attention to the falsity, irrationality or viciousness of a biased attitude. But if when individuals are taught strategies for overcoming their biases they can change, then provided such strategies are not extremely difficult or time-consuming, they might well be morally responsible if they continue to exhibit implicit biases, even if the biases are mere automatic associations.<sup>50</sup>

Such a view would seem to absolve subjects of fault who currently exhibit implicit bias, be they conscious or unconscious attitudes, but (through no fault of their own) have no learned strategies for eliminating that bias, of racism. But, it may be that an agent's current inability to control such attitudes does not actually mitigate her responsibility. We might ask how the subject came to possess such attitudes and if it was ever in her power to have developed different ones. To draw upon an earlier example, imagine that the landlord, having become aware of his racist attitude, sincerely desires to change it but finds that he is unable to. We might still be tempted to argue that he is at fault for his attitude, and therefore racist, if he could have, at some point in the past, formed a different one. Unlike a person

---

<sup>49</sup> See, for example, Kawakami, et al. (2000); Dijksterhuis & van Knippenberg (1996); Karpinski & Hilton (2001); Rudman & Borgida (1995); Dasgupta & Greenwald (2001).

<sup>50</sup> On the other hand, if implicit biases are not fundamentally different mental states from our consciously held attitudes (or they are, in fact, conscious attitudes), then it is possible that the fact that they are not subject to traditional modes of revision is indicative of a problem with those modes. Maybe our traditional understanding of the sort of control necessary for moral responsibility was misguided. Agents were always morally responsible for attitudes that they held because they failed to subject them to these sorts of revisionary strategies because the old strategies of revision were fundamentally misguided. Alternatively, one might argue that agents are no more morally responsible for any other attitudes than they are for implicit biases because, prior to the insights gleaned by modern psychology, they could not have revised any of their attitudes. Such revisions, when they do occur, are a matter of dumb-luck and not a consequence of rational reflection or meditation on one's moral values.

who was raised in a culture that was so thoroughly steeped in racist attitudes that she never had a chance to develop non-racist ones, the landlord cultivated his biases, though they are now impervious to change. The biases are connected, in the right way, to beliefs or attitudes that the agent was responsible for. Similarly, if those with implicit biases can no longer change them, they might still be responsible for them, and thus racist, if biases developed because the individuals held objectionable attitudes or beliefs. The biases might also turn out to be racist, even if they are mere associative tendencies, because they stem from some immoral feature of the agent who develops them.

Finally, if implicit biases are mere automatic associative tendencies and an agent with implicit bias is unaware of her bias, cannot control it and was not in control of the development of the biases, she still might be a racist if those biases are indicative of her deep, underlying character. Perhaps, when discussing the agent's moral responsibility for implicit biases (and moral responsibility, generally) the important question is not whether she is aware of the biases or could alter them, but whether or not those biases are indicative of whom she *really* is. The fact that the disassociation between an agent's explicit, self-reported biases and her implicit biases lessens when motivational factors like social desirability are controlled for would seem to tell in favor of the hypothesis that implicit biases are indicative of an agent's deeper character.

Implicit biases could be activated by the mere knowledge of cultural stereotypes, whether or not an agent endorsed them. If this were true, then implicit biases would not be indicative of an agent's deep character. Just knowing of the existence of a particular racial stereotype or that some people harbor negative attitudes towards a racial group would be insufficient evidence of anything objectionable in the agent herself, even if that knowledge

caused her to act in a discriminatory fashion. Of course, after becoming aware that they harbor biases because of their knowledge of common racist stereotypes and attitudes, and that such biases are harmful to particular racial groups, individuals might be racist if they do not take steps to excise those biases. They act with disregard for the victims of their discriminatory behavior. Once made aware of these biases and strategies for changing them, perhaps agents could be blamed for failing to change the features of their individual lives that make them more likely to act on implicit biases. And even if they could not, as individuals, change their behavior, perhaps they would nonetheless be morally required to advocate for institutional changes that would compensate the victims or mitigate the discriminatory effects of their bias. But the individuals would not, it seems, be racist because they harbored the biases or acted upon them, if mere knowledge were sufficient to give rise to them.

Thus, it is possible that, given what we currently know and suspect about the nature and cause of the implicit biases could be accommodated on traditional models of individual moral responsibility. Or, perhaps, those traditional models will need to be modified in light of the evidence from implicit bias. Additional empirical research will likely shed more light on this question and I do not want to rule out the possibility that, in light of further developments in the psychological and philosophical literature, including consensus on the nature of implicit biases, individual analyses may turn out to be correct. However, as I have argued above, there is one interpretation of the nature and cause of implicit biases for which traditional backward-looking (or ‘fault-finding’) models of moral responsibility cannot account. Neither the biases themselves, nor the individuals who harbor them, nor the behaviors they cause would properly be called racist. Thus, my aim for the remainder of this paper is to argue that even if this interpretation turns out to be correct, implicit biases are still

racist. In other words, *even if implicit biases turn out to be nothing more than automatic associative tendencies between concepts activated by mere knowledge of cultural stereotypes and the agents who harbor them are unaware that they have them and so cannot, prior to becoming aware, control them, implicit racial biases are still racist in that they stem from racist attitudes.* I argue that, under those conditions, implicit biased behavior would be a form of institutional, rather than individual, racism. This claim admits of several possible interpretations, which I will disambiguate in the following section.

#### **4. Implicit Bias as Institutional Racism**

To say that institutional analyses of implicit bias have been less pervasive than individual ones, is not to say they have been ignored completely in the current literature. Institutional theories, because of their close connection to justice (where justice is a feature of the arrangement of our basic institutions) are often refer to structures or systems. Sometimes it is assumed that simply because implicit biases contribute to racial inequities they must therefore be immoral or unjust; But, as I have argued elsewhere, that connection is too quick. More promisingly, feminist theorists have argued that racist stereotypes and attitudes are part and parcel of a dominant ideology of racial privilege that works to reinforce the status of those in power and perpetuate asymmetric power relations. Sally Haslanger (2000) for example, has argued that our race concepts, properly understood, just are expressions of power relations.<sup>51</sup> One's race is not just one's actual social position within

---

<sup>51</sup> Haslanger (2000) writes that "A group is racialized iff<sub>df</sub> its members are socially positioned as subordinate or privileged along some dimension (economic, political, legal, social, etc.), and the group is "marked" as a target for this treatment by observed or imagined bodily features presumed to be evidence of ancestral links to a certain geographic region (p. 44)."

that hierarchy but also includes prescriptive content.<sup>52</sup> It is inseparable from stereotypes and attitudes about that racial group.<sup>53</sup> It determines both how one ought to be treated and the meaning of the treatment one receives (Haslanger, forthcoming, ms., 30).

For example, members of certain racial groups are often, because of their racial group membership, deprived of full inclusion in the epistemic community (the community of knowers). Their epistemic authority is undermined in a variety of ways because of their racial group membership that marks them, among other things, as intellectually inferior. If justice requires (as it plausibly does) that individuals' access to knowledge not be restricted on the basis of their racial group membership, but also (and more importantly) that they not be deprived of the capacity for rational self-development (for which inclusion in the epistemic community is an essential precondition) then their exclusion is a grave injustice (Haslanger, forthcoming, ms., 19).

Importantly, racial subordination (or the creation of race) is not simply a matter of individual attitudes and beliefs. Centuries of material deprivation make it the case that 'equal' treatment with respect to race will nevertheless perpetuate these hierarchical relations. Moreover, even in cases where there are no biases, explicit or implicit, at work, one's racial group membership can determine the social meaning of an interaction (Haslanger, forthcoming, ms., 22-27). Actions will be interpreted through a dominant ideology of race as hierarchy, even if the individual actor does not act from racist attitudes,

---

<sup>52</sup>Haslanger (forthcoming) writes, "Racial frameworks situate groups delimited by 'color' in social hierarchy that carries presumptions concerning both moral and epistemic authority and a range of associated entitlements. These presumptions are enforced and enacted day to day between the races." (ms., p. 4)

<sup>53</sup> Haslanger (forthcoming) writes, "The social relations defining both gender and race consist in patterns of treatment towards bodies as they are perceived (or imagined) through collective frameworks of salience. Just as pink 'means' girl and blue 'means' boy in the contemporary US, the bodily markings presumed to be evidence of (relatively recent) ancestry in sub-Saharan African 'means' (among other things) intellectual inferiority..." (ms., p. 30)

conscious or unconscious. And individuals may not always be responsible for the knowing or controlling for the social meaning of their actions, even when these actions lead to injustice. So our concepts and background beliefs about races, our entire ideology of racialization, express reinforce and perpetuate unjust power relations for which no individual may be responsible.

On this view, implicit biases are racist because they stem from and contribute to the dominant ideology of racialization.<sup>54</sup> They help to maintain the unjust power structures. But they are only a small, and relatively unimportant part of the story. The larger picture is of individuals participating in institutions or structures that, without any racist attitudes or beliefs, conscious or unconscious, perpetuate racial injustice. Explanations of this sort are institutional or structural (as opposed to agential) because they posit these broad ideological forces as a (if not the primary) purveyor of racial injustice. Racial injustice is contained in the very concept of racialization and so built into the structure of our society.<sup>55</sup>

Even if we disagree with Haslanger that our race concepts are necessarily hierarchical, we could still agree that negative stereotypes and affective attitudes towards historically disadvantaged racial minorities are part of this dominant ideology. They are built into our language, social norms and customs and so, to the extent that they contribute to

---

<sup>54</sup> I use “racist” here as a synonym for racial injustice, though Haslanger and others would likely resist that move, reserving ‘racist’ to pick out the individual phenomenon.

<sup>55</sup> Haslanger rejects individualistic explanations of social phenomena in her Carus Lectures (2012) in favor of structural explanations, which focus on articulating the background ideology or schema (where a schema is a set of concepts and background beliefs that “help us interpret and organize information...and are the basis for various behavioral dispositions) (Handout from Lecture 2, p. 1). She applies that view to the education system arguing, “School...creates kinds of individuals through a process of discipline. It does not do this in a deterministic or mechanical way, of course, for part of the goal of the discipline is to form individuals who voluntarily enact the social structures that are to be perpetuated...[T]he process of racialization that occurs in schools prepares African-Americans for the subordinate status they can expect to occupy, and it does this very effectively. The ‘success’ of racialization is at odds with the explicit egalitarian ideology that many Americans hold dear, but this ideology does not hold the political, cultural or economic reins; and structures have pernicious effects without anyone intending or designing them (Haslanger, forthcoming, ms., p. 2).”

racial disadvantage, it may be the case that these harms are acted out by agents who, though they perpetuate this ideology, have no intention to act or awareness that they are acting in a racist way or perpetuating racial injustice, broadly understood.

This type of institutional analysis of the racism of implicit biases is quite different from the one I intend to draw on here. On the view articulated above, every part of a society that employed or expressed or otherwise participated in the creation and perpetuation of race categories with their corresponding power relations would be, in some sense, racially unjust. The racism is ‘institutional’ in the sense that it is not ‘agential.’ It is a property of a large, loosely-organized, societal-wide structure or system without aim or purpose beyond, one might hypothesize, its own perpetuation. However, even to say that is to take a step towards assigning agency to this structure that it importantly lacks.

Though I have argued against structural analyses like Haslanger’s as accounts of *racism* elsewhere, nothing I say here hangs on that argument (Mitchell). I submit that even if the concept of race is inseparable from racial attitudes and stereotypes and our culture is, as a result, shot through with racism through the perpetuation of a racial hierarchy that is the product of no one’s intentional action, there is still space to mark out agential racism (the bad attitudes and beliefs of individuals) as a particular sort of harm. Assuming that racism can also be expressed in the morally objectionable attitudes and beliefs of agents, the structuralist should be able to accept the analysis I offer here while still believing that there is work to be done by a broader, structural explanation of the phenomenon.

Thus, I want to focus on a different sort of institutional racism, where what I mean by ‘institution’ does not exclude, but instead relies upon, the concept of agency. By ‘institution,’ I intend to mark out discreet organizations that, I will assume for the purposes

of this paper, can be understood as possessing some sort of group or collective agency sufficient to ground moral responsibility.<sup>56</sup> This claim is obviously controversial, but I will not argue for it here.<sup>57</sup> Rather, I will assume that the sorts of group agents I will single out here, corporations and companies with organized structures and codified policies and procedures will count, if anything does, as group agents. That agency entails not only that they be capable of collective or group intentions (and thus, be able to perform actions that are properly thought of as theirs) but also that those intentions are appropriately subject to moral praise and blame. Institutions not only can act, but they can act immorally.<sup>58</sup>

---

<sup>56</sup> In other words, I will reject the methodological individualism embraced by Max Weber (1968), among others. But I will also steer clear of views on which mere aggregates of individuals or institutions can be said to act with intention. I take it that some organizational structure or other (I review several in the footnote 37) sufficient to, at the very least, transfer the intentions/attitudes of one individual so that it determines the behavior of another is necessary to give rise to group action (though not every case of intention transfer is an instance of group action).

<sup>57</sup> Like Cooper (1968), I take it as one piece of evidence in favor of group or collective action and moral responsibility that we often do talk as if groups can act, and be held responsible for those actions and that is difficult to see how the intentions, actions, and responsibility we assign to these collective agents could be broken down and analyzed in terms of individual action, intention and responsibility. This is not, of course, a definitive argument in favor of collective agency. We could simply be mistaken and/or employing incoherent concepts. Alternatively, groups may act as agents in a purely representation sense—a sort of Hobbesian Leviathan. There is some individual actor or corporate head to whom the actions are appropriately ascribed, and who bears moral responsibility for those choices, but he can be said to act for the rest of the group on account of their having transferred their wills to him. Group actions might, in some sense, supervene on the actions of the respective members of the collective, as does responsibility for those actions (Tuomela, 1989, 2005, 2006). Recent work by Pettit and List (2011) has made some headway in articulating yet another possibility for group agency, where groups can be organized in such a way that they are properly understood as having intentions that cannot be ascribed to any particular member of that group. They are irreducible group agents. For my purposes, any of these accounts of group agency is acceptable and I will not attempt to adjudicate between them here.

<sup>58</sup> Here, I think it is not inconsistent with my conclusion to think that, though organizations or institutions can act collectively, they do not possess moral agency in anything but a metaphorical sense. Thus, when we accuse a corporation of racism, it is possible that we are saying that its actions are of the sort that, were they performed by individuals, they would be morally condemnable as racist. If we are content to let this metaphorical sense of moral agency stand as a coherent and generally accepted use and employ it to praise and blame corporations and their actions, then I would be happy to accept that is all we mean when we say that a corporation is “racist.”



#### 4.1 The Agential Account of Implicit Bias as Institutional Racism

In this section, I argue that implicit biased microbehaviors (or IBMBs) are the actions of the set of major media organizations that constitute the modern American media landscape. By this I mean the entertainment and news media including, but not limited to, the companies who own and operate radio stations, television and film studios, magazines, newspapers and other forms of print media, as well as music production companies and record labels. While the particulars of the medium and content may differ drastically from one such organization to another, the basic aim of each is the same: produce media for consumption by the public. (I will hereafter refer to the aggregate of organizations that pursue this aim as ‘the American media industry.’)<sup>59</sup>

I argue that by using them in the content it produces the American media industry perpetuates awareness of common cultural stereotypes and vicious attitudes towards blacks. Having stipulated in the previous section that such an awareness is sufficient to give rise to implicit biases as associative tendencies in individuals, the media industry causes these biases to form in their consumers. The biases, in turn, cause those consumers to act out in the discriminatory ways described in the first section of the paper. I argue that the stereotypes and attitudes expressed by the media industry are racist and that consumers IBMBs, as a direct result of those racist attitudes and beliefs, are racist. However, I remain agnostic as to whether or not the corporations are at fault for the behaviors, and thereby

---

<sup>59</sup> The aggregate is not itself a group agent. They express similar attitudes and engage in similar behaviors making it appropriate to speak of them in a general way. But they do not possess the sort of relationships that I take to be necessary for group intentionality or moral responsibility. They act largely independently from one another. I will argue that implicit bias is a product of the racism of each, for which they are independently at fault. It may be that they also bear some collective moral responsibility as an aggregate, the way that Held (1970) argues that a mob might be collectively responsible for the harms they cause but this is not a line of argument I will pursue here. Consequently, when I speak of the American media industry I wish to indicate an aggregate of institutional agents each of whom is responsible, in some part, for the actions of implicit bias.

racist, or whether certain excusing conditions might apply that would mitigate their responsibility and thus, their racism.

The public that occupies the role of consumers within the organizational structure of a media production company is, in some sense, a merely hypothetical member of the organizational structure. Unlike other members, consumers are subject to no organizationally imposed sanctions if they fail to fulfill their role and consume the media produced. However, in another sense, their inclusion within that structure is very real insofar as the company directs its efforts towards influencing their choices and includes their preferences in its decision-making. They are unofficial but influential stakeholders in the corporate structure.<sup>60</sup>

In a more direct and obvious way, those media organizations also employ individuals in the roles of writers, directors, producers and other creatives. The individuals occupying those roles aim to produce content for consumption because that is the function of the role. So, a particular writer qua artist can intend to inspire through his art, but qua writer for a

---

<sup>60</sup> I acknowledge that my inclusion of ‘consumers’ within the institutional structure of media companies is controversial. Indeed, I think whether or not they count as a proper part of the institution will depend on the aims of a given company with respect to soliciting customer opinions and responding to information about what its customers want. If the actions of a media company are collective in the sense that they are the will of a single individual whom the rest have consented to follow or obey then tacit consent to the expression of the stereotypes by that actor must be read off the public’s consumption of that media. If the actions are collective because they are the product of voting procedures that arrive at a group intention then the ‘vote’ of consumers is likely hypothetical, as the opinions of the public who consume the media are not consulted as an official member of the production team. However, the opinions of consumers are often solicited through polls and screen testing, and in both entertainment and news media attempts are made to shape content and presentation to consumers desires. My tentative suggestion is that to the extent that this process is part of the policies and procedures of a company, the consumer is appropriately thought of as a role-holder in the organization’s structure. This makes it the case that some number of media companies who do not produce content with an aim of attracting consumers or who hope to attract viewers but have no procedures in place for monitoring and taking into account the desires of potential or actual consumers, might cause implicit bias by expressing stereotypes and negative attitudes towards a particular racial group but the IBMBs that result will not be properly thought of as institutional actions. Some IBMBs are caused by institutional expressions that are not behaviors of that institution and so are not institutional racism.

television company his aim, whether he acknowledges it or not, is to produce content for public consumption.

It is well established in the communications field that, in producing content for public consumption, writers or producers often exhibit stereotypes about and negative attitudes towards historically disadvantaged racial minorities, including blacks. At times, these producers merely mention such stereotypes and attitudes as when, in the course of covering the shooting of teenager Treyvon Martin, various news agencies reported that George Zimmerman could have held anti-black racist attitudes or stereotypes.<sup>61</sup> Frequently, however, they *use* these stereotypes, depicting blacks (and particularly black males) as criminal, lazy, unintelligent or aggressive in both entertainment and news media.<sup>62</sup>

When media outlets express anti-black stereotypes and attitudes (‘using,’ rather than ‘mentioning’ them) in their products, it is plausible to assume that they often (though perhaps not always) stem from the conscious or unconscious racist beliefs and attitudes of those creating the content.<sup>63</sup> The decision to focus on, for example, black men engaged in criminal

---

<sup>61</sup> My contention that using racial stereotypes is racist is importantly different from an argument that any mention of such stereotypes is racist. Mentioning stereotypes is, I think, allowable (after all, I have done so many times throughout this paper). And yet, mentioning stereotypes might also lead to implicit bias if awareness of common cultural stereotypes is sufficient to give rise to these biases. It seems to me that the dangers of creating a population with implicit bias ought to be weighed against the necessity of reporting on racism, including creative or historical pieces that employ these stereotypes to make salient the viciousness of racial prejudice, past and present. If mentions of racial prejudice are pervasive enough, it could be the case that the public continues to exhibit implicit bias as a result of their exposure to American media. But their behaviors would not be racist. They would not stem from the racist attitudes and beliefs of a discreet institution of which they are a part. So while individuals qua consumers would have the same associations and perform the same racially discriminatory IBMB, those behaviors would no longer be racist. I do not think this is as strange as result as it might appear. There are many behaviors that, though on the surface indistinguishable from each other, carry a very different moral weight when we discover that they are born of a vicious attitude or one of disregard vs. a conscious deliberative process.

<sup>62</sup> Just on the media portrayal of blacks as criminal see Dixon & Linz (2000); Dorfman and Schiraldi (2001); Sommers et al., (2006); Rome (2004); Covington, J. (2010).

<sup>63</sup> The frequency with which such stereotypes continue to be portrayed by the American media industry despite the widespread knowledge (and certainly industry common knowledge) that the engage in such behavior is one reason to suspect that those who are producing this content actually endorse such stereotypes. Another is the

activity is not an IBMB (IBMBs are microbehaviors or automatic responses) nor, I think, should it qualify as a case of non-culpable ignorance. Reasonably educated Americans, including news, television, film and radio writers and producers, have access to the data necessary to establish that blacks do not engage in more criminal activity than whites and that most crimes are intraracial, not interracial (Rennison, 2001). The pervasive portrayal of blacks as criminals, just one of many racial stereotypes and negative attitudes towards blacks by the media industry (and expressing the racist attitudes and beliefs of the producers) disseminates knowledge of the stereotype to the American public in their role as consumers. If, as stipulated, mere knowledge of cultural stereotypes is sufficient to give rise to implicit bias in the form of associative mental states, then the use of these stereotypes creates implicit biases.

By consuming media laced with stereotypes and negative attitudes towards blacks, more Americans (including children) form associations on which they unwittingly act. Irrespective of whether or not the *implicit biases* are racist, the causal chain linking the expression of racism by the corporation to the actions of the implicitly biased individual renders the discriminatory *microbehaviors* racist. While the biases themselves may be too minimal, as mental states, to be subject to moral evaluation, an institutional approach affords us the perspective that the IBMBs are importantly connected to racist cognitive or non-cognitive attitudes. These are not the attitudes of the individuals who harbor the biases, but those the content-producers who use racist stereotypes and attitudes and disseminate them to the general public. Because of this connection, the effects of implicit bias are racist even if

---

widespread explicit racism of many Americans. For example, an AP poll (2012) found that 51% of Americans expressed explicit anti-black racist attitudes.

the biases themselves are not. The harms they cause are not simply *racial* discrimination, but *racist* discrimination.

The institutional character of implicit bias is perhaps best captured by an analogy: Imagine a Senator who, acting on her antipathy for blacks, introduces a bill that would cause harm to that group and manages to get his bill passed as a law. An unwitting police officer is given the task of enforcing that law. Even if the officer is completely unaware that the law is an expression of the Senator's racism, or that it causes harm to a particular racial group, her enforcement of the law is, nevertheless, racist. She is acting out the racism of the Senator by enforcing his racist law.

The stereotypes and negative attitudes used by the American media industry are like a discriminatory bill written by a racist Senator. They express the racism of their creators while imbuing the consumers with implicit biases. Learning those racist stereotypes and attitudes gives rise to associative tendencies and so IBMSs. Like the police officer charged with enforcing the racist Senator's law, we all, qua consumers, carry out the racism of the American media industry when we perform IBMB. As individuals we are not racist (at least, not for carrying and acting on these implicit biases) just as a police officer need not be racist to commit a racist act when she enforces a racist law. She is carrying out the actions of a racist institution.

Naturally, there are some significant differences between the two cases: First, if implicit biases are caused by racist expressions in the media, the 'laws' that we follow are unconscious. Though consumers with implicit bias have knowledge of certain stereotype or attitudes, they are unconscious that those stereotypes hold force for them; they do not realize that they have, in some sense, internalized the association and will act upon it. This would be

like a police officer who is aware of a certain law but does not take herself to be in the business of enforcing that law. In fact, she might have explicitly decided not to enforce it. And yet, without realizing it, she continually does enforce it. But, that this would certainly be unusual behavior for any rational person to engage in on the macro-level does not mean we should reject this analysis of IBMBs. Instead, if anything, it should make us more wary of the potential effects that our media consumption can have on our actions.

Second, in the case of the police officer, every instance of her enforcement of a law is racist because her behavior expresses the Senator's racist attitude. The link between the attitude and the behavior is clear. If implicit biases are the mere consequence of knowing cultural stereotypes then any number of institutions and individuals could be responsible for any particular agent's implicit bias. So why think IBMB are the behaviors of the media industry rather than some other institution or individual? Why think that they are institutional at all if one could learn a cultural stereotype from a friend, a grandparent, or some random stranger on the street?

Unfortunately, given the myriad of ways on which the public is exposed to these stereotypes, it would be difficult, if not impossible, to prove the sort of direct connection between the media's use of stereotypes and the harms caused by these stereotypes that would definitively establish causal responsibility. There is no necessary connection between the development of implicit racial biases and the American media industry. But, imagine that we could establish that, as a matter of fact, the American media are major purveyors of anti-black racist stereotypes and attitudes and so, have had a large impact on the population's general awareness of negative racial stereotypes and attitudes. Certainly, given the ubiquity

of such stereotypes and attitudes in the media, and Americans' exposure to media outlets, it is plausible that they cause or perpetuate these biases.

In the end, the argument that media outlets are causally responsible for implicit biases might depend on what we mean when we say that implicit biases result from *awareness* of common racist stereotypes or attitude. Is it enough that someone hears a stereotype once to develop an implicit bias? I recall a college friend once told me that his uncle firmly endorsed the claim, "All Yankee girls have fat ankles" (a claim I am assuming is a false generalization, though I have no empirical evidence one way or another) but, I would be surprised to learn that on the basis of just becoming aware of that stereotype I now more strongly associate 'Yankee girls' with 'fat ankles' than any other group.

So, instead, we might argue that the media are causally responsible for implicit biased behaviors along the same lines that some claim they are responsible for making extreme thinness part of our common cultural standard of beauty or normalizing violence among American youth. There is a causal connection between those phenomena and the media industry's almost exclusive portrayal of extremely thin females and its casual approach to violence. Though the media are not solely responsible (they may not even be the originator of the stereotype or attitude) our continuing familiarity with these stereotypes and attitudes (especially at a young age) is due (in some significant part) to the ubiquity of their expression in the media.<sup>64</sup>

Of course, for the IBMBs to be racist, the attitudes that these institutions express must actually be racist and not just unfortunate. What if a content-producer's decision to include stereotypical depictions is not born of racism but is just a decision to show people what they

---

<sup>64</sup> These comparisons are not meant to convince the skeptic but to sketch a model for how one might go about arguing that IBMBs are the racist behavior of a modern media industry.

want to see? One can certainly imagine a studio executive arguing that while he is aware that blacks are no more criminals than whites and bears no ill will towards black people, his duty is to his bottom line. Consumers, consciously or unconsciously, want to see these racist stereotypes and attitudes expressed and they are more likely to consume media that presents them. The racism, he argues, is theirs not his.

To show the racist attitudes properly belong to the media industry, it must be the case that using these attitudes is racist regardless of whether or not the attitudes themselves are actually endorsed by the producers of the content. This argument is, I think, fairly intuitive and does not depend on showing that these stereotypes and vicious attitudes have harmful consequences (though they do): Imagine a scenario in which a group of colleagues are telling racist jokes. None of them endorse the attitudes or stereotypes the jokes' express, but they are neither telling them with either a sense of irony (which some might argue makes such jokes excusable) nor just reporting the joke. They are intending to elicit a laugh from the joke, not at the outrageousness of the stereotype it expresses, but at the racial group who is its subject. Now, further imagine that no member of the racial group will ever hear the joke and telling the joke will not cause any of the worrisome psychological effects that are the subject of this paper. I find it plausible that those telling, listening to, and laughing at the joke are all displaying an attitude of racist disregard to members of the racial group that is its subject. Even if one does not endorse the stereotype, using inferiorizing beliefs about or vicious attitudes towards a particular racial group, whether it is to get a laugh or sell one's product, is racist.

Nor does the caveat that one would use any racial or social group to get a laugh or sell products make it the case that one is not a racist. First, an attitude of disregard toward all



groups does not make it the case that one is not displaying disregard toward one. Second, given the long history of racial injustice towards blacks in this country, equally bad treatment does not express equal levels of disregard. One must be more careful with one's words and actions, not simply because of the consequences of those words and actions are different for different racial groups (though these consequences are also important) but because, given our racist history, the same words and actions express different attitudes depending on the racial group that this their object (provided that one is aware of that history). A colleague's insensitive habit of remarking on others' weight expresses a far more callous and coldhearted attitude when he does it to a person he knows struggles with an eating disorder than when he does it to someone who does not. Thus, it is not enough that a studio executive could claim that he would use any racist stereotype or attitude to sell his product. The ones he does use are still racist.

To say that IMBMs stem from the racist attitudes and beliefs of the content creators is not yet enough to show that the media industry is responsible for them. While the behaviors may be racist, perhaps the industry itself is not. The corporations that make up the media industry might express racist attitudes but not be at fault. As with individuals, the moral responsibility of institutions for IBMB would likely hang on questions of their awareness, control, and what we might call "centrality" (an analog to the individual case of "deep character") of the racist attitudes they express. Regardless, they are not excused simply because they may not be aware that their racist expressions cause implicit bias. As an individual, I am likely morally responsible for the consequences of expressing racist beliefs and attitudes even if I did not mean to cause those consequences, so long as I am responsible for the expression itself. And, unlike individuals with implicit bias, the industry can control

whether or not they express those attitudes and thus, influence whether or not IBMB occur as a consequence of their use of racist stereotypes and attitudes.

Even if the industry is racist, perhaps they have no moral duty to curb their racist behavior. For example, while their duty to their bottom line does not render the stereotypes and attitudes they express non-racist, it might still trump a moral duty not to show them. They are like a mortgage company that requires all home loan applicants to have a credit score above a certain threshold in order to be approved. Blacks, who for reasons of past racial injustice are less likely than whites to meet that threshold, are therefore less likely to receive home loans. However, the mortgage company would do worse and perhaps even fail, if they offered home loans to applicants with lower credit scores. They cannot, it would seem, be forced to shoulder the burden of past racial injustice if doing so would contribute directly to their failure as a company. Similarly, a media company who exhibited a greater diversity of female figures or refused to air gratuitous violence might be acting well, but, one might argue that their doing so can hardly be morally required if it interferes (to some extent of another) with the aim of making money.

I think, however, there may be an important disanalogy between the two cases. With respect to the mortgage industry, it is at least plausible that there is nothing intrinsically wrong with using credit scores as a means of differentiating between acceptable and unacceptable loan applications.<sup>65</sup> Were it not that past racial injustice renders the process unfair for black applicants, we would have no objection to the mortgage company (or the industry as a whole) engaging in this practice. But perhaps there is something intrinsically wrong with a media industry portraying only extremely thin women or showing gratuitous

---

<sup>65</sup> I say “plausible” so as to leave open the possibility that a thoroughgoing critique of capitalism could show this practice to be intrinsically morally reprehensible.

violence. And, I think it quite likely that there is something intrinsically wrong in using anti-black racism to sell a product. It is, as I argued above, at least indicative of an attitude of disregard.

So, it may be that the industry is morally culpable for the racism it expresses. Regardless, I did not set out to prove that claim. My aim was only to show that the behaviors of implicit bias are racist even if implicit biases are mere associations caused by knowledge of common cultural stereotypes. Though, on an institutional approach, the implicit biases themselves are not racist (they are too minimal as mental states), the behaviors they directly cause (IBMBs) are. They are the racist behaviors of an institution. Vetted, approved and eventually produced the attitudes these behaviors stem from are the expressions of American media companies. Individuals qua consumers act out the racism of institutions of which they are a part.

## **5. Conclusion**

So, given what I have said here, the next obvious question is what, if anything, ought to be done about implicit bias? To some extent, this will depend on whether or not implicit biases actually are mere associative tendencies caused by awareness of common cultural stereotypes and racist attitudes or if individuals are at fault for their own biases. If implicit bias is a form of institutional rather than individual racism, then new avenues open up for thinking about and dealing with its discriminatory effects. For example, one might argue that racist stereotypes and attitudes ought to be excised from our media entirely, given their potentially harmful effects. Or maybe they ought to be regulated, with warning labels placed on media that contains racist stereotypes and attitudes, just as some have argued we should regulate violent video games or stick-thin models. However, in the case of racism, many of

the products infected with racist stereotypes and attitudes are classic works of literature, music and film. How should we handle those cases? Equally important are considerations of corporate responsibility. Going forward, are media outlets responsible for fixing the racism they have perpetuated or as individuals are we responsible for our own consumption? To say that the implicit racial bias is a form of institutional racism is not, after all, to abdicate all individual responsibility. The police officer who realizes she has been enacting a racist law can refuse to enforce it. She can resign. As consumers, perhaps we have a similar responsibility to resign from participation in mass media in order to stop acting on implicit racial bias. These are, no doubt, tricky questions and ones that ought to be thought through carefully. But knowing that implicit biases are racist, regardless of what they are and where they come from, will aid in that process. When we are tempted to throw up our hands in frustration at the pervasiveness of the problem and its seeming insurmountable, it is helpful, I think, to know that our struggle is not just against some trick of human psychology or insoluble structural injustice, but it is a fight against racism.

### **3. The Search for Formal Equality of Opportunity: A Case for the Preferential Treatment of Black Americans**

#### **1. Introduction**

Andre and Wesley, current U.S. high school seniors, are applying for federally funded college scholarships. They were admitted to the same school with the same SAT scores and extracurricular activities. But Andre has a slightly lower GPA than Wesley. On one form, they are asked to identify their respective racial backgrounds. Andre, who is black, checks one box, while Wesley, who is white, checks another. Despite his lower GPA, is Andre more likely to get the loan because he is black? And if so, should he?

Arguments that seek to answer this question in the affirmative, presenting an argument in favor of preferential treatment for black Americans, fall into one of two basic categories: the backward and the forward looking. Backward-looking arguments are generally compensatory; They are intended to repay (or partially repay) blacks for racist harms they have suffered. Forward-looking arguments are, as the name suggests, aimed at securing future goods. They aim to bring about a better, more just society through preferential treatment. Both present substantial difficulties. The backward-looking approach has theoretical and practical problems that make it difficult to clearly articulate and implement. Meanwhile, forward-looking arguments tend to focus on attaining controversial

social goals and in doing so, lose the support of many who, though liberal, disagree about both the nature and relative importance of freedom and equality. Moreover, forward-looking approaches tend to send the wrong message about preferential treatment to whites and blacks alike because they focus on ends that are only tangentially related to preventing anti-black discrimination.

In what follows, I will present a forward-looking argument that avoids these two complaints. I will assume merely formal equality of opportunity (rather than any more substantive form of equality), a minimum principle to which most (though not all) liberals would assent.<sup>66</sup> This principle is often taken to ground anti-discrimination laws and policies but not preferential treatment. I will show how it can be extended such that it necessitates the latter as well. By minimizing these substantive commitments, my hope is to bypass some of the theoretical problems often associated with forward-looking arguments including disputes over the content and demands of freedom and the role of equality within a theory of justice. My sense is that convincing as many people as possible that, in light of their own ideological commitments, they ought to support affirmative actions is likely an effective strategy for inducing social change. And, if this minimal requirement is all that is needed, then to deny blacks a presumption in favor of preferential treatment, one would have to deny the principle of formal equality of opportunity. Furthermore, because the forward-looking argument I

---

<sup>66</sup> The argument I make in these pages is heavily indebted to recent work on the value of equality in liberal societies and the need for racial integration in order to achieve that ideal, particularly Anderson (1999, 2002, 2010). I think sketches of the argument I offer here can be found in several places in her 2010 work *The Imperative of Integration*. If the argument is indeed prefigured or even explicitly made in Anderson, my modest aim is to draw it out and make it more explicit in hopes of showing exactly how significant of a contribution it is to the literature on affirmative action. By connecting it with Hill's work on affirmative

action's message, I also aim to draw out how it importantly differs, in spirit and message, from other justifications Anderson offers for affirmative action, especially those in which she regards it as a tool for promoting a (generally) more democratic civil society.

offer concentrates on securing protection for blacks against racist discrimination, it is not vulnerable to the critique that it expresses a disregard for concerns that are of central importance to the black community. Finally, I show how this argument can be extended. It is not only a critique of colorblindness but also an argument against certain forms of (seemingly non-objectionable) statistical discrimination.

### **1.1. The Backward-Looking Approach**

If the two had exactly equal qualifications, we might think preferential treatment for Andre falls straightforwardly out of principles of redress and the reality of past anti-black discrimination. When we consider that there were times in the not-too-distant past when any indication of Andre's racial background would have disqualified him from a great number of opportunities (including college admission and scholarships) it seems only fair that Andre should be first in line to get the money. Despite the fact that Andre is, on some metric, a slightly less qualified of a candidate, it still seems he is owed something because of the harms that were perpetuated against black people, including his ancestors. Moreover, we might reasonably assume that, had Andre's ancestors been afforded all the same advantages as Wesley's ancestors, Andre might have had the same GPA as Wesley. Therefore, we might conclude that he is just as deserving of the money on grounds of merit and entitled to it in order to compensate for past injustice.

Unfortunately, in studying the case a little longer, two complications arise that might mitigate the force of our initial intuition.<sup>67</sup> First, let's assume that these scholarships are a

---

<sup>67</sup> I will set aside some other complications including who donated the money for the scholarships, whether the university is public or private (and whether or not that matters) and if the effects of a more diverse student population would be positive or negative. These issues are not unimportant and they are further indicators of exactly how theoretically and practically exhausting backward-looking arguments can be. My hope is that the

finite resource such that, if Andre receives the money, then Wesley does not. Though it is perhaps true that some of Wesley's success (including, we might think, his GPA) is attributable to the many advantages of white privilege that he and his family enjoyed which Andre's family did not, Wesley himself had nothing to do with the racism Andre's family endured. Why should Wesley lose out on an opportunity in order to make up for harms that he did not cause?

By way of analogy, imagine that instead of competing for a scholarship, Wesley and Andre are running a race. If a spectator jumped out and tripped Andre, causing him to lose the race, it would be difficult to deny that Andre was harmed. But if to make up for the harm done to Andre we were simply to award him 1<sup>st</sup> place in the race, then it seems Wesley might also have the standing to argue that he has been harmed by us. After all, he did not trip Andre. Consequently, if it is true that Andre is owed redress for harms perpetrated against his ancestors then giving him the scholarship might be a way of trying to make up for those harms. But we might still be left feeling that Wesley has been treated unfairly. We might even worry that his right to equal opportunity has been violated. It is he that has been judged for the color of his skin, rather than the content of his application.<sup>68</sup>

---

two complications I present introduce sufficient worries for the backward-looking approach without having to explore additional complexities.

<sup>68</sup> There are a number of disanalogies between the two cases that might undermine the intuitive strength of this objection. For example, college scholarships, unlike most races, can consider any number of factors in determining who the winner should be. The rules governing scholarships are generally more complicated and less universal than those governing races. So, we are not actually taking from Wesley when we give the scholarship to Andre. Wesley was never entitled to the scholarship. Those who fight this claim usually argue that scholarship competitions (or hiring and admissions decisions) ought to be decided purely on the basis of (what they call) "merit." Like races, the rules ought to be such that those who meet some general criteria will win, regardless of their race. They forget that affirmative action is actually a way of doing just that; Affirmative action is one tool for equalizing the centuries of privilege that white people have enjoyed, often at the expense of blacks. Like a "head start" in a running race, preferential treatment can be a way of making the fight fair and giving each opponent an equal change at winning.



Second, even if we were to decide that the effects of past racial injustice, irrespective of Wesley's participation in it (and thus, Andre and Wesley's racial group membership) are a relevant consideration in who should get the scholarship, what should we make of the claim that *Andre* might have done better if he and his family had not been victims of racist laws? Even if we can find some actually culpable person or institution to offer up compensation for past harms (a tricky business in itself), how can we know for sure that Andre, *in particular*, was harmed by past racial discrimination in such a way that he is owed this scholarship? His lower GPA might not be the result of racial discrimination but some other, non-redressable harm (or perhaps, no harm at all). Maybe Andre would not have been better off had his family received all the same benefits as Wesley's family. Andre's ancestors might have squandered the money. After all, there are plenty of white families who did have all of the same advantages and did worse than Wesley's family. Proving the connection between a particular black American's current socio-economic status (or, GPA) and past racial injustice is especially difficult in cases where a great deal of time has passed between the injury and the attempt to offer reparations. The counterfactuals ('had person A not been a member of oppressed group Y then she would not have suffered harms r, s and t...') are likely to be difficult to evaluate. If we were to wholeheartedly endorse compensatory preferential treatment for black Americans we might end up rewarding those who do not deserve it in addition to taking from those who did nothing wrong.<sup>69</sup>

---

<sup>69</sup> In a somewhat related worry, philosophers with such divergent views as Goldman (1976) and Nagel (1973) argue that compensatory arguments for race-based preferential treatment programs like affirmative action are unjustified because they reward the most qualified and so primarily help those who were least harmed by past racial injustice. However, as Boxill (1992) points out, the least harmed members of an oppressed racial group could still be badly harmed by past racial injustice. Though Goldman and Nagel's critiques might justify expanding compensatory preferential treatment, they do not give us reason to abandon it.

## 1.2 The Forward-Looking Approach

So, even if the compensatory argument is theoretically straightforward, it is beset by a number of practical complications that are unlikely to be resolved.<sup>70</sup> Recognizing this, we could consider a different strategy. Perhaps, Andre deserves preferential treatment not because he was harmed, but instead because such preferential treatment is the best way to achieve a just society now. As liberals, we share a general commitment to the value of individual freedom. It could be that, in order to be truly unconstrained, members of a society must be given an equal opportunity to develop their capabilities and preferential treatment for Andre is a means to achieving that aim.

This sort of forward-looking argument could assert that since black Americans were excluded from many opportunities in the past, society is currently less free than it otherwise might be. The best way of achieving a more free society now is to engage in affirmative action and offer the scholarship to Andre. This view is forward-looking because, though it recognizes that racial harms occurred that lead to current inequality, those harms do not ground the preferential treatment. Rather, preferential treatment for blacks is justified solely because it is expected to produce good outcomes, namely increases in individual freedom. If

---

<sup>70</sup> By which I do not mean to say that they are *irresolvable*. I am generally sympathetic to backward-looking arguments for affirmative action and unconvinced by the worry that some blacks who were not directly harmed by slavery or segregation might end up unfairly benefiting from a reparations schema and whites who were not at fault might end up losing advantages they earned. But I suspect that lack of concern probably stems from a general commitment to principles of distributive justice and not from a fully worked out response to this position that does not depend on any substantive views about the nature of justice that those who endorse it could not accept. And, as my goal is here is to convince those with radically different views about the nature of justice that by their own lights they ought to endorse affirmative action, relying on my own commitments to dismiss this worry would run counter to my aims. However, that is not to say that an argument that does defeat these objections on the objector's own terms is not possible. In fact, I believe that Boxill (1992) offered one, as does Ta-Nehisi Coates (2014) in his recent article. My larger point is that haggling over compensatory schemas, as this sort of objection invites us to do, might be philosophically exhausting and pragmatically pointless.

it turned out that offering the scholarship to Wesley and other white males would lead to better outcomes than preferential treatment would be given to them instead.<sup>71</sup>

But this strategy, too, is beset by a number of complications. Is it true that individual freedom requires this sort of substantive equality? And if so, how much and what kind of equality is required to achieve individual freedom? And what is freedom anyway? Even among those who agree about the priority of freedom, there is substantial divergence with respect to what that agreement entails. And though nearly everyone will assent to the claim that the just state treats its citizens equally or with equal respect, it is far from clear what equality requires.

Elizabeth Anderson (2010), for example, has recently argued in favor of affirmative action for blacks as a tool to promoting integration. Against luck egalitarians, she claims that in the equal distribution of resources, the state should not make exceptions for bad circumstances that are an individual's fault (see note 12). Justice in a democratic society requires social equality, meaning that citizens must be able to stand and relate to one another as equals. To do this, individuals must be guaranteed a minimum standard of living (you cannot, for example, stand before others as an equal if you cannot afford clothing or food) and equality of opportunity to secure positions beyond that minimum.<sup>72</sup> Without these equalities in place, racial segregation will continue, racial stigmatization and discrimination will necessarily ensue, and a great number of injustices, racial and otherwise, will result from both segregation alone and discrimination as its consequence.

---

<sup>71</sup> This point is made especially clear by Hill (1995).

<sup>72</sup> As Anderson (2010) writes, "The distinctive normative feature of democratic societies is social equality. All of the members of a democratic society have a just claim to stand in relations of equality with their fellow citizens" (p. 17).

But is Anderson right? Is a minimum standard of living actually required to guarantee that citizens enjoy democratic equality? And should we even be aiming for democratic equality? The range of possible positions on the value, nature, and extension of equality and the myriad of disputes between those who agree on many of the major points, including the fact that blacks are entitled to some sort of preferential treatment in order to achieve a given social aim, is both distracting and potentially debilitating. Once entrenched in this debate, we find we are only nominally discussing whether Andre is entitled to preferential treatment. Instead, we are caught up in a deep (and seemingly intractable) dispute about the nature of justice, which is unlikely to be resolved in favor of one side over the other anytime soon.

We might be tempted to dig in here and spend our remaining philosophical strength spelling out and defending a particular conception of freedom or equality that would entail preferential treatment. But in so doing, we risk losing sight of the fact that debates over preferential treatment for black Americans are not merely (or even primarily) a matter of abstract political principles but calls for policy changes that have the potential to drastically affect the life chances of actual individuals. Moreover, even if we could agree on a particular social ideal that we hope to achieve (a seemingly impossible task in itself) we might also worry about the message that a purely forward-looking approach sends to blacks and whites alike. As Thomas Hill (1995) points out, concentrating only on the overall social benefits of preferential treatment might express a lack of concern for the suffering of black Americans like Andre and little confidence in his abilities as an individual. It says to him: you are receiving this award not because you deserve it but because giving it to you is likely to bring

about some good for all of us. To Wesley it says that, though he has not been found undeserving of a scholarship, his desires must be sacrificed for the good of all.

Given the problems with both approaches, it would be nice if we could find an alternative; Either a backward-looking approach that doesn't rely on establishing the truth of a complicated set of counterfactual claims or a forward-looking justification that does not send the message to blacks or whites that they are being used as a means to some greater societal end, which they may or may not think is an appropriate end. Instead, while our forward-looking argument aims to achieve a good for all, it should do so in such a way that the clear message sent to black Americans is a concern for their liberty, in particular, not just democratic ideals generally.<sup>73</sup> Furthermore, the aim should be one to which a large swath of liberals can agree. Without ruling out the possibility of the former, I will concentrate my energies on developing the latter—a forward-looking approach that takes racial justice to be of chief concern and appeals to liberals across the political spectrum by requiring little in the way of controversial ideological commitments.

## **2. Preferential Treatment for the Vulnerable**

To arrive at the sort of forward-looking argument we desire—one that takes racial justice for blacks as the primary social good secured by preferential treatment—we must first show that preferential treatment is required to treat a certain kind of racial injustice against blacks. In this section, I will identify the conditions that make individuals initial candidates for preferential treatment (they are vulnerable) and then identify one type of vulnerability

---

<sup>73</sup> Hill (1995) has suggested that we look at the *message* of our affirmative action policies as an alternative to forward and backward looking proposals—What do we want them to *say*? I wholeheartedly endorse this approach and nothing I say here is meant to contradict it. My aim is to throw into the ring another alternative with broad political appeal and my hope is that expresses a more respectful message than more traditional forward-looking approaches and one that is more likely to be well-received by both blacks and whites.

that justifies that treatment. In the section that follows, I will argue that blacks are members of this type of vulnerable group.

On its face, the justification for preferential treatment can appear fairly simple. Plainly, not everyone enjoys an equally level playing field with respect to attaining her conception of the good. Some of us are more constrained than others. Indeed, all of us enter the world less well off than the adults around us who generally have use of their rational capacities and the ability to communicate their needs to others. As infants we are vulnerable insofar as we suffer from conditions that render us helpless in scenarios in which adults likely are not. The same is often true of the elderly or physically infirm. For example, assuming most of us share the goal of not suffering a stroke, it is the elderly who are more likely to succumb to one during an unexpected heat wave. Because they are, by and large, less physically robust than younger people, the elderly are more vulnerable to certain sorts of harms. Consequently, a fire department might extend preferential treatment by keeping a list of elderly people in the neighborhood and checking on them first during a heat wave. If justice requires that, as much as possible, all persons have an equal chance at attaining a reasonable conception of the good then preferential treatment is justified for those who suffer from some deficiency that lowers the likelihood that they will succeed.

These individuals, who are more susceptible to future harms (or, if you prefer, less likely to be able to attain their conception of the good) because they experience some condition that most human beings do not, are the vulnerable.<sup>74</sup> That condition need not

---

<sup>74</sup> 'Vulnerability,' then, is a statistical relation. It makes a claim about the goods or abilities a particular person or group enjoys relative to the rest of the population and takes for granted (at least for our purposes) a certain set of reasonable conceptions of the good. So, in general, American children are less vulnerable relative to African children, though all children are vulnerable, in some sense, relative to adults. Michael Kottow (2003) distinguishes this population from the vulnerable, referring to them as 'susceptible.' He argues that all human beings are essentially vulnerable and that vulnerability is the basis of the state's protection of human rights: "In a nutshell, the vulnerable are intact but at risk, in the same way a fine piece of porcelain is unblemished but

actually interfere with their ability to contribute to society or pursue their conception of the good. Rather, their vulnerability is a matter of the likelihood that they will be frustrated in their goals. By way of analogy, we might imagine a community whose houses are all built into the side of a mountain; the vulnerable are those whose houses stand nearest to the edge. It is they who are most likely to slide into the abyss if special care is not taken to protect them.<sup>75</sup>

When preferential treatment is justified, it is generally in response to such vulnerability. Preferential treatment is a way of protecting a group against the harms they are more likely to suffer because of an injury or impairment. In a sense, then, the name is somewhat misleading—“preferential treatment” for vulnerable groups is meant to raise them closer to, not above, the level of the non-vulnerable. However, the use of “preferential” does capture an important distinction between this and anti-discrimination laws and practices. Preferential treatment is not the mere extension of a particular treatment to a new group but the recognition that some greater level of care is required in order to achieve an equal outcome—to ensure that the vulnerable group is equally unconstrained. It is a way of moving their houses back from the edge, or helping to reinforce them against calamity.

Vulnerability can pertain to individuals, but also to socially salient groups. In many cases, when we identify a vulnerable group, the individuals in that group will merely exhibit a tendency towards experiencing some condition. The elderly, for example, have a tendency

---

highly vulnerable to being damaged. The susceptible are already injured, they already suffer from some deficiency that handicaps them, renders them defenceless and predisposed to further injury; their wounds lower the threshold to additional suffering” (p. 464). While I do not wish to deny that the human condition is a vulnerable one, I want to preserve the term here to refer to those in a special position of susceptibility *vís a* *vís* the majority.

<sup>75</sup> This ‘cliff analogy’ is adopted from one used by Jones, et al. (2009) in discussing the social determinants of children’s health, including racism.

towards comparative physical weakness while the young have a tendency towards an underdevelopment of rational capacities. There are exceptions in most vulnerable groups. However, we pick out these individuals and identify them as vulnerable by their group membership because there is a greater likelihood that a member of the group will experience the relevant condition.<sup>76</sup> Perhaps, given constraints of time or resources, we cannot realistically differentiate between those group members who actually have the condition and those who do not, so we rely on the fact of their group membership. Or, perhaps the mere fact that there is that tendency towards a particular condition is enough to make them susceptible to future harms. So, for example, although a particular elderly person may be in fine physical health the general tendency of elderly people to poor health justifies identifying the elderly as a vulnerable population. Even the most robust elderly person might be more likely to suffer a mugging because the general tendency of the elderly towards physical infirmity marks her out as a target for muggers. But demonstrating that a population is vulnerable is not, on most theories of liberal justice, tantamount to demonstrating that they are entitled to preferential treatment. Intuitively, the vulnerability must be of the right kind. It must make them likely to be prevented from receiving something to which we agree that they are entitled, such as a basic liberty.

---

<sup>76</sup> In many instances, the vulnerable group's tendency towards a particular injury or impairment is causally related to its group membership, but that need not be the case. The elderly are more likely to be physically weak because of problems related to aging but Iowans might be more likely to develop cancer without their living in Iowa bearing any causal relation to that tendency. In such cases, it would not make sense to identify Iowans as the vulnerable group. There is nothing about Iowa that contributes to their tendency to develop cancer. However, as I will argue later in the paper, even in cases where there is no causal connection, groups can sometimes have a social salience that makes it the case that the larger group, and not simply the population that is actually injured, is properly identified as vulnerable. In these situations, it is the existence of the tendency itself combined with the salience of the social group that creates a causal connection between an individual's group membership and the likelihood that she will suffer future harms.



## 2.1 Formal Equality of Opportunity and the Problem of Distinguishing Vulnerability

As we have seen, part of the argument for justifying preferential treatment rests on identifying an individual as a member of a vulnerable group. She must be part of a group that exhibits a tendency to some condition that makes her susceptible to future harms. But vulnerability alone is not enough; the vulnerability must be of the right sort. Intuitively, not every inequality deserves to be rectified or regulated against and there is substantial disagreement between liberals with respect to this issue. In general, everyone agrees that when a current vulnerability is the result of a definite past injustices, the individuals who suffer ought to be compensated so that the vulnerability does not interfere with their ability to achieve their aims. But given the practical problems involved in securing preferential treatment for black Americans on those grounds, we have stipulated that our account will be forward-looking. So what sort of vulnerability, experienced by black Americans, could inspire a wide swath of liberals to agree that preferential treatment is justified? What basic liberty are black Americans are unfairly denied?

One way to draw a distinction between those inequalities that would be acceptable for most liberals to regulate against and those that would not is by asking what caused a group's vulnerability. Imagine two groups, one whose members choose to build their houses close to the edge of the precipice because they enjoy the view and another whose members build there because they have no alternative. The first group, we might think, is not entitled to our help. They have chosen to make themselves vulnerable and so any consequences they experience are not our concern.<sup>77</sup>

---

<sup>77</sup> The intuitive idea behind this distinction is that, taking liberty as the primary good to be secured for each, intervention by the state (including preferential treatment) is justified only to the extent that it increases liberty. However, what counts as an interference to one's liberty is a subject of much debate. Most, though certainly

But many of us have the intuition that, at least in some circumstances, members of the second group would be entitled to our help.<sup>78</sup> For example, imagine that those who live on the edge are forced to do so because they are members of a despised social group, scorned for no morally defensible reason. Certainly, *some* intervention is justified to protect those individuals from the threat of harms caused by others. They are vulnerable for reasons that have nothing to do with their individual merit. The constraints they experience as a result of their condition are not imposed by nature, but by other agents.<sup>79</sup> So, to protect them from the condition they suffer (membership in a despised social group) we introduce formal equality of opportunity. We hold that, as John Rawls (1999) puts it, “all have at least the same legal rights to all advantaged social positions” (p. 62). Of course, this does not yet introduce any sort of preferential treatment. It is merely a guarantee of equality under the law.

There are, however, problems with adopting this position. The most important for our purposes is that if we accept that markets ought to be regulated to rule out disadvantages or constraints imposed by other agents (to make it the case that careers are actually open to natural talents) then we might not be able to stop at guaranteeing mere formal equality of opportunity. As Rawls (1999) famously argues, the same principles that lead us to endorse formal equality of opportunity ground a more substantive alternative, which he calls

“equality of fair opportunity” (sect. 12). Recognizing that the extent to which individuals are

---

not all, liberals want to rule out state interference to correct for inequalities resulting from voluntary choices. Proponents of this view include “luck egalitarians” like Roemer (1994) Nagel (1991), Dworkin (1981), Cohen (1989) and Arneson (1997) as well as negative liberty theorists like Kramer (2003) and Carter (1999). Anderson (1999) who defends “democratic equality,” is a notable exception to this view. She argues that labeling some individuals as “at fault” for their misfortune and denying them aid on that basis is disrespectful.

<sup>78</sup> Liberal theorists would diverge with respect to whether individuals ought to be protected against or compensated for the effects of bad luck, generally, or only harms that are product of another’s intentional action.

<sup>79</sup> Kramer (2003) flags the intuitive force of this distinction though he admits there may be no non-question-begging way to establish it.

able to cultivate their natural talents is largely determined by accidents of their birth (the economic and social position of their parents, for instance) and that those “accidents” (really, the intentional and unintentional actions of other agents) place significant restraints on the freedom of individuals, merely eliminating formal sources of discrimination will not actually achieve the goal of awarding opportunities on the basis of natural talent. Instead, it appears that actually ensuring formal equality of opportunity would require much more drastic measures than intervention into the private choices of individuals and a large-scale redistribution of wealth. If individuals would endorse a policy of formal equality of opportunity because, given the constraints of the original position, they decide that allocating desired positions on the basis of individual merit (and not by the arbitrary, and perhaps immoral, preferences of others) is the most effective means of maximizing their primary goods, then they ought to endorse an equality of fair opportunity that aims to correct the many inequalities that interfere with the development and exercise of individual talent.<sup>80</sup>

Those liberals who accept Rawls’ line of argument from formal to substantive equality of opportunity, but balk at his conclusion, might be tempted to reevaluate the force of the initial intuition that some forms of discrimination ought to be regulated. If we cannot draw a clear distinction between discrimination on the basis of characteristics such as race or gender, which seem clearly imposed by an coercive agent and the many other constraints which we normally attribute to luck (but are also, in reality, the result of the choices of other

---

<sup>80</sup> This argument can be taken farther, as Rawls does, to criticize not only the role of the social lottery in our prospects but the “natural lottery” as well. The talents that one is born with are as much a result of luck as the social position into which one is born, even if the latter is determined by the actions of other agents while the former is not. This further complexity, however, is unnecessary for the conclusion that merely formal equality of opportunity is unstable as an ideal and so I leave it aside.

agents) then perhaps we ought to reject both equality of fair opportunity and formal equality of opportunity.<sup>81</sup>

After all, there is a different way of framing the distinction between the vulnerable and non-vulnerable groups. To return to the earlier example of a mountain-side community, rather than asking what caused the disadvantages suffered by a vulnerable group, we might instead wonder whether or not those that live farther away from the edge have unjustly gained the advantages they enjoy. We cannot protect those at the edge without taking resources from those who are farther in, and that appropriation is only justified, some would argue, if they do not have a right to those advantages they possess. A Nozickian libertarian would claim that so long as those advantages are the result of free transfers, then they are the legitimate property of those who possess them.<sup>82</sup> Taking these resources would violate the rights of the owners, so discrimination in the public sphere on the basis of personal preferences, even morally objectionable preferences, must be permissible.<sup>83</sup> On this view, all attempts to level out the playing field, even anti-discrimination laws, would amount to preferential treatment.

The two sides may be right to insist that there is no tenable position between them; I do not have space to argue that point here. Rather, I want to simply flag my suspicion that many people do, in fact, endorse a middle position. They have different reactions to a sign outside a grocery store that says “No blacks, no Jews” than one reading, “No people who

---

<sup>81</sup> For one example see Cavanaugh (2002).

<sup>82</sup> See Nozick (1974), p. 149-182 for the complete argument.

<sup>83</sup> Some proponents of this view argue also that, as a matter of fact, the market will not support racial or gender discrimination and so no state regulation is necessary. This claim is highly contested, but for my purposes, nothing rides on its truth or falsity. If it is the case that no state intervention is necessary to guarantee non-discriminatory practices, then the state will not intervene, but if state regulation is required, then formal equality supports a state obligation to intervene.

cannot afford to pay.” They tend to find the first sign morally offensive and the sort of thing that ought to be regulated against. Perhaps this is a simple failure on their part to appreciate the extent to which the economic and social positions that individuals occupy are largely determined by circumstances beyond their control. Or, alternatively, they may not be adequately attuned to the libertarian argument that, like the benefits that come from the possession of natural talents, individuals are entitled to the benefits that come with having been born a favored race in a racist society.

Though it is possible that a justification exists for extending formal equality of opportunity, the rest of my argument does not depend on one. Rather than trying to convince the advocate of formal equality of opportunity that she ought to modify her position in one of the two directions outlined above, I aim to meet her where she stands. I argue that given a commitment to formal equality of opportunity on which individuals’ legal right to achieve advantaged social positions ought not be limited by the prejudice of others, she ought to endorse not just anti-discrimination legislation, but also preferential treatment for black Americans.

### **3. Blacks as a Vulnerable Population?**

If the preceding discussion of vulnerability and the types of constraints that can appropriately be regulated against (those that violate formal equality of opportunity) is correct, the next step is to demonstrate that blacks are a vulnerable population. This is not yet to argue they ought to be treated preferentially, only that they are vulnerable in some way that offends against formal equality of opportunity. In an obvious sense, many American blacks suffer from conditions that many whites do not: blacks are more likely to be poor,

under-educated and unemployed.<sup>84</sup> Each one of these conditions makes it less likely that an individual will be able to achieve a desired position in society or pursue her own conception of the good, and thus, renders her vulnerable.<sup>85</sup>

However, this unfortunate reality is not enough to establish that blacks as a whole are a vulnerable class, let alone that they are entitled to protection as such. That blacks are more likely to be poor or under-educated, and that those conditions make those who suffer from them susceptible to future harms, is only sufficient to establish that the poor and uneducated are vulnerable because they are poor and uneducated. Why think that blacks who are rich, highly educated and gainfully employed are also part of a vulnerable group because they are black? After all, large numbers of black Americans are perfectly capable of attaining their goals (as evidenced by the fact that they do so). So what evidence do we have that blacks as a whole reside nearer to the precipice than whites because they are black?

Recall that in the previous section, I showed that an individual does not need actually to experience consequences as the result of her condition to count as vulnerable. She can stand at the edge of a precipice but still be located on fully solid ground, just as someone can have a symptomless illness that nevertheless puts her at risk for catching some further disease. The important feature of her situation is that whatever condition she suffers, regardless of whether or not it now affects her, makes her more susceptible to a future harm. It increases the likelihood that she will be harmed in some way. And as a member of a

---

<sup>84</sup> According to a recent study, in 2009, 46% of traditionally college-aged white Americans enrolled in university, compared to only 35% of blacks (Young, 2011, p. 2). A different study revealed that 40% of black children, as opposed to 13% of white children, were living in poverty and the unemployment rate for black men in 2011 was more than double that for white men (Acs, 2013, p. 6 and 18)

<sup>85</sup> There are, of course, exceptions. If one wishes to be an ascetic monk then being born into poverty may be no bar to one's future goals, but in the usual case, these are the sorts of conditions that make it more difficult to fulfill one's aims.

socially salient group she need not actually possess the condition. It is sufficient that the group as a whole exhibits some tendency towards a condition that makes each of the members (whether or not she possesses the condition) vulnerable to some future harm.

This, however, might seem like a non-starter. Blacks do not, in virtue of some shared characteristic, exhibit a group tendency to poverty, under-education, or unemployment. The tendency of the elderly to physical infirmity (which makes them more susceptible to muggings because muggers are more likely to target them) is caused by their age. Blacks are not susceptible to the consequences of poverty or under-education because they are black. As we recognized above, it is not even the case that all blacks are poor or under-educated. It seems that there is no property that all blacks share that makes them vulnerable, as a group, to future harms.

Here, one might be tempted to protest. It seems obvious that blacks do, in fact, share a common feature that makes them vulnerable (as blacks) and it is not a tendency towards poverty or under-education. Rather, as a result of their blackness, all blacks are vulnerable to anti-black racism (to some degree or another). Thus, regardless of whether or not they are poor or under-educated, being black in a society with anti-black racism makes blacks as a whole susceptible to future harm, in the form of racial discrimination.

The problem with pursuing this line of argumentation, in which black vulnerability follows from racial group membership and the existence of anti-black racism, is that those who admit that anti-black racism is a source of black vulnerability could claim that blacks already have protection against future harms sufficient to level the playing field and are not due any preferential treatment. They already enjoy formal equality of opportunity in this regard. They could cite the existence of anti-discrimination laws and practices that impose

legal sanctions on individuals whose public transactions are racially biased. Even evidence that current anti-discrimination laws are ineffective would not be enough to justify preferential treatment, rather than an increase in anti-discrimination laws or litigation. Justifying preferential treatment, as I intend to do, requires more than the assertion that blacks are currently subject to anti-black racism and must be protected. Instead, it must be shown that protecting blacks from discrimination requires treating blacks differently from (and in preference to) whites. The harm they suffer must offend against formal equality of opportunity in a different way, such that the means to securing it is preferential treatment, rather than mere anti-discrimination legislation. The following section gives a theory of black vulnerability that will be used to ground that claim.

### **3.1 A New Theory of Black Vulnerability**

Black vulnerability is importantly linked to the development of negative racial stereotypes. Though the particulars of stereotype development are complex, they appear to rely at least somewhat on the process of sorting individuals into social categories based on what appear to be salient differences (much in the same way we use classification systems to make sense of the natural world). Many of the social categories with which we interact are passed along to us by the culture in which we live. For example, ‘black’ and ‘white’ are significant social categories in the United States and ‘man’ and ‘women’ appear to be important groupings worldwide. Once such categories are established, intragroup differences tend to be experienced as less than they actually are and intergroup differences as greater, giving rise to stereotypes about the members of those groups (Bobo & Massagli, 2001, p. 102).



While some “in-grouping” and “out- grouping” along socially salient lines may be inevitable, the particular stereotypes associated with a social category are still responsive to social conditions. It need not be the case that the category ‘black’ is associated largely with negative stereotypes and ‘white’ with positive ones. Black African civilizations might have coherently held negative stereotypes about whites. Instead, the role or position that a group occupies in society will likely determine the stereotypes associated with that group (Bobo & Massagli, 2001, p. 102). Recent research shows that socio-economic status is a major factor influencing the development and content of stereotypes about the character of particular social groups (Bobo & Massagli, 2001, p. 102). So, existing economic or social inequalities can lead individuals to develop and maintain negative stereotypes about that group. As sociologists Bobo and Massagli (2001) observe, “...all else equal, the more economically successful the members of a group are perceived to be (rich versus poor), the more favorable the other trait beliefs will be (intelligent, self-sufficient, easy to get along with, and speaking English well)” (p.102). In other words, in addition to the already difficult position faced by those who are poor, under-educated or unemployed, the mere fact of their poverty can lead others to develop negative trait beliefs (which are often affectively-laden) towards them.<sup>86</sup>

However, it is not just those who are actually poor who are subject to these stereotypes. In the U.S., ‘black’ is a salient social category that, as noted above, tends to

---

<sup>86</sup> Anderson (2010) recognizes this phenomenon as well. She writes, “Given that practices of social closure make race highly salient as a social category and identity marker, people will try to make sense of the observed effects of segregation by constructing stereotypes about racial groups. The group inequalities generated by segregation provide much of the content of these stereotypes. Thus, people will tend to perceive blacks as (relatively) poor, on welfare, uneducated, idle, prone to form single-parent families at a young age, unlikely to keep up their property, and liable to engage in criminal activity. Such group stereotypes are then used to make inferences about the likely characteristics of individual blacks” (p. 44). On their own these are merely cognitive inferences and not yet racial stigmatization (a term which Anderson prefers to “racism,” which she claims invokes too vicious an attitude and tends to shut down conversation). But, the individuals who hold these stereotypes tend to attribute confirming behavior to dispositions in the stereotyped group, which, in turn, rationalizes negative attitudes towards members of that group. Anderson writes, “The tendency to attribute negative stereotypes dispositionally yields stigmatization of disadvantaged groups” (p. 46).

correlate with lower economic and educational achievement. If one is black, one will tend to occupy a lower socio-economic status than if one is white. Because of this tendency, an association between blackness and poverty develops and negative stereotypes about blacks emerge or already existing stereotypes are confirmed. (This is in keeping with a general correlation between racial group stereotypes and the respective socio-economic position occupied by those groups.) Consequently, the continued existence of large racial disparities between whites and blacks, (most notably economic gaps, but also their attending causes and consequences, including lack of education and employment) contributes to the development and perpetuation of anti-black stereotypes in the United States.<sup>87</sup> In other words, not only because they are black (and some people possess racist attitudes and beliefs) but because blacks tend to be poor, under-educated and unemployed, they are susceptible to the anti-black racism.

But what if (as is clearly not the case in the actual world) black poverty were entirely the fault of those who experience it? If that were true, wouldn't blacks be vulnerable in the same way as those who choose to build their houses close to the edge of the cliff so that they can enjoy the view? They would not deserve protection against the harms of these stereotypes.

My response is twofold. First, even if racial inequalities between blacks and whites were not at all the result of past prejudice, but entirely the fault of those who experience them, we still must admit that not all black people are responsible for black poverty. Some blacks are rich, educated, and employed. Thus, there are at least some blacks who are

---

<sup>87</sup> This assumes that the only or major contributing factor to the creation of particular negative stereotypes is economic achievement. Likely there are other ways in which the social status of certain minority groups influences which stereotypes develop.

completely blameless for current racial inequalities and yet, because of their blackness and the effects of racial disparities, they would be susceptible to the harms caused by increased anti-black racism in their society.

Second, we might be tempted to think that these economically privileged blacks are the only innocent victims of increased anti-black racism and so, harms against the vulnerable which deserve our protection are relatively small. But the rest of the black population is also vulnerable and the nature of their vulnerability is the sort to which all those who endorse formal equality of opportunity would object.

If they are poor, under-educated and unemployed for reasons that are entirely their fault, blacks may deserve many of the consequences that follow from their situation. A law that is applied equally to all but disproportionately disadvantages the poor because they are poor may be perfectly acceptable under formal equality of opportunity, so long as the poor deserve to be poor. We might even condone holding negative attitudes towards the poor, because they have been irresponsible or lazy. These could be character defects that warrant our general disapproval. However, as we have seen, racial inequalities do not breed negative attitudes only towards the poor, but give rise to negative racial stereotypes. Such negative stereotypes about blacks are generally considered central to our ordinary conception of anti-black racism and racism cannot be deserved. We might discriminate against the poor because they are lazy but we ought not to discriminate against them because they are black. To discriminate on the basis of their blackness would be exactly the sort of prejudice that formal equality of opportunity purports to guard against. Thus, a tendency for blacks to be poor, uneducated and unemployed is injurious to all blacks as black because the continued existence of those disparities has a direct causal impact on the development and maintenance

of anti-black racism. It is on the basis of this shared vulnerability to anti-black racism caused by racial disparities (and not anti-black racism, full stop) that blacks ought to be treated as a single, vulnerable group.

#### **4. Protecting the Vulnerable from Harm Through Preferential Treatment**

Having shown that blacks are vulnerable because they are subject to anti-black racism is insufficient to prove that they are entitled to preferential treatment. Even if that anti-black racism stems from pervasive social and economic inequalities, one might still argue that the effects of this racism (namely, violations of blacks' rights to formal equality of opportunity) could be treated through anti-discrimination laws. To show that guaranteeing formal equality of opportunity is sufficient to justify preferential treatment, it must be the case that mere equality of treatment (anti-discrimination laws) would not suffice to protect blacks from the effects of anti-black racism. In this final section, I argue that this is, in fact, the case.

Notice, first, that state policies play a role in maintaining social and economic disparities between blacks and whites. When enacting some resource distribution schema, be it for public education, tax incentives, Medicaid vouchers or unemployment benefits, the state can decide to take race into account or not. And those policies that do not take race into account (those that are 'colorblind') sometimes perpetuate racial disparities. (When they do not, it is likely because they are taking some other factor into account that strongly correlates with race). So, for example, if federal aid for education is done in a race neutral way (say, for simplicity's sake every public school gets \$10,000) then schools that were already underfunded (because they were paid for by the low property taxes in impoverished black neighborhoods) will still have substantially less than those (white) schools which were already well funded. The under-education of black Americans will continue, as will blacks'

poverty and under-employment. By default, colorblind policies that perpetuate these disparities also maintain or further develop anti-black racist attitudes in the form of negative stereotypes.<sup>88</sup> Thus, the state is complicit, if not actively engaged, in the creation of a more racist population.

Conversely, enacting policies that provide preferential treatment for blacks and thereby reduce economic and social disparities is a way for the state to counteract these racist attitudes. Of course, some may object that this sort of attempt at creating a more virtuous citizenry is an illegitimate state action. But, there is an important distinction to be drawn between the cultivation of virtue and state production of vice. A state that attempts to make its citizens more virtuous could call to mind the sort of worries Karl Popper (1945) famously leveled against Plato's republic. To do so effectively might require far-reaching limitations on freedom of thought and expression, and any number of other interferences in the personal lives of its citizens. In Plato's case, it meant placing restrictions on the ideas to which citizens were exposed. But avoiding actively producing vice is a quite different affair. If the state is charged with creating just and fair institutions that protect its citizenry, then it might want to avoid actively cultivating those attitudes which, if acted upon, are likely to cause some portion of its citizens harm. Imagine, for example, that the state is the only entity empowered to print books, for technical rather than legal reasons. In such a situation, refusing to print books that express certain beliefs that contradict principles endorsed by the

---

<sup>88</sup> There is, of course, an argument (often advanced by opponents of affirmative action policies) that preferential treatment for blacks actually breeds resentment and anger amongst whites and contributes to a net increase in anti-black racism. Thus, they would deny my claim that by perpetuating economic disparities the state contributing to the development of anti-black racism. The question of who's right is an empirical one and so I will not attempt to adjudicate it here. However, I do wish to flag my suspicion that piecemeal preferential treatment policies which do not so much reduce racial disparities as create a small, elite class of well-off blacks while leaving a much larger group at the low end of the socio-economic ladder relatively untouched may well increase racial animosity but such evidence does not tell against the theory that a substantial reduction in economic disparities overall might result in a decrease in anti-black racism.

state, such as the value or patriotism or civil service, could be seen an unjustified interference on the part of the state in the name of creating a more virtuous citizenry. But, given the opportunity to submit some number of books for publication each year, the state could choose not to write books that it thought likely to produce a more vicious citizenry.

So, limiting our discussion to the class of cases in which some state intervention is legitimate (a class which will vary depending upon where we fall along the liberal spectrum), in each instance there is an option for the relevant law to dictate equal or preferential treatment to blacks. The state's justification for intervention might stem from concerns completely unrelated to guaranteeing formal equality of opportunity to blacks.<sup>89</sup> For example, with respect to voter ID laws, there is a question about whether or not the state can legitimately pass any legislation requiring that voters show ID, which is not *necessarily* grounded in worries about discrimination (though of course, in the US case the legitimacy of such laws is often undermined by their disproportionate impact on black Americans). Then, there is a further question about whether or not such a law ought to be colorblind or dictate preferential treatment for blacks. If a voter ID law is legitimate, should it apply equally for blacks and whites? Perhaps, if blacks and whites are both subject to voter ID laws, racial disparities in voter turnout rates will continue and white interests will be disproportionately represented leading to continued disparities in wealth between whites and blacks. But, if only whites are IDed, then these disparities will be reduced. If colorblind treatment in this case causes the perpetuation of economic and social disparities between whites and blacks, leading to the maintenance or development of anti-black racism, then maybe the state instead

---

<sup>89</sup> There is a further argument about the circumstances in which the state would be justified in expanding either colorblind or preferential treatment to new areas that we will not explore here.

ought to adopt a policy of preferential treatment for blacks and not require that they produce IDs to vote.

Naturally, some will object that I am running together an increase in actual discrimination with the development of anti-black racist attitudes. The two are not necessarily the same. Why ought the state care about the racism of its population? While the state can properly regulate discriminatory behavior, it does not have the right to police the attitudes of its citizens.<sup>90</sup> The state is charged with guaranteeing that individual liberty is respected, not with restricting the beliefs and attitudes of its citizens. A person may hold racist beliefs and attitudes without engaging in any discriminatory behavior. Thus, there is no justification for policies designed to reduce racial disparities and thus, the continuation and/or cultivation of anti-black racism, only the maintenance or development of effective anti-discrimination laws to prevent private attitudes from spilling over into public actions.

Such an argument fails to properly understand the potential dangers blacks face due to continued or increased anti-black racism. Among concerns for equality and liberty, the liberal state is also committed to stability, a mandate that includes the ability to effectively enforce the laws and policies it creates.<sup>91</sup> There are real worries about the current ability of anti-discrimination laws to operate effectively (as evidenced by well known racial disparities in arrest and conviction rates, sentencing, and hiring, to name a few) and if racial disparities

---

<sup>90</sup> Anderson (2010) declines to refer to these cognitive or affective racial biases as “racism,” preferring to reserve the term for more serious expressions of vice. However, she does argue that racial ‘stigmatization’ can cause a myriad of serious harms to the black population including (1) damaging the reputation of blacks, and thus, their public standing; (2) creating a self-fulfilling prophecy for blacks; (3) encouraging confirmation bias, wherein parties do not accurately assess the evidence but pay more attention to evidence that confirms the stereotype than that which denies it (p. 53-57).

<sup>91</sup> Rawls (1999) is perhaps most explicit about this demand writing, “An important feature of a conception of justice is that it should generate its own support. Its principles should be such that when they are embodied in the basic structure of society men tend to acquire the corresponding sense of justice and develop a desire to act in accordance with its principles” (p. 119). The argument I offer here is an attempt to spell out what features are necessary for a principle of formal equality of opportunity to generate its own support.

continue or widen over time, along with anti-black racist attitudes, those laws might become even less effective. In order to guarantee that anti-discrimination laws are enforceable, some (not insignificant) portion of the population must be committed to ensuring those laws are carried out. And for that to be the case, it must either be in the self-interest of the individuals who make and enforce the laws to ensure that they do not discriminate or those individuals must be committed to the principle of justice anti-discrimination laws uphold. If these conditions are not met, then there is little reason to believe the law will be enforceable.

The recent controversy over the Voting Rights Act of 1965 touches on exactly this issue. During the Supreme Court hearing, critics of the law argued that a decrease in racist attitudes in the affected regions made it so that federal oversight was no longer necessary. This argument speaks to the historical reality that at the time the VRA was passed, despite the implementation of anti-discrimination laws in those counties, elected officials and poll workers could not be trusted to regulate their own voting practices. Racist attitudes were too pervasive. If racist attitudes, though perhaps less explicitly expressed, are just as pervasive today in places with large racial disparities as they were in the past, we ought to be concerned that there are no parties capable of performing the regulatory role necessary to ensure that institutional behavior is non-racist. And even if racist attitudes are less virulent than they once were, we might well wonder whether or not they have sunk below the level necessary to ensure that anti-discrimination legislation is stable. Furthermore, if racial disparities persist without a concerted effort to stem the attitudes to which they give rise, the population is likely to become more racist, less committed to the principle of non-discrimination against



blacks and consequently, less able to regulate institutional policies that call for them to act in a non-discriminatory fashion.<sup>92</sup>

This argument does depend on the assumption that blacks are, in the absence of preferential treatment, likely to remain poorer, under-educated and under-employed. If blacks were moving away from these conditions, then there would be no reason to offer preferential treatment. Racist attitudes would decrease and anti-discrimination laws would stabilize without government intervention. Other socially salient groups, one might argue, have also been at the receiving end of the negative attitudes and stereotypes associated with poverty, unemployment and lack of education. But no preferential policies were necessary because, within a few generations, those groups managed to achieve equal status to whites and those stereotypes largely disappeared. Anti-discrimination laws, then, ought to be sufficient to achieve our ends.

Given historical trends, I find the assumption that racial disparities between blacks and whites will continue quite plausible. Black unemployment has remained twice as high as that of whites for over 50 years (Fletcher, 2013, para. 15). The black poverty rate is still three times as high as that of whites and it has ceased declining (Fletcher, 2013, para. 16). And unlike backward-looking arguments that must show that these disparities are caused or maintained by racial injustice, my theory makes no such demands. Regardless of why these disparities persist, if the state wishes not to contribute to creating a more racist citizenry or wants to ensure it has enforceable anti-discrimination laws, then blacks are entitled to preferential treatment so as to reduce those disparities.

---

<sup>92</sup> A recent AP poll (2012) shows a slight uptick in the percentage of the U.S. population that holds explicitly anti-black racist attitudes (from 48% in 2008 to 51% in 2012).

There is a possible alternative to preferential treatment. Perhaps, rather than ensuring that a state's citizenry is adequately committed to the principles of justice that ground anti-discrimination laws, stability can be maintained entirely through self-interest. James Madison (1787), for example, thought that factions could be managed so long as the country remained large enough and representatives were chosen from a varied enough body of voters to guarantee that the interests of one party could be pitted against the interests of another. Such a scheme would tend to result, Madison claimed, in the election of worthy candidates committed to upholding justice. But, in cases where corrupt candidates gained office, the size of the republic and the many groups contained therein would prevent that corruption from spreading too far.

Unfortunately, Madison's proposed solution is ineffective against a majority who are all committed to circumventing a particular law or policy (or who are not sufficiently committed to ensuring that it is upheld). If the empirical research is correct, then when racial disparities are pervasive throughout the United States and/or there is general awareness of such disparities, negative stereotypes are also likely to be widespread. Those attitudes will tend to manifest in discriminatory behavior by those who hold them because the principles that ground a commitment to formal equality of opportunity (that race is not a morally salient feature and does not affect individual's merit) will not be endorsed by a majority of citizens. Though it will always be in the interest of blacks that race-based anti-discrimination laws be enforced, if a black minority (stymied by discrimination in voting practices) is unable to make its interests heard through the election of anti-discrimination candidates to key positions, then such interests are likely to be overrun by a racist majority. Moreover, discrimination is notoriously difficult to prove, making it more difficult for the blacks to

avail themselves of the general check on corrupt laws and practices offered by the judicial system.<sup>93</sup> An interest in the stability of anti-discrimination laws, then, would seem to demand that the state guard against widespread racist attitudes in the citizenry.

In offering this argument, I do not presume that in all cases a state must or should offer preferential treatment to its vulnerable black citizens. There may well be instances in which colorblind treatment is justified for moral or pragmatic reasons. Rather, the potential harm caused to blacks as a vulnerable population should be included in calculations determining whether or not to extend colorblind or preferential treatment. Moreover, it is not just when considering enacting colorblind policies that the state ought to guard against this harm. The same problem can arise in cases of seemingly legitimate discrimination. As I will argue in the next section, statistical discrimination can also threaten black citizens' right to formal equality of opportunity.

## **5. What's Wrong with Rational Racial Profiling?**

*There is nothing more painful for me at this stage in my life than to walk down the street and hear footsteps and start to think about robbery and then look around and see it's somebody white and feel relieved.*

*-- The Reverend Jesse Jackson*

In the previous sections, I argued that colorblind policies that perpetuate economic and social inequalities between whites and black threaten blacks' formal equality of opportunity. They cultivate racist attitudes in the population and destabilize anti-discrimination laws. Equally troubling are non-colorblind public policies that take race into account when they ought not. Of these, many are wrong for obvious reasons. By

---

<sup>93</sup>See, for example, Semli (2001). Bartlett (2009) argues that excessive legal control and pressure is likely to be less effective at preventing workplace discrimination than promoting commitment to nondiscriminatory norms and doing the latter involves fostering good intentions.

discriminating on the morally arbitrary feature of race, they threaten blacks' equal right to formal equality of opportunity. But within the set of policies that discriminate on the basis of race is a troubling subset that appear to be free of the obvious flaws; these are instances of what is often called 'statistical discrimination' (SD).

In its most general form, SD occurs when some easy-to-see property of individuals correlates (imperfectly) with some other property that, for some reason or other, is harder to detect. Rather than search for the more subtle property, we discriminate between individuals using the more visible one (Maitzen, 1991, p. 23).<sup>94</sup> By itself, SD is a fairly normal and relatively uncontroversial phenomenon. Likely, we all engage in some form of SD on a daily basis. But seemingly more problematic cases of SD arise when an individual is *discriminated against* on the basis of statistical information about a *socially salient group*, like a racial group, of which she is a member. The use of racial group membership as the basis for discrimination I will refer to as 'racial racial profiling' (RRP).<sup>95</sup> Even more worrisome, if only because their effects are likely greater, are state level policies that engage in racial racial profiling.

A host of previous analyses have attempted to articulate what, exactly, is wrong with RRP. Some have argued that it is intrinsically, rather than contingently morally wrong because it violates certain rights or fails to treat its subjects as individuals. I am

---

<sup>94</sup> For example, a middle school basketball coach who wants to create the strongest possible team might walk around the school encouraging very tall students to try out, because being tall is statistically, but not perfectly, correlated with basketball talent (especially, one might think, among young children.)

<sup>95</sup> One potential problem with SD that has not gotten much attention is the worry some socially salient groups may be of a special sort, such that it is inappropriate to assign individuals to them if they do not first assign themselves. If race is a socially salient group of this sort then we would not know, in many cases, whether or not an individual is of a particular race and so it would be wrong (and perhaps racist) to draw such inferences. Of course, holding this view would require subscribing to a sort of racial voluntarism that perhaps just amounts to denying that racial groups are socially salient categories and can ever be used for any purposes. Though this would not just be a problem for RRP but for any policy that uses racial categories, it would explain why RRP is an especially worrisome form of SD.

unconvinced.<sup>96</sup> But, those analyses that have concentrated on articulating the contingent moral wrongness of RRP have also fallen short. In particular, they have failed to capture an important moral distinction between RRP when used against blacks and when used against other socially salient racial groups in the U.S. Although I recognize that RRP can be used against any racial group to bad effect, I argue that its tendency to perpetuate disparities threatens blacks' right to formal equality of opportunity. Thus, there is an additional (albeit

---

<sup>96</sup> In this I follow Lippert-Rasmussen (2007) who considers and rejects three attempts to show that SD is intrinsically morally wrong. They are as follows: (1) we all have a right to be treated as individuals (on the basis of our individual properties) and SD violates that right because it treats us on the basis of properties that pertain to a socially salient group to which we belong. Thus, SD is a failure to respect each person as an individual; (2) SD violates a right to equal treatment. When SD is practiced, individuals who have equal abilities are less likely to receive some benefit or avoid some harm for the irrelevant reason (given their equal abilities) that they do not belong to the socially salient group; (3) SD violates a right to fair treatment. Although I can be treated worse on the basis of what I choose to do, it violates my rights to treat me worse on the basis of what others choose to do. I am not responsible for their behavior and so I should suffer no ill treatment because of it. Lippert-Rasmussen rejects the first by presenting several instances in which an individual is subject to SD without an apparent violation of his rights. For example, a police investigation of a hate crime might concentrate on interviewing suspects who are a different race from the victim because statistically speaking, it is unlikely that a hate crime would be committed by a member of the victim's own race. Those suspects have been subjected to SD and yet their rights do not appear to be violated. Lippert-Rasmussen also contends that most regulations are based on generalizations (regulations involving the age at one can drive, rent a car, or see a certain kind of film) and it would be strange to think that all of these regulations are violating some fundamental right to be treated as individuals. While, as Lippert-Rasmussen acknowledges, it might be morally preferable to treat each other as individuals so that we all receive the treatment we deserve, there is no fundamental right to such treatment. Against the second suggestion, Lippert-Rasmussen argues that unequal treatment is not always morally objectionable because at times unequal treatment serves to promote equal opportunities. Though this may seem to take on board unnecessary baggage about nature of equality (indeed, Lippert-Rasmussen cites Dworkin, who is hardly an uncontroversial subject in the literature on equality and equal opportunity), if the argument I offered above is right then guaranteeing minimal equal treatment (for example, the opportunity to a fair trial regardless of one's race) might, at times, require differential treatment (treating blacks preferentially with respect to the distribution of goods and services). Against the third suggestion, that statistical discrimination is unfair and therefore intrinsically morally wrong, Lippert-Rasmussen argues that there are instances in which we think it is right, and not unfair, for a particular group to accept some small, uncompensated burden in order to secure a much greater benefit for another group. Lippert-Rasmussen takes this point as intuitive, but I'm not sure that it is. However, on all but the most extreme forms of libertarianism, there is some caveat allowing for small, uncompensated harms to one group in order to avoid catastrophic harms to another. And, as I argued above, any conception of justice that takes equality of opportunity as one of its ideals must admit that there are times when securing these rights for all necessitates imposing some burdens on a few (making them less well off than they would have been were the rights of others not fully respected) and this is not thought to be unfair.

contingent) reason not to practice RRP against blacks given current social and economic conditions.<sup>97</sup>

## 5.1 The Contingent Moral Wrongness of Statistical Discrimination

Those who favor democratic equality have suggested that RRP is wrong because it stigmatizes an already marginalized or disadvantaged group. It either expresses that we do not, or makes it the case that we cannot, relate as equals to one another.<sup>98</sup> For example, when the police single out black Americans as the likely perpetrators of certain crimes it creates or emphasizes an already existing association between blackness and criminality.

---

<sup>97</sup> Lippert-Rasmussen (2007) offers the insightful suggestion that the wrong of certain forms of SD may be a matter of recognizing how the facts upon which the instance of SD relies came to be true. He presents two ways of understanding this worry. The first way is where SD becomes a self-fulfilling prophecy. For example, suppose police officers believe that blacks are more likely than other groups to be hostile towards and use force against police officers. Thus, police officers approach blacks more aggressively and are more likely to use forceful methods of containment against blacks than against other racial groups. Observing and experiencing the hostility of these officers in their interactions with blacks angers and scares blacks and so they begin to respond to police officers with more hostility. It would be inappropriate, Lippert-Rasmussen argues, for the police to appeal to this statistical fact as the reason for their discriminatory policy because it is the result of unjustified differential treatment. The second problem with using some social facts as a basis for SD is that they may obtain as the result of some other, unjust action. For example, suppose that blacks are more likely to sell illegal drugs than other groups because living in a racist society has severely limited their employment opportunities. Though targeting blacks on the basis of these statistical facts will greatly increase the chances that the police will catch drug dealers, it would be inappropriate for them to do so because some of those who benefit from this statistical discrimination (whites) will have also played an integral role in unjustly bringing about the fact that blacks are more likely to deal drugs. Though I think Lippert-Rasmussen is right to argue that these are worries for statistical discrimination and I am generally sympathetic to his analysis, in trying to justify the harm of RRP to negative liberty theorists his argument runs up against a number of potentially fatal problems. Perhaps most devastating is the following: so long as the people employing (or benefiting from) the statistical data did not themselves perpetrate the original injustice, why would it be inappropriate for them to rely on that information? One can agree that blacks ought to be compensated for past injustices but still maintain, while it might be supererogatory to refrain from acting on those social facts, given their origin it is not required. That some of the people who perpetrated the initial injustice were also police or also white does not seem sufficient reason for others who are police or white to abstain from using SD. And, even if such a solution is workable in the case of police in negative feedback loops (because they are all members of a discreet institution and contributed directly to the particular social fact upon which they are now acting) without a complex, and I think implausible, theory of collective responsibility whites could not be thought to be responsible for the crimes of past whites and the link between particular social facts and past injustices would need extensive spelling out (though I find such links intuitively quite plausible).

<sup>98</sup> For a comprehensive articulation of the ideal of democratic equality and its relation to worries of racial stigmatization see Anderson (2010), especially footnote 11.

Consequently, others might be more likely to see blacks as criminals even when the statistics do not support such a perception. The harms this might cause are too numerous to mention, but some recent victims of such stigmatization likely include Trayvon Martin, Jonathan Ferrell and Jordan Davis. Nevertheless, there may be instances in which the benefit of singling out a particular racial group for RRP would outweigh the harm caused by stigmatization and its contribution to inequality.<sup>99</sup>

The objection I offer to RRP is a species of this more general worry about the dangers of stigmatization. However, unlike other critiques of this kind, it relies on a fairly minimal view of what a commitment to formal equality of opportunity entails—It must be the case that laws that declare the equal legal status of racial groups are enforceable.

RRP against blacks necessarily entails singling blacks out (because they are black) for some worse treatment than other racial groups. This may well lead to some overall societal benefit but it causes some harm to the individual black person (even though that harm may be perfectly justified, in a broader sense). For example, imagine (contrary to fact) that blacks are more likely than whites to possess and distribute small quantities of marijuana. Targeting blacks for this crime would be case of RRP, but it would also burden those blacks who are arrested with legal fees, court dates, fines, jail time and criminal records. This, in addition to contributing directly to the perpetuation of negative stereotypes or stigma against black Americans (for example, the perception that blacks are naturally more criminal), would also likely perpetuate social and economic inequalities between blacks and whites. For example, regardless of whether or not they are convicted, blacks who have to appear in court might miss work, sacrificing wages or even their job, and pay expensive legal fees, diminishing

---

<sup>99</sup> Though I disagree, there were certainly some who argued this was the case for Middle Eastern individuals boarding planes in the days after the September 11, 2001 attacks.

savings or increasing debt. Perpetuating these inequalities will, in turn, contribute to creating or maintaining a more racist population. And, as with colorblind policies, the dangers of these persistent racist attitudes include ongoing or increased discrimination and the destabilization of anti-discrimination laws.<sup>100</sup>

Naturally, as in the case of colorblind policies, the effect that any particular state policy might have on the racist attitudes of the population would need to be weighed against other liberties. If formal equality of opportunity is indeed a demand of liberal justice then it cannot be sacrificed wholesale for some other social good. But, for pragmatic purposes, there may be times when it must be traded to some limited extent in order to secure some other, equally or more important social good. Thus, in times of national emergency, for example, some increase in racist attitudes might be acceptable (provided it does not sink below a point where the laws become unstable) in order to protect the population from a greater threat. For example, the police might have engaged in RRP by pulling over only black men, if they had known that the DC sniper was black. In that case, the potential benefit of stopping the murders could have outweighed the harm of perpetuating racial disparities (if, for example, pulling over more black men led to a greater number of arrests of black men during that time period). Importantly, not every racial group faces this threat when it is profiled, and so the danger of profiling blacks may be greater than the danger of profiling other racial groups. Racial disparities between blacks and whites are greater and more

---

<sup>100</sup> One might object that the worry I have identified for RRP and colorblindness is not just a problem for these policies but could be leveled against any policy likely to perpetuate economic and social disparities between whites and blacks. I agree. If I have not exhausted the logical space of potential public policies by describing those that employ colorblindness, preferential treatment, racism and RRP and there remains some other set of policies that could perpetuate or exacerbate economic and social disparities between blacks and whites then it too would be subject to these worries. In singling out colorblindness and RRP my goal is to highlight two types of policies that are sometimes thought to pose no threat of racism (or the racist status of which is highly contested) and argue that, in fact, they do. But I remain open to the suggestion that there might be other policies that are problematic in the same way.



persistent than between whites and other racial groups and so, RRP should be exercised with particular caution against blacks. Indeed, it may be, and likely is the case, that anti-discrimination laws are already unstable for the black population (and so preferential treatment is needed to correct this injustice).

This account, like others that emphasize the potential problems RRP can cause for equality of opportunity also explains why RRP may be wrong when carried out by the state but need not be wrong (and is certainly less wrong) when performed by individuals. In other words, RRP is an institutional failure and perhaps also, in some circumstances, a form of institutional racism. When racist attitudes increase to the point that anti-discrimination laws are no longer stable, the state has failed in an important duty to its black citizens. The state expresses disregard for the rights of its black citizens whether it contributes to these attitudes by maintaining racial disparities or simply sits by idly, watching as they develop. The same is not true for the individual nor the private corporation. Though they contribute to the stigmatization of blacks when they engage in RRP, they are not tasked with maintaining a sufficient level of non-racist attitudes of the citizenry. (Although they may, like Jackson, take it on a personal moral duty and so feel guilt or pain when they contribute to stigmatization, if only by expressing it in their own minds.)<sup>101</sup>

---

<sup>101</sup> I lump private citizens and corporations together here, but I am not sure that is appropriate. It may be that corporations have a duty not only to obey the laws of the country in which they operate, but also not to perform actions that indirectly undermine those laws. Though I do not have space to pursue this complication here, I want to flag it as a potential disanalogy between, say, the confession by Rev. Jackson that began this section and Cornel West's famous anecdote that he was unable to catch a cab in NYC, despite the fact that he was a fairly wealthy, well-dressed, Harvard professor.

## 6. Conclusion

Many will find this hard to swallow that assent to mere formal equality of opportunity commits one to preferential treatment for black Americans, And those who want to hold onto formal equality of opportunity might refuse to move from one to the other, just as some could, as I admitted earlier, deny the force of Rawls' argument from formal to fair opportunity. But, it is important to see that while extending preferential treatment to black Americans in the way that I have suggested here would certainly entail a somewhat different distribution procedure for resources than is currently employed, it is not nearly as radical a redistribution as Rawls'. Assenting to equality of fair opportunity requires those who have done well in order to provide for their children's future realize that their children are no more deserving of the money they have saved for them than any other children. My argument requires that individuals agree to the more modest claim that, because they occupy a privileged position, they are not entitled to the same share of some government goods and services as those who do not. They might still be entitled to any private advantages they have accrued. Showing that they are not would require additional argumentation.

While the intention of this article, which was to advance an argument that appeals to a wide spectrum of liberal thinkers, focused on the contribution that colorblind policies and RRP make to racist attitudes and the destabilization of anti-discrimination laws, I want to end by acknowledging that there are other, serious concerns one might have with a state that creates policies which it knows foster racist attitudes against its black citizens. Even if such a state were able to effectively protect blacks against any increases in discrimination and enforce anti-discrimination legislation, we might still worry that such policies are extremely disrespectful. We can imagine a state in which all of the non-black citizens harbor anti-black

racist attitudes but only exhibit those attitudes in settings in which no blacks are present. They hold private rallies, perhaps, where they chant anti-black racist slogans and reenact lynchings. The proceedings are secret and blacks incur no harm, psychological or otherwise, from their occurrence. Perhaps some would argue, in this hypothetical scenario, that blacks experience no actual disrespect. But, I think we might still have worries about the character of a state that not only tolerates these activities but, supplies the slogans and the rope. When the state takes actions that create and foster anti-black racist attitudes, it seems to me that it disrespects black people even if it does them no harm.

## REFERENCES

- Acs, G. (2013). *The Moynihan report revisited*. Washington, DC: The Urban Institute.
- Adams, M., Bell, L. A., Griffin, P. (1997). *Teaching for Diversity and Social Justice: a Sourcebook*. New York: Routledge.
- Anderson, E. (1999). What is the point of equality? *Ethics*, 109, 2, 287–337.
- Anderson, E. (2002). Integration, affirmative action, and strict scrutiny. *NYU Law Review*, 77, 1195–1271.
- Anderson, E. (2010). *The imperative of integration*. Princeton, NJ: Princeton University Press.
- Arneson, R. (1989). Equality and equal opportunity for welfare. *Philosophical Studies*, 56, 77-93.
- Banaji, M. R., Lemm, K. M., & Carpenter, S. J. (2001). Automatic and implicit processes in social cognition. In A. Tesser & N. Schwartz (Eds.), *Blackwell Handbook of Social Psychology: Intraindividual Processes* (pp. 134-158). Oxford: Blackwell.
- Bartlett, K. T. (2009). Making good on good intentions: The critical role of motivation in reducing implicit workplace discrimination. *Virginia Law Review*, 95, 1893-1972.
- Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal: A Field Experiment on Labor Market and Discrimination. *The American Economic Review*, 94, 4, 991-1013.
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *Journal of personality and social psychology*, 81, 5, 828-841.
- Blum, L. A. (2002). *I'm Not a Racist, But—: The Moral Quandary of Race*. Ithaca: Cornell UP.
- Bobo, L. (2001). Racial Attitudes and Relations at the Close of the Twentieth Century. N. Smelser, W. J. Wilson, & F. Mitchell (Eds.), *America Becoming: Racial Trends and Their Consequences* (pp. 262–299). Washington, DC: National Academy Press.
- Bobo, L., & Massagli, M. (2001). Stereotyping and urban inequality. *Urban inequality: Evidence from four cities*, 89-162.
- Boxill, B. (1992). *Blacks and social justice, revised edition*. Lanham, MD: Rowman &

- Littlefield.
- Brownstein, M. and Saul, J. (Eds). Forthcoming. *Implicit Bias and Philosophy: Volume 1, Metaphysics and Epistemology*. Oxford University Press.
- Brownstein, M. and Saul, J. (Eds). Forthcoming. *Implicit Bias and Philosophy: Volume 2, Moral Responsibility, Structural Injustice, and Ethics*. Oxford University Press.
- Carmichael, S. (Kwame Ture) & Hamilton, C. (1967). *Black Power*. New York: Random House.
- Carter, I., (1999). *A measure of freedom*. Oxford: Oxford University Press.
- Cavanagh, M. (2002). *Against equality of opportunity*, Oxford: Clarendon.
- Coates, T.-N. (2014, May 21). The case for reparations. *The Atlantic*. Retrieved from <http://www.theatlantic.com/features/archive/2014/05/the-case-for-reparations/361631/>.
- Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics*, 99, 906-944.
- Cooper, D. E. (1968). Collective Responsibility. *Philosophy*, 43, 258-268.
- Covington, J. (2010). *Crime and racial constructions : cultural misinformation about African Americans in media and academia*. Lanham, Md.: Lexington Books.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, 15, 12, 806-813.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of personality and social psychology*, 81, 5, 800-814.
- Dijksterhuis, A., & Van Knippenberg, A. D. (1996). The knife that cuts both ways: Facilitated and inhibited access to traits as a result of stereotype activation. *Journal of Experimental Social Psychology*, 32, 3, 271-288.
- Dixon, T. & Linz, D. (2002). Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News. *Journal of Communication*, 50, 2, 131–54;
- Dorfman, L. & Schiraldi, V. (2001). *Off Balance: Youth, Race & Crime in the News*. Washington, D.C.: Building Blocks for Youth. Retrieved from <http://www.buildingblocksforyouth.org/media/media.pdf>.

- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology: Special Issue on Unconscious Processes in Stereotyping and Prejudice*, 33, 510-540.
- Dovidio, J. F. & Gaertner, S. L. (2000) Aversive Racism and Selection Decisions: 1989 and 1999. *Psychological Science* 11, 315-319.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of personality and social psychology*, 82, 1, 62-68.
- Dovidio, J. F. & Gaertner, S. L. (2004). Aversive Racism. *Advances in experimental social psychology*, 36, 1-52.
- Dworkin, R. (1981). What is equality? Part 2: Equality of resources. *Philosophy and Public Affairs*, 10, 283-345.
- (1973). *Education and Racism: An Action Manual*. Washington, D.C.: National Education Association. Retrieved from <http://files.eric.ed.gov/fulltext/ED073214.pdf>.
- Emerson, M.O. (2011). Review of 'The Imperative of Integration' by Elizabeth Anderson. *American Journal of Sociology*, 117, 1, 317-319.
- Faucher, L., & Machery, E. (2009) Racism: Against Jorge Garcia's moral and psychological monism. *Philosophy of the Social Sciences*, 39, 1, 41-62.
- Fletcher, M. A. (2013, August 27). Fifty years after march on Washington, economic gap between blacks, whites persists. *The Washington Post*. Retrieved from [http://www.washingtonpost.com/business/economy/50-years-after-the-march-the-economic-racial-gap-persists/2013/08/27/9081f012-0e66-11e3-8cdd-bcdc09410972\\_story.html](http://www.washingtonpost.com/business/economy/50-years-after-the-march-the-economic-racial-gap-persists/2013/08/27/9081f012-0e66-11e3-8cdd-bcdc09410972_story.html).
- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Frye, M. (1983). *The Politics of Reality: Essays in Feminist Theory*. Freedom, California; The Crossing Press.
- Garcia, J. L. A. (1996). The heart of racism. *Journal of Social Philosophy*, 27, 1, 5-46.
- Garcia, J. L. A. (2011). Racism, psychology and morality: Dialogue with Faucher and Machery. *Philosophy of the Social Sciences* 41, 2, 250-268.
- Gawronski, B., Hofmann, W. & Wilbur, C. J. (2006). Are "implicit" attitudes

- unconscious? *Consciousness and Cognition* 15, 3, 485-499.
- Gehring, W. J., Karpinski, A., & Hilton, J. L. (2003). Thinking about interracial interactions. *Nature Neuroscience*, 6, 12, 1241-1243.
- Gilliam, F. & Iyengar, S. (2000). Prime Suspects: The Influence of Local Television News on the Viewing Public. *American Journal of Political Science*, 44, 3, 560–573.
- Glasgow, J. (2009). Racism as Disrespect. *Ethics*, 120, 64-93.
- Goldman, A. (1976). Affirmative action. *Philosophy and Public Affairs*, 5, 178–195.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M.R. (2007). The presence of implicit bias in physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22, 1231–1238.
- Greenwald, A., McGhee, D. & Schwartz, J. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 6, 1464–80.
- Greenwald, A., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97, 17–41.
- Haslanger, S. (2000). Gender and race:(what) are they? (What) do we want them to be?. *Nous*, 34,1, 31-55.
- Haslanger, S. (2004). Oppressions: Racial and other. M. P. Levine & T. Pataki (Eds.), *Racism in Mind*. Ithaca: Cornell UP.
- Haslanger, S. (2012). Ideology is a moral issue. [Lecture handout]. *Doing Justice to the Social*. The Carus Lectures. The American Philosophical Association.
- Haslanger, S. (Forthcoming). Studying White Black: Trust, Opportunity and Disrespect. *DuBois Review*.
- Held, V. (1970). Can a random collection of individuals be morally responsible? *The Journal of Philosophy* 67, 14, 471-481.
- Hill, Jr., T. E. (1995). The message of affirmative action. *The Affirmative Action Debate*. New York: Routledge.
- Hoyt, C., Jr. (2012). The pedagogy of the meaning of racism: reconciling a discordant discourse. *Soc Work*, 57, 3, 225-234.

- Hyers, L. L., & Swim, J. K. (1998). A comparison of the experiences of dominant and minority group members during an intergroup encounter. *Group Processes & Intergroup Relations*, 1, 2, 143-163.
- Jones, C. P., Jones, C. Y., Perry, G. S., Barclay, G., & Jones C. A. (2009). Addressing the social determinants of children's health: A cliff analogy. *Journal of Health Care for the Poor and Underserved*, 20, 1-12.
- Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002) Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental Social Psychology*, 38, 586-602
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of personality and social psychology*, 81, 5, 774-788.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *Journal of personality and social psychology*, 78, 5, 871-888.
- Kelly, D., & Roedder, E. (2008) Racial Cognition and The Ethics of Implicit Bias. *Philosophy Compass*, 3, 522-540.
- Kim, Y. M. (2011). *Minorities in higher education: Twenty-fourth status report*. Washington, DC: American Council on Education.
- Kottow, M. (2003). The vulnerable and the susceptible. *Bioethics*, 17, 460-471.
- Kovel, J. (1984). *White racism : a psychohistory* (Columbia University Press Morningside ed.). New York: Columbia University Press.
- Kramer, M. H. (2003). *The quality of freedom*. Oxford: Oxford University Press.
- Lippert-Rasmussen, K. (2007). Nothing personal: On statistical discrimination\*. *Journal of Political Philosophy*, 15, 4, 385-403.
- List, C. & Pettit, P. (2011). *Group Agency: The Possibility, Design and Status of Corporate Agents*. Oxford Scholarship Online.
- Madison, J. (1787, November 22). The Federalist, no. 10. *Daily Advertiser*. Retrieved from <http://www.constitution.org/fed/federa10.htm>.
- Maitzen, Stephen (1991). The ethics of statistical discrimination. *Social Theory and Practice*, 17, 1, 23-45.



- McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37, 435-442.
- Mills, C. (2003). 'Heart attack:' A critique of Jorge Garcia's volitional conception of racism. *The Journal of Ethics*, 7, 29-62.
- Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109, 41, 16474-16479.
- Nagel, T. (1973). Equal treatment and compensatory discrimination. *Philosophy and Public Affairs*, 2, 348-363.
- Nagel, T. (1991). *Equality and partiality*. Oxford: Oxford University Press.
- Nextions. (2014) Written in Black & White: Exploring Confirmation Bias in Racialized Perceptions of Writing Skills. *Yellow Paper Series*. Retrieved from [http://www.nextions.com/wp-content/files\\_mf/13972237592014040114WritteninBlackandWhiteYPS.pdf](http://www.nextions.com/wp-content/files_mf/13972237592014040114WritteninBlackandWhiteYPS.pdf)
- Nier, J. A. (2005). How Disassociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach. *Group Processes Intergroup Relations*, 8, 1, 39-52.
- Nozick, R. (1974). *Anarchy, state, and utopia*, New York: Basic Books.
- Payne, B. K. (2006). Weapon Bias: Split-Second Decisions and Unintended Stereotyping. *Current Directions in Psychological Science*, 15, 287-291.
- Pettit, P. (1997). *Republicanism: A Theory of Freedom and Government*, Oxford: Clarendon Press.
- Phelps, E. A. O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore J. C. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, 12, 729-738.
- Philips, M. (1984). Racist acts and racist humor. *Canadian Journal of Philosophy*, 14, 1, 75-96.
- Pierce, A. (in press). Structural Racism, Institutional Agency and Disrespect. *Journal of Philosophical Research*.
- Pinderhughes, E. (1989). *Understanding Race, Ethnicity and Power*. New York: The Free Press.
- Popper, K. (1945). *The open society and its enemies, Volume 1*. London: Routledge.

- Price, J., & Wolfers, J. (2010). Racial discrimination among NBA referees. *The Quarterly Journal of Economics*, 125, 4, 1859-1887.
- Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press, rev. ed. 1999.
- Rennison, C. (2001). Violent Victimization and Race, 1993–98. *Bureau of Justice Statistics Special Report NCJ 176354*. Washington, D.C.: U.S. Department of Justice. Retrieved from <http://permanent.access.gpo.gov/lps19659/www.ojp.usdoj.gov/bjs/pub/pdf/vvr98.pdf>.
- Roemer, J. E. (1998). *Equality of opportunity*. Cambridge: Harvard University Press.
- Rome, D. (2004). *Black demons : the media's depiction of the African American male criminal stereotype*. Westport, CT: Praeger.
- Ross, S., & Agiesta, J. (2012, October 27). AP poll: Majority harbor prejudice against blacks. *AP*. Retrieved from <http://bigstory.ap.org/article/ap-poll-majority-harbor-prejudice-against-blacks>.
- Rudman, L. A., & Borgida, E. (1995). The afterglow of construct accessibility: The behavioral consequences of priming men to view women as sexual objects. *Journal of Experimental Social Psychology*, 31, 6, 493-517.
- Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, 17, 4, 437-465.
- Saad, L. (July 18, 2013). In U.S., 52% of Blacks Unhappy With Societal Treatment. *Gallup*. Princeton, NJ: Gallup, Inc. Retrieved from <http://www.gallup.com/poll/163553/blacks-unhappy-societal-treatment.aspx>
- Semli, M. (2001). Why are employment discrimination cases so hard to win? *Louisiana Law Review*, 61, 555-575.
- Shelby, T. (2002). Is Racism in the "Heart"? *Journal of Social Philosophy*, 33, 3, 411-20.
- Shelton, J. N. (2000). A reconceptualization of how we study issues of racial prejudice. *Personality and Social Psychology Review*, 4, 4, 374-390.
- Sommers, S. R., Apfelbaum, E. P., Dukes, K. N., Toosi, N., & Wang, E. J. (2006). Race and media coverage of Hurricane Katrina: Analysis, implications, and future research questions. *Analyses of Social Issues and Public Policy*, 6, 1, 39-55.

- Tatum, B. D. (1997). *'Why Are All the Black Kids Sitting Together in the Cafeteria?' and other conversations about race*. New York, NY: Basic Books.
- Tuomela, R. (1989). Actions by collectives. *Philosophical Perspectives*, 3, 471-496.
- Tuomela, R. (2006). Joint intention, we-mode and I-mode. *Midwest Studies in Philosophy* 30, 1, 35–58
- Weber, M. (1968). *Economy and Society: An Outline of Interpretative Sociology*. (G. Roth & C. Wittich, Eds.). Berkeley: University of California Press.
- Wellman, David T. (1977). *Portraits of White Racism*. New York, NY: Cambridge University Press.
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of experimental social psychology*, 10, 2, 109-120.
- Zheng, R., & Brownstein, M. (forthcoming). Attributability, Accountability, and Implicit Bias. J. Saul (Ed.), *Implicit Bias and Philosophy: Volume 2, Moral Responsibility, Structural Injustice, and Ethics*. Oxford University Press.