

César Cabezas

Temple University

This is a post-peer-review, pre-copyedit version of an article that is published in *Ethics*. Please cite the published version, which is available at <https://www.journals.uchicago.edu/doi/10.1086/727273>.

## **Is Conceptual Inflation a Problem for a Theory of Institutional Racism?\***

### Abstract

I address the objection that the concept of racism has become overly inflated. The charge of conceptual inflation is often leveled against conceptions of racism that go beyond the traditional understanding of racism as race-based ill-will or disregard. Theories of institutional racism are a common target of conceptual inflation critics, especially when they ascribe racism to institutions partly in virtue of their impact. Conceptual inflation critics argue that theories of institutional racism engage in untoward conceptual inflation insofar as they undermine our moral understanding of racial phenomena, hinder our ability to explain the causes of racial inequality, and even undercut struggles for racial justice. I develop an original account of institutional racism that is immune to all three versions of the conceptual inflation challenge.

### Introduction

There is a growing complaint that the concept of racism has become overly inflated. Conceptual inflation critics worry about the negative consequences of employing the term too expansively and without the proper analytical rigor. “Racism” was originally coined to designate the doctrine of racial classification and hierarchy advanced by the Third Reich. Over time, the term came to refer to racially hostile and discriminatory behavior that is willful, conscious, or endorsed by individual agents. This relatively narrow definition tends to be the reference point for those who worry about the conceptual inflation of “racism”. Today, “racism” is

increasingly applied to a much larger variety of contexts. Jokes, sports team mascots, monuments, institutions, and even whole societies may now be described as racist. Not only has the class of potentially racist phenomena expanded, but the requirements for ascribing racism have also been relaxed. Today, for an action to count as racist, it need not be driven by a racist motive, nor need the person carrying out the action consciously endorse any racist beliefs. In fact, research on implicit bias suggests that a person may explicitly reject racist beliefs and still act in a racist manner.<sup>1</sup> Moreover, actions do not even need to express overt racial discrimination to count as racist. Research on microaggressions suggests that unintentional slights, snubs, and insults, as well as verbal and non-verbal messages that invalidate a racialized group's identity or lived experience may also be racist.<sup>2</sup> These discursive and theoretical developments point to a wide-scope conception of racism that is common among anti-racist activists and has increasingly taken hold of lay discourse.<sup>3</sup>

Philosophers have been less quick to embrace the shift towards a wide-scope conception of racism. Some worry that conceptual inflation has turned “racism” into a mere catch-all phrase for everything that one condemns or disagrees with in the domain of race. They call for a narrow-scope conception of racism that reins in the indiscriminate inflation of the term.<sup>4</sup> Others embrace the conceptual inflation of “racism” and defend a wide-scope conception of racism.<sup>5</sup> This paper contributes to this ongoing philosophical debate in three ways. First, I identify three different versions of the conceptual inflation challenge. Philosophers have raised two worries about the conceptual inflation of “racism”. First, it can undermine our moral understanding of racism. Secondly, it can hinder progress towards racial justice by undermining the prospect of interracial dialogue and reconciliation. However, there is a third version of the conceptual inflation challenge that philosophers have ignored so far. According to sociologist Robert Miles—whom philosophers critical of conceptual inflation often cite—the problem with the conceptual inflation of “racism” is that it weakens the causal-explanatory power of the concept.<sup>6</sup> For Miles, the goal of a theory of racism is to provide an accurate description of the phenomena that fall under it, and to explain how such phenomena contribute to relevant

social processes—in particular, to the persistence of racial inequality. In Miles’ view, a narrow-scope conception of racism is better suited for this causal-explanatory role, and therefore conceptual inflation should be avoided. Thus, one can criticize the conceptual inflation of “racism” on at least three grounds: moral-explanatory, pragmatic, and causal-explanatory.

The philosophical debate on the conceptual inflation of “racism” has focused on the moral-explanatory and pragmatic questions. Proponents of the wide-scope conception of racism have defended it against the claim that it undermines our moral understanding of racism and hinders the pragmatic goal of attaining racial justice. However, a defense of the wide-scope conception of racism on causal-explanatory grounds is still lacking. This paper’s second contribution to the debate is to fill that gap. I address Miles’ causal-explanatory critique of the conceptual inflation of “racism”. To do so, I focus on his causal-explanatory critique of a key element of the wide-scope conception of racism—the concept of institutional racism. Pace Miles, I argue that the concept of institutional racism, properly understood, is well-suited to accurately describe racist phenomena at the level of institutions, and to explain how such phenomena contribute to the persistence of racial oppression. Miles is not the only conceptual inflation critic to target the concept of institutional racism. Garcia and Blum also criticize this concept on moral-explanatory and pragmatic grounds. I also defend the concept of institutional racism from these two other versions of the conceptual inflation challenge. Hence, this paper’s third contribution to the debate is to develop an original theory of institutional racism that can withstand all three versions of the conceptual inflation challenge.

The paper proceeds as follows. In section 1, I examine a model for theorizing racism that is influential among proponents of the wide-scope conception of racism. This model, which I call output-based, centers the analysis of systems of race-based oppression and their maintenance via the interaction of individual, ideological, and institutional forms of racism. I briefly show how the output-based model theorizes individual and ideological racism. In section 2, I focus on institutional racism and argue for a wide-scope theory that recognizes both input-based and output-based cases of institutional racism. Input-based cases of institutional

racism correspond to institutions whose goals, rules, or application of rules are infected by race-based ill-will or disregard. Output-based cases correspond to institutions that contribute to a system of racial oppression and are legitimized by racist ideology. In section 3, I tackle the conceptual inflation challenge to theories of institutional racism. I show that the real target of conceptual inflation critics is not the concept of institutional racism per se, but the output-based model of institutional racism in particular. Conceptual inflation critics favor a narrow-scope theory of institutional racism that only recognizes cases of input-based institutional racism. I lay out and counter all three versions of the conceptual inflation critique of the output-based model of institutional racism: moral-explanatory, pragmatic, and causal-explanatory. Furthermore, I argue that recognizing this kind of institutional racism is key for furthering our understanding of the mechanisms that sustain racial oppression.

### 1. The Output-Based Model of Racism

According to Hardimon, the narrow-scope conception defines racism as racially hostile and discriminatory behavior that is willful, conscious, or endorsed by individual agents.<sup>7</sup> While this articulation of the narrow-scope conception focuses on racist actions, one may also extend it to racist individuals and racist institutions. An individual who willfully or consciously engages in racially hostile or discriminatory behavior (or who affirms that behavior upon reflection) is racist. An institution is racist if its goals and rules are racially hostile or discriminatory. Alternatively, if the institution's goals and rules are not racist, the institution may still be racist if its members engage in racist behavior while carrying out their institutional roles. What all these versions of the narrow-scope conception have in common is that they ascribe racism to actions, individuals, and institutions in virtue of their input. An action is racist if it stems from the racial ill-will or disregard of an agent. An agent guided by such motivations is racist, and an institution whose agents are guided by such motivations (or whose goals and rules are designed or enacted by agents guided by such motivations) is also racist. As Garcia puts it, racism is defined by the vices of race-based ill-will and disregard, which can infect

actions, individuals, and institutions so as to render them racist.<sup>8</sup> In the narrow-scope conception, actions, individuals, and institutions are racist in virtue of their input (i.e. race-based ill-will or disregard).

The wide-scope conception is a reaction to theorizing racism purely in terms of input. This conception starts from the intuition that we can ascribe racism to actions, individuals, and institutions partly in virtue of their output—more specifically, in virtue of their contribution to the maintenance of a system of racial oppression. As its name suggests, the wide-scope conception seeks to expand the meaning of “racism”. It does not deny the importance of race-based ill-will and disregard in ascriptions of racism. However, it aims to complement this input-based model of racism with an output-based model that centers the analysis of systems of racial oppression. By contrast, defenders of the narrow-scope conception of racism only endorse the input-based model and reject the output-based model on grounds of conceptual inflation. Since critics of conceptual inflation take the input-based model for granted, this section focuses on articulating and defending the output-based model. However, following the wide-scope conception, I contend that both models are important for a thorough understanding of racism.

In this section, I offer a philosophical reconstruction of the output-based model of racism.<sup>9</sup> The key feature of the output-based model of racism is that all racist phenomena are interrelated parts of a larger system of racial oppression. A system of racial oppression is a dynamically stable social system partially structured by hierarchical racial categories and whose output is the distribution of (economic, political, social, symbolic, aesthetic, among others) advantages and disadvantages along racial lines.<sup>10</sup> Global white supremacy (as well as its many national and sub-national variants) is arguably the most important empirical example of a system of race-based oppression and will be the focus of my analysis. Systems of race-based oppression change over time to adapt to shifting political, cultural, and economic circumstances. For example, the social system of white supremacy in the post-Civil War South reflected the economic and political needs of the old Southern planter elite.<sup>11</sup> This racist system was legitimized by the doctrine of biological racism. By the second half of the 20<sup>th</sup> century, the growing economic and political power of the Black community led to the collapse of Jim

Crow institutions, as well as the turn away from biologically racist ideas in mainstream culture. The U.S. racist system evolved to adapt to these new circumstances. It started to rely more on color-blind and informal (rather than race-conscious and state-sponsored) practices to continue its racially unequal distribution of advantages and disadvantages. It also started to justify racial inequality in terms of the cultural—rather than biological—inferiority of non-whites. Because of its dynamism in the face of challenges to it, the U.S. racist system has been able to endure over time and keep its output—the unequal distribution of resources to the benefit of whites and to the detriment of non-whites—relatively stable.<sup>12</sup> This is the sense in which racist systems are dynamically stable.

Under systems of racial oppression, most agents participate (directly or indirectly) in social practices that adversely impact members of disadvantaged racialized groups. The prevalence of these harmful practices is partially explained by the presence of racist ideologies that inform agents' beliefs (explicit and implicit) and actions (fully intentional and habitual). Racist ideologies consist of beliefs, habits, social meanings, symbols, scripts, stereotypes, and interpretive frames that function to legitimize systems of racial oppression. In addition, agents who reliably benefit from the racially disparate impact of these practices acquire an interest in their continuation. This interest stems from their advantageous position in the racialized distribution of advantages and disadvantages under a system of racial oppression. These “positional interests” offer an additional incentive for racially privileged agents to participate in social practices that reproduce racial oppression.<sup>13</sup>

In the output-based model, “racism” refers primarily to systems of race-based oppression. This does not mean that the model emphasizes social systems at the expense of individuals and their actions, beliefs, and desires. Individual racism is part and parcel of larger systems of race-based oppression, which also include ideological racism and institutional racism. A full-fledged theory of racist systems ought to examine racist phenomena at all levels of the system (e.g. individual, ideological, institutional) as well as the connections among them. Such an ambitious task is outside the purview of this paper. Instead, I focus my attention on

theorizing institutional racism. My goal in this section is to provide a sketch of the output-based model of racism in order to contextualize the wide-scope theory of institutional racism that I defend in the rest of the paper. Before turning to my theory of institutional racism in section 2, I briefly explain how the output-based model conceives of individual racism (1.1) and ideological racism (1.2).

### *1.1. Individual Racism*

My account of output-based individual racism is informed by psychological theories of racial prejudice. Racial prejudice is a psychological attitude, and therefore has cognitive, behavioral, and affective components.<sup>14</sup> Racial prejudice was traditionally understood as an explicit attitude consisting of conscious and deliberate processes. Today, racial prejudice is also thought to encompass implicit attitudes (also known as implicit biases) in which the cognitive, affective, and behavioral processes can be unintentionally activated or operate outside of conscious awareness. The cognitive component of prejudice is often called stereotype. Racial stereotypes are associations and beliefs about the characteristics of a racialized group and its members, which shape how people think about and respond to the group. The affective component of racial prejudice picks out negative emotions (e.g. fear, hatred) towards members of other racialized groups, which can shape agents' behavior towards those groups and their members.<sup>15</sup>

Finally, the behavioral component of racial prejudice captures actions that maintain or reinforce advantage and disadvantage for racialized groups in accordance with the racial hierarchy in place. These actions are informed by the cognitive and affective components of racial prejudice. Psychologists call this feature of racial prejudice “discrimination”, but I prefer the term “functionally racist behavior” because it captures the broad set of actions that play a functional role in reinforcing systems of race-based advantage and disadvantage (e.g. discrimination, opportunity hoarding, racial apathy, and active ignorance).

The term “functionally racist behavior” also keeps a positive feature of the psychological account of racial prejudice, which is to highlight the racial impact of our actions rather than our intentions. This allows me to get around a common criticism of psychological accounts of racism, which is that they require proof of

intent to harm non-whites in order to ascribe racism to individuals.<sup>16</sup> For example, Stickers argues that this sets the bar for racism too high insofar as it “establishes white people as the sole arbiters of what counts as “racist” because, after all, only they know their intentions”.<sup>17</sup> While I am sympathetic to Stickers’ worry, it is possible to bypass it if ascriptions of individual racism focus on actions that reinforce systems of racial oppression rather than actions that are motivated by race-based ill-will or disregard.<sup>18</sup> On the output-based model of individual racism, the key question is not “Am *I* a racist?” but rather “is *my action* (or inaction) racist insofar as it reinforces racial oppression?” On this model, “individual racism” refers primarily to functionally racist behavior—that is, to the actions (and inaction) of individuals that contribute to the reproduction of a system of race-based advantage and disadvantage, regardless of the agent’s intentions.<sup>19</sup>

## *1.2. Ideological Racism*

My account of ideological racism draws on the Western Marxist theory of ideology. Summarizing this tradition, Shelby defines ideology as a set of widely held beliefs, associations, attitudes, dispositions, habits, and cultural products that misrepresent significant aspects of social reality, and that function to support an unjust social order in virtue of these misrepresentations.<sup>20</sup> Shelby argues that an ideology is racist when its content misrepresents “races” and race relations and that misrepresentation functions to maintain a system of racial oppression. One example of racist ideology is the doctrine of biological racism, which was the referent of “racism” when the term was first coined in the 1930s. According to this set of beliefs, the world population can be classified into hierarchically ordered “races”, which share a biological essence that explains similarities in visible characteristics (e.g. skin color, hair, facial shape), as well as behavioral, cultural, and moral tendencies. According to biological racism, these racial essences are fixed, passed on through bloodlines, and justify the domination of superior “races” (i.e. whites) over inferior “races” (e.g. Blacks). This set of beliefs is ideological because it misrepresents significant aspects of social reality (e.g. it posits the existence of biological and hierarchically ordered “races”) and its misrepresentations serve to justify the oppression of allegedly inferior “races”. “Ideological racism” does not only refer to racist beliefs. Racist ideology also encompasses cognitive



states and processes such as implicit bias and racial stereotypes, as well as affective states such as fear of Black men and indifference towards race-based inequality and non-white suffering.

It is tempting to think of racist ideology purely in terms of individual psychology. In fact, the examples I have given thus far fall under the cognitive and emotional features of racial prejudice (see 1.1). Indeed, one way to think of ideological racism is in terms of mental states and processes that causally contribute to functionally racist behavior (i.e. individual racism). The relevant beliefs, associations, attitudes, dispositions, and habits may lead an individual to engage in functionally racist behavior. In some cases, racist ideology can cause functionally racist behavior independently of the agent's awareness or endorsement of the action. For example, a woman who automatically clutches her purse whenever she encounters a Black man on the street engages in functionally racist behavior even if she does not endorse or recognize her automatic reaction. Still, this act of individual racism is caused by racist ideology because what causes her automatic behavior are her cognitive/affective associations between Black men and criminality/danger.

However, ideology is not reducible to mental states and processes. It also includes cultural products, such as slogans, gestures, rituals, film, music, TV shows, and advertisement. When racist ideology becomes embedded in a society's culture, it acquires a materiality that cannot be explained purely in terms of individual psychology. Thus, we must also theorize racist ideology as a cultural phenomenon. According to Haslanger, we can theorize culture as a web of social meanings or schemas that enable social coordination and action.<sup>21</sup> Culture is akin to a toolkit of habits, skills, and styles from which we draw (often spontaneously and unthinkingly) as we interact with others and with the world. Cultural schemas eventually make their way into the mental states of individual agents (e.g. as implicit or explicit beliefs) but they are not reducible to them. As Haslanger argues, the network of semiotic relations that make up a culture may include non-discursive elements, such as objects and actions that have a public meaning.<sup>22</sup>

Racist ideology as a cultural phenomenon helps explain the prevalence of racist ideology as a psychological phenomenon. In turn, racist ideology as a psychological phenomenon helps explain the

prevalence of individual racism (i.e. functionally racist behavior) and institutional racism. For example, depictions of Black people as violent criminals in films, TV shows and the news (cultural ideological racism) reinforce cognitive and affective associations of Black people with criminality and danger (psychological ideological racism), which in turn helps explain white people's tendency to call the police on Black people for minor disagreements or baseless suspicions (individual racism), as well as the systematic anti-Black violence enacted by police departments across the US (institutional racism). In this way, racist ideology plays the key role of connecting the macro (institutional racism) and micro (individual racism) components of systems of race-based oppression. As a cultural-cum-psychological phenomenon, racist ideology helps explain why individuals engage in functionally racist behavior within and outside institutional contexts.

According to the output-based model, "racism" refers primarily to systems of race-based oppression. Secondly, "racism" may also refer to the components of the system—namely, individual racism, ideological racism, and institutional racism. I define individual racism as actions (and inaction) of individuals that contribute to the reproduction of race-based advantages and disadvantages, regardless of the agent's intentions. I define ideological racism as widely held beliefs, associations, dispositions, habits, and cultural products and schemas that distort our experience and perception of racialized groups and race relations. Crucially, for such distortions to be ideological, they must also function to support, stabilize, or legitimate a system of race-based oppression. We are now ready to turn to the last component of systems of race-based oppression—institutional racism.

## 2. Institutional Racism: Input-Based and Output-Based

In this section I develop a theory of institutional racism in accordance with the wide-scope conception of racism. As such, my theory identifies and explains two models of institutional racism, input-based and output-based. Input-based cases of institutional racism correspond to institutions whose goals, rules, or application of rules are infected by race-based ill-will or disregard. Output-based cases correspond to institutions that contribute to a system of racial oppression and are legitimized by racist ideology. Unlike

advocates of a narrow-scope conception of racism that only recognize the input-based model, I defend a wide-scope theory of institutional racism that accommodates both models. In section 3, I defend this theory of institutional racism from the challenge of conceptual inflation.

So, what is an institution? According to Shelby, institutions are “formal system[s] of roles and rules that enable and regulate sustained cooperative action for some specified purpose”.<sup>23</sup> Institutions rely on the continuous participation of individual agents for their existence. Institutions are recursively constituted by individual agents who coordinate their collective action in accordance with the institutional roles they occupy and the rules associated with those roles. While causally dependent on individual agents, institutions also shape their agency. They allow individuals to occupy institutional roles that simultaneously enable and constrain their action. Institutions open possibilities for action for individuals that they would not have outside the system of roles and rules that make up the institution. However, this same system of roles and rules also sets up behavioral expectations that are formally and informally enforced in order to ensure the stability of the institution.<sup>24</sup> For example, a university is an institution that enables agents who occupy the student role to enroll in courses, attend professors’ office hours and lectures, earn a college degree, and so on. However, the agency of the students is also constrained by the institutional rules that apply to the student role, which may or may not be formally codified (e.g. formal rules regarding academic honesty, the informal rule that students sit rather than stand during class).

### *2.1 The Input-Based Model*

What makes an institution racist? In the input-based model, “institutional racism” refers to institutions whose goals, rules, or application of rules are infected by race-based ill-will or disregard.<sup>25</sup> Garcia offers a prominent account of input-based institutional racism. According to Garcia, an institution is racist when it is infected by the moral vices of race-based ill-will or disregard.<sup>26</sup> For Garcia, these are primarily individual vices—they “lie in the hearts of individuals”—and constitute the core of racism. However, racism can also spread to institutions if racist intentions (i.e. race-based ill-will or disregard) inform the actions of individuals

“when they make and execute the policies of the institutions in which they operate”.<sup>27</sup> For Garcia, institutions are racist because of their input (i.e. race-based ill-will or disregard that has spread from individuals to the institution).

Shelby helps us explain how institutions become infected with racist intentionality.<sup>28</sup> First, the goals of an institution may be racist if it aims to harm the members of a racialized group and those aims stem from race-based ill will or disregard. Racist institutional goals need not be explicit or public; in fact, institutions that appear to have race-neutral goals may be covertly designed to subordinate a racialized group. An institution so designed is racist even if it fails to attain its ends. In other words, whether an institution is racist on the input-based model is not a matter of its output (e.g. negative impact on a racialized group), but of its input (i.e. racist intentionality).

Second, an institution can be racist if its rules incite institutional agents to act with race-based ill-will or disregard. To function appropriately, an institution requires its members to enact the role assigned to them and the rules associated with the role. Institutional rules may be racist even if they don't formally or overtly express race-based ill-will or disregard. These vices may be present in informal or implicit institutional rules. The Ferguson Police Department is an example of an institution with racist rules. The DOJ's Ferguson Report states that one of the implicit rules followed by police officers was to presume Black citizens to be “criminals” and “lacking personal responsibility”.<sup>29</sup> This rule incites FPD police officers to act with race-based disregard for Black citizens, and therefore constitutes the FPD as a racist institution on the input-based model.

Finally, race-based ill-will or disregard can infect the application of institutional rules. Even if the goals and rules of an institution are free of these racist vices, an institution may be racist if its members systematically fail to apply institutional rules equally and consistently because of personal race-based ill-will or disregard. Racially discriminatory application of institutional rules may be conscious or unconscious, but it is always the result of race-based ill or disregard. For instance, even if the FPD reformed its institutional rules, it would remain racist so long as race-based ill or disregard motivates police officers to use excessive force against

Black citizens. Insofar as police officers would systematically fail to apply the correct (non-racist) rule for the use of force when they interact with Black citizens, the FPD would continue to be a racist institution on the input-based model.

To sum up, institutions are racist in the input-based model if their goals, rules, or application of rules are infected by the vices of race-based ill-will or disregard. In this model, institutional racism is parasitic on individual vices. Racism originates in the hearts of individuals and it eventually infects the constitutive features of the institutions in which those individuals participate. Garcia is willing to acknowledge that institutions can be racist in virtue of their input, i.e. individual race-based ill-will or disregard that infects the institution's goals, rules or application of rules. However, he rejects the main insight of the output-based model of racism, which is that institutions can also be racist in virtue of their output. Since my wide-scope theory of institutional racism encompasses both models, I now turn to the output-based model.

## *2.2 The Output-Based Model*

In my output-based model, an institution is racist if (a) it reliably contributes to the reproduction of a system of race-based oppression and (b) it is legitimized by racist ideology.<sup>30</sup> The first condition establishes what kind of output is relevant when ascribing racism to institutions. Systems of race-based oppression are dynamically stable social systems structured by hierarchical racial categories that distribute advantages and disadvantages along racial lines. Therefore, an institution meets condition (a) if it reproduces race-based advantage and disadvantage. Social systems are only able to distribute race-based advantage and disadvantage thanks to the actions of individuals. In my output-based model, "individual racism" focuses on the actions (and inaction) of individuals agents outside institutional contexts. For example, the woman who clutches her purse whenever she encounters a Black man on the street contributes to the symbolic disadvantage of Black men by perpetuating racist stereotypes of Black men as criminals. However, this woman does not perform the action while occupying an institutional role, nor is her action enabled or constrained by an institution.

By contrast, a police officer who racially profiles Black men contributes to the same symbolic disadvantage, but in this case the action is enabled and motivated by his role as police officer and the rules of his police department. A police department that has racial profiling as one of its operative rules will reliably contribute to various forms of Black disadvantage—not just symbolic, but also material because of the institution’s involvement in the criminal justice system. This output will be reliable because the institution’s set of rules and roles will motivate the same institutional action regardless of who occupies the role of police officer. Therefore, the police department in this scenario meets condition (a). This police department would also count as racist on the input-based model insofar as its racial profiling rule can be traced back to race-based ill-will or disregard in the institution’s past or present participants. This is not a problem because the input-based and output-based models of institutional racism are not mutually exclusive. In fact, input-based racist institutions that work as intended will often perpetuate race-based advantage and disadvantage, and therefore will also count as racist on the output-based model.

In order to show the distinctiveness of the output-based model, I introduce an example of an output-based racist institution that would not count as racist on the input-based model. Word-of-mouth hiring is a widely used hiring practice that allows employers to channel job offerings through the personal networks of current employees. This practice is a cost-efficient way for employers to find suitable job candidates because current employees tend to refer people who they are willing to vouch for and who they think would be a good fit for the organization. Let us imagine a small company, Noodle, whose current employees are mostly white and whose CEO decides to use word-of-mouth hiring for recruiting new employees. Let us further imagine that Noodle’s CEO is not using this recruiting technique as a covert way of hiring more whites. In fact, he thinks of himself as “colorblind”—he does not care about the racial composition of his workforce as long as they do their job well. His reasons for adopting word-of-mouth hiring are in no way influenced by race-based ill-will or disregard, but rather by a desire to find suitable candidates. Moreover, the goals, rules, and application of rules at Noodle are also uninfected by race-based ill-will and disregard.

By stipulation, Noodle does not fit the criteria of input-based institutional racism. However, it contributes to the reproduction of race-based advantage and disadvantage by erecting an institutional barrier to employment for Black people. Due to residential and social segregation, Black people are often excluded from white social networks, which due to past and present racist practices, tend to accrue more social capital. Blacks' exclusion from white social networks that channel job offerings through word-of-mouth bars them from one of the most important mechanisms for success in the labor market and is among the causes for the long-standing Black-white unemployment gap.<sup>31</sup> In the context of a racially segregated society in which network effects disadvantage Black job candidates, Noodle's adoption of word-of-mouth hiring has the unintended effect of reinforcing race-based economic disadvantage.<sup>32</sup>

Noodle meets condition (a) in my output-based model of institutional racism. However, having a negative impact on non-whites is not sufficient for ascribing output-based institutional racism. In other words, not all institutional negative impact on non-whites is *racist* impact. In order for the impact to be racist, we must also be able to situate the institution within a broader system of racial oppression. Here is where the connection between institutional racism and another key component of systemic racism—ideological racism—becomes crucial. The function of racist ideology in systems of racial oppression is to ensure the stability of the system by motivating agents to engage in individual racism and input-based institutional racism, and by legitimizing output-based institutional racism. If we can establish that racist ideology supports an institution in either of these ways, we have a reason to believe that the institution belongs to the larger system of race-based oppression. In the input-based model, the connection between racist institutions and racist ideology is straightforward. Racist ideology qua cognitive and affective states (e.g. implicit bias, racial stereotypes, fear of Black men, indifference to Black suffering) motivates race-based ill-will and disregard among individual agents, which eventually infects the institution's goals, rules, and procedures. In the pure case of the output-based model of institutional racism, racist ideology does not motivate race-based ill-will or disregard. However, racist ideology is still operative by distorting our perception of the existence and causes

of racial inequality,<sup>33</sup> and thereby legitimizing the institution's racially disparate impact. In other words, condition (b) for ascribing racism to an institution in my output-based model is that racist ideology legitimizes the institution's contribution to race-based advantage and disadvantage.

According to sociologist Eduardo Bonilla-Silva, one of the main racist ideologies fulfilling this legitimizing function today is colorblind ideology.<sup>34</sup> As an ideal, colorblindness promotes the end of racial distinctions due to their historical use as tools for furthering inequality, stigmatization, and discrimination. As an ideology, colorblindness depicts our society as having already attained racial justice, with the result that race-conscious policies aimed at redressing racial inequality are no longer deemed necessary. Colorblind ideology distorts agents' interaction with the world such that they come to notice little to no racial inequality. To the extent that colorblind ideology recognizes the persistence of racial inequality, it explains it as the result of non-whites' presumed cultural inferiority, or the unfortunate (but otherwise just) result of race-neutral dynamics. These explanations work to legitimize the current racial order as not needing any race-conscious intervention to ameliorate the persistent effects of racial inequality on the life chances of non-whites.

To return to the example of Noodle, colorblind ideology legitimizes its word-of-mouth hiring practice by minimizing or outright masking its causal role in reproducing the Black-white unemployment gap. Rather than situating the racially disparate impact of word-of-mouth hiring in the context of a racially segregated society in which Black social networks carry little social capital due to unaddressed racial harms, colorblind ideology explains the Black-white unemployment gap in terms of Blacks' lack of individual motivation and personal responsibility. By distorting agents' perception of the causes (or even the very existence) of racial inequality, colorblind ideology legitimizes the racially disparate impact of Noodle's word-of-mouth hiring practice. Insofar as colorblind ideology hinders social critique of this practice's impact on Black job applicants, it also thwarts support (both within the institution and in society at large) for efforts to correct this impact. Absent a robust social critique of this practice, Noodle is likely to keep it and will thus continue to contribute to the Black-white unemployment gap. Thus, Noodle meets both conditions for institutional racism in my



output-based model. It (a) contributes to the reproduction of race-based advantage and disadvantage and (b) its racially disparate impact is legitimized by racist ideology.

The goal of this section was to offer a philosophical reconstruction of the concept of institutional racism. The wide-scope theory I defend here encompasses input-based and output-based cases of institutional racism. As I show in the next section, the real target of conceptual inflation critics is the output-based model of institutional racism. These critics often defend a narrow-scope theory of institutional racism that only recognizes the input-based model. By contrast, I argue that a proper understanding of institutional racism must take the perspective of the wide-scope conception and recognize both models. Since conceptual inflation critics already acknowledge the input-based model of institutional racism, my defense will focus on the output-based model.

### 3. “Institutional Racism” and the Challenge of Conceptual Inflation

According to historian George Fredrickson, the term “racism” was sporadically used in the 1920s to describe German theories of racial purity and superiority.<sup>35</sup> However, it was the ascent of Hitler to power and the Third Reich’s embrace of theories of racial essentialism and hierarchy that sparked the widespread use of the term. Magnus Hirschfeld was the first to give real currency to the term “racism”. In his book *Racism*, he set out to refute pseudo-scientific arguments for the existence of hierarchically ordered “races”.<sup>36</sup> Hirschfeld’s *Racism*, written in 1933-34 and posthumously published in 1938, was part of a wider effort by academics and activists to call into question the Third Reich’s use of the “race” idea to justify the exclusion and extermination of Jews. This theoretical-political effort culminated with UNESCO’s 1950 statement on “The Race Question”, which defined racism as a doctrine that “falsely claims that there is a scientific basis for arranging groups hierarchically in terms of psychological and cultural characteristics that are immutable and innate”.<sup>37</sup> It was in this context of political contestation that the term “racism” was originally used to refer to a set of false beliefs used to legitimize a system of racial oppression—in other words, to an ideology. The concept of racism arose in response to a practical problem—in this case, the need to prevent theories of racial difference and hierarchy

from being used to legitimize systems of racial oppression.<sup>38</sup> It was only in the 1960s that popular usage of the term became associated with the narrow-scope conception of racism—that is, to racially hostile and discriminatory behavior that is willful, conscious, or endorsed by individual agents.<sup>39</sup>

Blum, Garcia, and Miles criticize theories of institutional racism on grounds of conceptual inflation. However, only Miles conceives of racism as an ideology, which as we have established was the original meaning of the term. Whereas Garcia's volitional account identifies racism with ill-will and disregard towards another race, Blum has a dual definition of racism as race-based antipathy and inferiorization. Therefore, Blum's and Garcia's respective accounts also inflate the concept of racism with respect to its original meaning. How then could they argue against the conceptual inflation of "racism" without falling into a contradiction? My interpretation is that they distinguish between good and bad cases of conceptual inflation, and that their preferred accounts of racism fall under the good cases, while "institutional racism" falls under the bad cases. If I am correct, Blum and Garcia are not opposed to conceptual inflation per se; their criticisms only target cases of bad conceptual inflation. Therefore, the next step is to identify the criteria for assessing different cases of conceptual inflation.

In this section, I identify and analyze three such criteria: moral-explanatory power (3.1), pragmatic considerations (3.2), and causal-explanatory power (3.3).<sup>40</sup> I draw on Blum, Garcia, and Miles to make explicit all three grounds on which the output-based model of institutional racism has been criticized as a case of bad conceptual inflation. I show that the output-based model of institutional racism is immune to all three versions of the conceptual inflation challenge. In so doing, I reject Blum, Garcia, and Miles' push for a narrow-scope conception of institutional racism that only recognizes the input-based model. By contrast, I defend a wide-scope conception that encompasses input-based and output-based cases of institutional racism. Furthermore, I show that a wide-scope conception of institutional racism positively contributes to moral and causal explanations of racial phenomena, and to the practical anti-racist project of eliminating or mitigating racial oppression.

### *3.1. Moral Explanation*

As moral philosophers, Garcia and Blum are interested in the moral significance of racism. While Garcia argues that an adequate conception of racism must “clarify why racism is always immoral”, Blum claims that a theory of racism must contribute to our moral understanding of race-related wrongs.<sup>41</sup> Underlying this perspective is the idea that “racism” is a term of strong moral condemnation whose main use is to denounce severe racial wrongs. Accordingly, bad conceptual inflation occurs when we use “racism” to refer to race-related phenomena that do not qualify as severe moral ills. Blum rejects the use of the term “racist” to cover all cases of racial insensitivity, racial ignorance, same-race socializing, and racial discomfort and anxiety, regardless of their degree of wrongness.<sup>42</sup> According to him, only cases that involve race-based inferiorization or antipathy are deserving of the strong moral condemnation entailed by the label “racist”. Race-related wrongs do not always reach the degree of moral severity that warrants an ascription of racism.

For Blum, the overuse of the epithet “racist” in everyday race talk is detrimental to our moral understanding of race-related wrongs for two reasons. First, by calling all race-related ills racist, it obscures the great variety of race-related wrongs and hinders the development of a nuanced moral vocabulary for characterizing such wrongs. Second, it encourages misunderstanding of the moral significance of lesser race-related wrongs. This misunderstanding pulls in two directions. One option is to overstate the moral condemnability of lesser racial ills, such as racial ignorance and insensitivity. The second is to deny that these racial ills have any moral significance at all insofar as they fall below the threshold of moral opprobrium required to be considered racist.

Blum’s interest in an adequate moral understanding of the plurality of racial ills informs his criticism of theories of institutional racism. Blum understands ascriptions of institutional racism as primarily concerned with denouncing race-related wrongs at the level of institutions rather than individuals. He worries that such theories obscure the importance of individual moral wrongs in the domain of race.<sup>43</sup> Blum argues that theories of institutional racism reduce racism to a “system of injustice and unequal advantage or power”, with the

consequence of eclipsing individualist theories of racism that focus on individual attitudes and behavior.<sup>44</sup> For him, “institutional racism” replaces an incomplete individualist conception of racism with an incomplete institutionalist conception. Instead, Blum argues, we should recognize the categorial plurality of racism without privileging any one manifestation of it.

However, Blum is not dismissive of the concept of institutional racism altogether. He argues that institutions that express racial inferiorization or antipathy should count as racist. His problem is with defining institutional racism as “a practice that is itself free of racial bias but in its implementation has a disproportionately negative effect on subordinate racial groups”.<sup>45</sup> In other words, the real target of Blum’s conceptual inflation critique is the output-based model of institutional racism. This is clear when he recommends narrowing the scope of “institutional racism” to the input-based model—namely, institutions that are “explicitly racially discriminatory”, “intend a racially discriminatory result”, or whose “actual functioning involves racism or racial discrimination” even if they are not guided by an “official or intentional policy of racism or racial discrimination”.<sup>46</sup> Similarly, Garcia is okay with the concept of institutional racism so long as we limit its referent to institutions whose constitutive features are infected by race-based ill-will or disregard, i.e. input-based institutional racism. For Garcia, the output-based model’s focus on institutions that reproduce Black disadvantage obscures the moral insight that what makes an institution racist is whether it is informed by racist ill-will, regardless of its consequences.

The problem with this critique is that it assumes that all accounts of racism must have moral explanation as their primary aim. However, there are other questions that an account of racism may focus on. For instance, a causal-explanatory approach may focus on identifying the causal mechanisms behind the reproduction of social problems afflicting subordinate racialized groups, such as mass incarceration, police brutality, unemployment, and concentrated poverty.<sup>47</sup> The output-based model of institutional racism is primarily interested in this causal-explanatory project. Of course, questions of moral explanation and causal explanation are both important for achieving a holistic understanding of racism.<sup>48</sup> The causal-explanatory

project that informs the output-based model of institutional racism is compatible with and complements Blum's and Garcia's moral-explanatory project. Given that the output-based model's primary aim is to answer the causal-explanatory question, criticizing it for failing to answer the moral-explanatory question is misguided.

Blum and Garcia could reply that the output-based model of institutional racism does not merely fail to contribute to the moral-explanatory project, but actually works against it. For example, Blum argues that the output-based model reduces racism to a system of injustice, thus ignoring racism at the individual level. If Blum is correct, the output-based model wrongly narrows the reasons for condemning racism, which undermines the moral-explanatory project. This charge casts doubt on my earlier picture of the moral-explanatory and causal-explanatory projects as compatible paths towards a richer understanding of racism. However, the output-based model that I articulate in sections 1 and 2 is immune to this objection. My output-based model does not deny the existence of individual racism. To the contrary, it calls for an analysis of the interactions between individual racism, ideological racism, and institutional racism as a means for explaining the stability of systems of racial oppression. Paying attention to individual, ideological, and institutional racism is necessary for a thorough causal explanation of systemic racial oppression.

Furthermore, although the output-based model emphasizes causal-explanatory questions, the insight that an institution can contribute to racial oppression even if its individual members are not motivated by racial ill-will has important moral implications. It shows that an institution may be the subject of moral concern even if its members do not fit traditional criteria for assigning moral condemnation. The output-based model of institutional racism shifts our moral concern away from the practice of interpersonal moral condemnation. Instead, it focuses our moral attention on institutions that contribute to social problems faced by non-white communities (e.g. mass incarceration, housing insecurity, racial inequality) regardless of the intentions of the institutions' members. Of course, the output-based model does not deny that actions driven by race-based ill-will or disregard should be condemned. Rather, it expands the scope of our moral concern to include both personal and political morality.<sup>49</sup>

The input-based model takes the perspective of personal morality, which emphasizes the ethics of individual character and interpersonal conduct. In this model, the guiding moral question is whether the individual agents that make up the institution deserve the strong moral condemnation associated with the label “racist”. Traditionally, interpersonal moral condemnation requires drawing a causal link between the action and the agent’s sphere of intentional control. Only when the action can be shown to stem from race-based ill-will is the agent thought to deserve the strong moral condemnation that accompanies the label “racist”.<sup>50</sup> By contrast, the output-based model shifts our moral attention to matters of political morality, which deals with the justice of society’s institutional arrangements. From a political morality perspective, we should focus on the actions and processes that contribute to the reproduction of systems of racial oppression whether or not we can attribute those actions and processes to race-based ill-will. Political morality takes the perspective of the victims of racial oppression and directs our moral attention to the social problems that they face in their everyday lives, such as mass incarceration, police brutality, housing insecurity, and racial inequality.<sup>51</sup> Rather than limiting our moral understanding of racial matters, the output-based model of institutional racism actually expands the scope of our moral attention beyond interpersonal moral condemnation. In so doing, it also advances the moral-explanatory project.

Blum also worries that the output-based model implies that all institutions that reproduce racial disadvantage are driven by racist motives and are deserving of moral opprobrium, which would rule out that they could ever be morally acceptable.<sup>52</sup> For Blum, this blanket moral assessment of all institutions that have a negative impact on non-whites prevents a careful weighing of the benefits against the race-related costs of such institutions. For instance, the seniority system is a practice that is itself free of racial bias, but which has the effect of contributing to the Black-white unemployment gap.<sup>53</sup> In spite of its contribution to a system of racial disadvantage, Blum argues, seniority is a valuable institutional rule insofar as it provides job security, protects workers from management arbitrariness, and discourages destructive competitiveness among

workers. However, once seniority is considered a case of institutional racism, it is placed beyond the pale before we can examine its ethical pros and cons.

Pace Blum, ascribing racism to an institution in the output-based model does not imply that the institution is beyond the pale. The input-based model of institutional racism restricts ascriptions of racism to the most severe racial wrongs, and therefore tends to place the institution beyond the pale. However, this standard does not apply to the output-based model. Ascriptions of racism in the output-based model convey varying degrees of condemnatory force.<sup>54</sup> Thus, to claim that an institution is racist partly in virtue of reproducing racial advantage and disadvantage does not by itself preclude a further assessment of its moral pros and cons. It may be that the seniority system advances workers' rights while simultaneously being racist on account of contributing to racial oppression. Following the paradigm of political morality, the decision of whether the practice should be overhauled or abandoned altogether will hinge on a practical assessment of its overall impact on social justice. The output-based model of institutional racism merely provides the descriptive information required for making an informed assessment of this kind. It does not settle the normative question one way or the other.

Garcia also thinks that the output-based model of institutional racism gets in the way of an adequate moral understanding of race-related wrongs. In particular, he argues that its focus on systems of racial oppression encourages us to see the vice of racism as something in the world rather than in ourselves.<sup>55</sup> According to Garcia, the output-based model is wedded to the idea that racism requires the existence of a system of power that disadvantages people of color. This model entails that racism is only morally significant when it is tied to power, which fails to appreciate that individuals can be racist in the absence of a system of racial oppression. As a result, it lets individuals who harbor ill-will against members of a racialized group, but who may not live in a system of racial oppression, off the hook.

My output-based model does not dismiss individual racism. To the contrary, it considers individual racism as one of the key components of systems of racial oppression, along with ideological racism and

institutional racism. However, it is true that individual racism is more morally significant when it is part of a system of racial oppression. We must keep in mind that individual racism in the output-based model is not restricted to actions driven by race-based ill-will. Earlier, I defined individual racism as actions (and inaction) of individuals that contribute to the reproduction of a system of race-based advantage and disadvantage, regardless of the agent's intentions. Some individual actions that would count as racist in the output-based model would seem of little moral significance unless we draw their connection to a larger system of racial oppression. For example, a professor that consistently mixes up the names of his two Black students may seem to commit a mostly harmless microaggression. However, if we consider that the professor's action takes place on a college campus in which Black students are constantly harassed by campus security and have their intellectual abilities constantly challenged by faculty and other students, the professor's microaggression no longer appears as an incidental snub, but as part of a broader pattern of individual and institutional actions that systematically harm Black students on that college campus. The fact that the professor's microaggression is part of a larger system of racial oppression gives his action a moral significance it would not have outside of that system. More generally, the output-based model morally evaluates individual acts of racism not as isolated, one-off occurrences, but as part of a larger system of racial oppression that is deeply entrenched in our society.<sup>56</sup>

Let us now turn to an example of individual racism motivated by race-based ill will. Imagine a relatively powerless racist who commits an atrocious racist act—an anti-Asian hate crime involving murder.<sup>57</sup> If this act were to take place in contemporary America, it would count as racist in the output-based model insofar as it partakes in a larger system of anti-Asian oppression that often includes acts of violence, as the recent wave of Anti-Asian hate crimes attests.<sup>58</sup> It would also count as racist in the input-based model insofar as the act is clearly motivated by race-based ill-will. What about its moral significance? Since it involves murder, the act belongs to the highest rank of moral wrongs. The fact that it is also a hate crime adds to its moral significance. According to the Department of Justice, hate crimes are “crimes motivated by bias against race, color, religion,



national origin, sexual orientation, gender, gender identity, or disability”.<sup>59</sup> In other words, these crimes have a higher severity because they are motivated by ill-will against vulnerable social groups in our society.

Would this racist act have the same moral significance in a counterfactual society in which race is not a social category around which a system of oppression is organized? In the input-based model, the answer may be yes. The added severity of the murder stems from its being motivated by the vice of race-based ill-will. Arguably, this vice would be equally morally significant even if the perpetrator of the hate crime lived in a society in which being Asian had as much social relevance as being left-handed has in our actually existing society.<sup>60</sup> By contrast, the answer in the output-based model is no. This is because the fact that we live in a racially oppressive society is morally significant for the output-based model. We care about racist acts partly because they partake in and reproduce a larger system of oppression that constrains the life chances of non-whites. These constraints range from almost harmless (e.g. Asian students whose names are constantly mixed up by their professor) to deadly (e.g. anti-Asian hate crimes).

The counterfactual scenario would be akin to a scenario of a murder motivated by anti-left-handed ill-will in our actually existing society. This act would certainly warrant strong moral condemnation qua murder, but it would not have the additional moral severity associated with hate crimes, which as we have seen is restricted to animosity against recognizably oppressed groups. The anti-left-handed ill-will motivating the murder would be considered a psychological anomaly of little significance beyond that isolated case. By contrast, the racial ill-will that motivates anti-Asian hate crimes in our society is not idiosyncratic, but the product of widespread anti-Asian stereotypes and prejudices. To sum up my reply to Garcia, the output-based model of institutional racism does not let individual racists off the hook in a counterfactual non-racist society. The relatively powerless racist is still morally responsible for murder and for the ill-will that motivates his action. What this model does is highlight the added moral significance that individual racism has when it takes place in a social context of systemic racial oppression.

Garcia also rejects the output-based model of institutional racism because he associates it with the idea that institutions can be racist merely because of their consequences. For Garcia, the idea that the effects of an institution can render it racist is unsustainable. To show this, he draws on a common criticism of race-based affirmative action; namely, that it breeds racist resentment and acts of vengeance and reinforces negative stereotypes against its beneficiaries.<sup>61</sup> If affirmative action institutions have the effect of perpetuating racist beliefs and acts, they should be considered to be racist on the output-based model. But it is preposterous to argue that well-meaning efforts to redress racial injustice, such as affirmative action, are racist merely because they also have the effect of reinforcing racist beliefs and actions. For Garcia, the fact that the output-based model must consider affirmative action institutions to be racist shows that it is unacceptably harsh and indiscriminate in its application of the term “racism”.

Garcia is correct that the output-based model ascribes racism to institutions partly in virtue of their consequences—namely, their causal contribution to a system of racial oppression. However, as I argued in section 2, a further necessary condition for an institution to be racist on the output-based model is that it be legitimized by racist ideology. Since affirmative action is not legitimized by racist ideology—in fact, affirmative action is fiercely opposed by the racist ideology of colorblindness—it is not correct to categorize it as racist. Thus, Garcia’s affirmative action *reductio* does not apply to the output-based model of institutional racism.<sup>62</sup> Still, I agree with him that institutions cannot be racist merely because of their consequences (i.e. negative impact on non-whites). They must also be legitimized by racist ideology to properly count as racist.

### *3.2. Pragmatic Considerations*

The output-based model of institutional racism has also been criticized as a case of bad conceptual inflation for pragmatic reasons. For Blum, the output-based model is part of a broader tendency in contemporary race talk to overuse the epithet “racist”. This tendency is problematic because the label “racism” carries an implication of strong moral condemnation, which makes it a “conversation stopper”.<sup>63</sup> Applying the label “racist” to lesser moral wrongs leads to an “inhibiting fear” among whites of being called racist.<sup>64</sup>

Moreover, ascriptions of institutional racism may imply that the racial disparities caused by institutions are ultimately rooted in the individual prejudice of their members.<sup>65</sup> Accordingly, the overuse of the term “racism” encourages defensive reactions among whites, and prevents meaningful conversation across racial lines.

According to Blum, the tendency to overuse the epithet “racist” is rooted in a widespread assumption that race-related phenomena are either “racist” or morally insignificant.<sup>66</sup> Faced with the choice of “racism or nothing”, those seeking to bring attention to race-related wrongs are tempted to label all such wrongs as “racist”, regardless of their degree of severity. However, this strategy is counterproductive because it weakens the moral force of the term and contributes to the banalization of racism and other racial wrongs. It encourages those accused of “racism” to not take the charge to heart, which has the undesirable consequence of undermining the condemnatory power of the term.

The pragmatic version of the conceptual inflation challenge focuses on the practical consequences of our use of the term “racism”. From its inception in the 1930s, the concept of racism has always served a pragmatic goal—namely, to confront systems of racial oppression. More generally, all (critical) theories of racism share the pragmatic goal of eliminating or mitigating the effects of racial oppression.<sup>67</sup> I take it that, for Blum, this pragmatic goal is best served by “facilitat[ing] communication between [racialized] groups about the character, forms and extent of racism (and other race-related ills)”.<sup>68</sup> Insofar the output-based model of institutional racism can alienate whites and foreclose meaningful cross-racial conversations, it can also undermine the anti-racist goal of eliminating or mitigating the effects of racial oppression. This leads Blum to criticize the output-based model of institutional racism as a case of bad conceptual inflation on pragmatic grounds.

Blum’s pragmatic critique rests on the unstated premise that the pragmatic anti-racist goal of eliminating or mitigating the effects of racial oppression is best served by facilitating cross-racial communication and conciliation. However, there are other equally plausible strategies for eliminating or mitigating racial oppression that do not center the feelings of white people.<sup>69</sup> For instance, in their

foundational account of institutional racism, Ture and Hamilton focus on the interpersonal and institutional harms that are systematically inflicted on Black people.<sup>70</sup> Their anti-racist strategy prioritizes diagnosing and solving pressing social problems in Black communities, such as food and housing insecurity, and substandard health care. Reflecting on their failed attempts at building coalitions with white liberals that truly aimed at Black liberation, Ture and Hamilton reject the idea that the prospect of alienating white allies should guide the anti-racist tactics of Black people.<sup>71</sup>

In addition, theories of racism may also prioritize non-pragmatic goals, such as achieving moral understanding of race-related wrongs or explaining the causal mechanisms behind the reproduction of racial oppression. In the output-based model, the primary aim of the concept of institutional racism is to provide causal explanations of the reproduction of racial oppression at the level of institutions. This causal-explanatory project also has a pragmatic import. We care to explain the causal mechanisms that reproduce systemic racial oppression in order to devise better intervention strategies to confront that system. Thus, the output-based model of institutional racism is also committed to the pragmatic anti-racist goal of ending or ameliorating racial oppression. Unlike Blum, the output-based model is not pre-theoretically committed to the strategy of promoting cross-racial conciliation as the primary means to fulfill the anti-racist goal. The output-based model's preferred means to pursue the anti-racist goal is under constant revision and rests on an assessment of the pros and cons of different strategies in light of our best available explanation of the workings of systemic racial oppression. Given the plurality of strategies for attaining the anti-racist goal of ending or ameliorating racial oppression, Blum is wrong to criticize the output-based model of institutional racism for failing to live up to his preferred strategy of promoting cross-racial conciliation. This is especially problematic because he assumes that cross-racial conciliation is the best strategy for ameliorating or eliminating racial oppression without first engaging in the causal-explanatory work of explaining how systemic racial oppression works. As Miles reminds us, "if the analysis is wrong, then it is likely that the political strategy will not achieve the intended objectives".<sup>72</sup>

### *3.3. Causal Explanation*

Miles' conceptual inflation critique predates Garcia's and Blum's. As a sociologist, his main worry is that the conceptual inflation of "racism" weakens the causal-explanatory power of the concept.<sup>73</sup> For Miles, the goal of a theory of racism is to provide an accurate description of the phenomena that fall under the concept and an explanation of how such phenomena contribute to socially relevant processes—in this context, to the persistence of racial inequality. Miles argues that a theory of racism as ideology is the best strategy to attain this causal-explanatory goal. He defines racism as an ideological, representational phenomenon that can be distinguished from other ideologies in virtue of its content, which includes among other things, the sorting of human populations according to biological and/or somatic characteristics, and the attribution of negatively evaluated characteristics (biological or cultural) to out-groups and positive ones to in-groups.<sup>74</sup>

Like Blum and Garcia, Miles does not reject the concept of institutional racism altogether. He recommends a narrow definition of institutional racism as exclusionary practices that embody (or institutionalize) a racist discourse to the extent that they a) were originally informed by such a discourse but are no longer justified by it, or b) are legitimized by a racist discourse that was formerly explicit but has since eliminated its explicit racist content while still carrying the same meaning.<sup>75</sup> For Miles, "institutional racism" should not refer to exclusionary practices per se—this is the trap of ascribing racism purely based on impact. Rather, the term should be restricted to cases in which we can demonstrate that racist ideology had a determinate influence on the exclusionary practice. While "determinate influence" is a very open-ended locution, Miles seems to have in mind cases in which racist ideology infects the wishes, desires, and intentions of institutional actors with race-based ill-will.<sup>76</sup> Thus, as with Blum and Garcia, Miles defends a narrow account of institutional racism that only recognizes the input-based cases in which race-based ill-will infects the goals, rules, and procedures of an institution. The real target of his causal-explanatory conceptual inflation critique is the output-based model of institutional racism.

This essay does not intend to fully substantiate the causal-explanatory power of the output-based model of institutional racism. The causal role of institutions in the reproduction of systemic racial oppression relative to other causal factors (e.g. sub-individual, individual, cultural, structural) is an empirical question best addressed by social science and public policy. It could turn out that focusing on racist institutions is not the best way to causally explain the persistence of race-based advantage and disadvantage. In the end, the output-based model of institutional racism must vindicate its causal-explanatory credentials in the tribunal of empirical social science and public policy. However, Miles' critique does not address the empirical track record of the output-based model. He criticizes the very conceptual viability of the model's causal-explanatory project. In other words, Miles depicts the output-based model of institutional racism as unable to pose conceptually coherent explanations of racial advantage and disadvantage. It is at this level of analysis that I reply to Miles. In what follows, I vindicate the output-based model of institutional racism as a viable causal-explanatory project.

The output-based model of institutional racism emphasizes the role that institutions play in sustaining race-based advantage and disadvantage. For Miles, the problem with using "racism" to refer to institutions that perpetuate Black disadvantage is that it excludes institutions that aim to perpetuate such disadvantage but fail to realize this goal. I agree that the output-based model is unable to accommodate such cases, but this is where the input-based model comes in. Miles is correct that an institution with racist goals, whether or not these become realized, should count as racist. He is also correct that the output-based model of institutional racism focuses on the causal role that institutions play in the reproduction of racial oppression. That said, insofar as the wide-scope conception that I favor recognizes both models of institutional racism, it is perfectly capable of accurately describing Miles' example (i.e. an institution that has racist goals, but fails to realize them) as racist.

Miles also worries that the output-based model assumes that racist institutions affect Black people exclusively and in virtue of their Blackness.<sup>77</sup> He claims that for the output-based model of institutional racism

to have causal-explanatory power, it must prove, and not merely assume, that the institutional disadvantage faced by Blacks is not also faced by other groups.<sup>78</sup> Miles illustrates this point with the practice of word-of-mouth hiring. According to Miles, the output-based model would consider an institution that adopts this practice to be racist in virtue of its negative impact on Black applicants. Insofar as the pool of potential applicants is made up of predominantly-white social networks, most Black applicants are effectively excluded from the job search. However, Miles argues, the practice also excludes other groups that are underrepresented in the workplace. For example, if women, Jewish or Irish people are also underrepresented in the workplace, they will also be disadvantaged by the network effects of word-of-mouth hiring. If Blacks are not the only group disadvantaged by word-of-mouth hiring, it is misleading to call such a practice racist, as if they were the only group that was kept out of the hiring process. As Miles puts it, “by what logic does one identify institutional racism as a specific phenomenon when other people are also excluded by the identical practice?”<sup>79</sup>

Here, Miles misconstrues the output-based model of institutional racism, which focuses on institutions that reproduce racial disadvantage, not just Black disadvantage. Racial oppression does not operate on a Black-white binary. There are other racialized groups besides Blacks and whites that face varying degrees of race-based advantage and disadvantage. Thus, it is false that this model assumes that racist institutions affect Black people exclusively. However, there is a grain of truth in Miles’ assertion that the output-based model claims that racist institutions affect Black people in virtue of their Blackness. Institutional racism is part of a system of racial oppression that distributes advantages and disadvantages along racial lines. Thus, it is no accident that racist institutions have a negative impact on non-whites. This is because racist institutions target non-whites in virtue of their membership in a subordinate racialized group.<sup>80</sup>

That said, racist institutions can also disadvantage groups other than non-whites. In his discussion of word-of-mouth hiring, Miles mentions three groups (women, Irish people, and Jewish people) that are not currently considered to be non-white, but who are also disadvantaged by the practice.<sup>81</sup> In the case of Irish and Jewish people, both have a history of being treated as members of inferior racialized groups in American

society. Miles uses this fact to claim that the output-based model cannot account for racism against the Irish and Jews because it restricts racism to domination of Blacks by whites. Beyond the fact that the concept of institutional racism does not follow this Black-white binary, it also does not rule out that the Irish and Jews were victims of systemic racial oppression in the United States back when they were treated as subordinate racialized groups. However, more recently, both groups have for the most part been assimilated into whiteness and do not face the same degree of social and residential segregation as Blacks, Latinxs and Native Americans.<sup>82</sup> Given that racial segregation is a crucial mechanism whereby word-of-mouth hiring reproduces racial disadvantage, the case of Irish and Jewish people is not a good counterexample. Insofar as both groups now partake in white social circles, they are not subject to the exclusionary dynamics that result from word-of-mouth hiring through predominantly-white social networks.

The case of women is more complicated because of the role played by intersecting systems of oppression. Evidently, non-white women are affected by a similar dynamic as non-whites more generally when it comes to exclusion from white social circles and the concomitant exclusion from word-of-mouth job offerings.<sup>83</sup> However, I read Miles as arguing that the job opportunities of women—regardless of race—are negatively affected by word-of-mouth-hiring because of their lack of representation in the workplace qua women. Miles is right that women’s job opportunities are negatively affected by word-mouth-hiring insofar as women are systematically excluded from men’s social networks (i.e. old-boy networks), which also accumulate the most social capital. In fact, there are mechanisms other than social segregation that lead women not to be referred for job offers channeled through personal connections (e.g. gender bias, sexist stereotypes, and deflated judgment of women’s skills, competence and potential).<sup>84</sup> Thus, word-of-mouth hiring also undermines women’s job prospects, independently of the mechanisms that explain its negative impact on non-white job prospects. However, all this shows is that word-of-mouth hiring is a racist *and* a sexist practice. Pace Miles, the output-based model of institutional racism does not assume that a racist institution cannot participate in other systems of oppression. A racist institution need not target non-whites exclusively. Given



the complex nature of social dynamics, a racist institution may also turn out to negatively affect social groups along other axes of oppression.

Miles also criticizes the output-based model of institutional racism for denying that “intentionality or motivation are measures of the presence or absence of racism”.<sup>85</sup> By focusing on impact, the output-based model marginalizes the determination of causality in ascriptions of institutional racism. In other words, since identifying negative impact on non-whites suffices for ascribing racism to institutions on the output-based model, the further task of determining the causes for that impact (e.g. racist ideology, labor market dynamics, Blacks’ agency) is deemed unnecessary. However, this task is crucial because different intervention strategies will be required if the causes are ideological beliefs or race-based ill-will as opposed to the unintended outcome of social processes.<sup>86</sup> “If the analysis is wrong, then it is likely that the political strategy will not achieve the intended objectives”.<sup>87</sup>

Miles is correct that the output-based model of institutional does not emphasize racist intentionality and motivation. However, the input-based model can account for institutions whose goals are driven by racist motivations or whose members act out of racist intentions (i.e. race-based ill-will). Unlike critics of conceptual inflation who want to restrict the concept of institutional racism to the input-based model, the wide-scope conception I defend recognizes that the input-based and output-based models are both key to theorizing institutional racism. Furthermore, it is not true that the output-based model engages in explanatory reductionism. A focus on the racist impact of institutions does not absolve us from analyzing the various causal processes that generate that impact. On the contrary, this task is key to the causal-explanatory project of the output-based model. After all, the point is to show that institutions can contribute to the reproduction of racial disadvantage even if they are not guided by racial ill-will. Accordingly, the causal-explanatory goal of the output-based model is to identify the various causal mechanisms that explain how these institutions can have a reliable racially disparate impact. In this essay I have highlighted word-of-mouth hiring, but I could have also discussed other causal mechanisms such as statistical discrimination, residential and social

segregation, unequal education opportunities, spatial mismatch, the shift in the economy from manufacturing to service industries, and so on.

If the goal of the causal-explanatory project is to identify the causal mechanisms that reproduce racial advantage and disadvantage at the institutional level, why is it important that we apply the term “racism” to the relevant institution?<sup>88</sup> In other words, what causal-explanatory value does the label “racist” add once we have figured out all the causal pathways that account for the maintenance of systemic racial oppression? The answer to this question lies in a key feature of my output-based model of racism, which is that all racist phenomena are interrelated parts of a larger system of racial oppression. Thus, ascriptions of institutional racism on the output-based model entail two causal-explanatory claims. First, that the relevant institution partially explains the persistence of race-based advantage and disadvantage. And second, that in order to appreciate the institution’s causal-explanatory power, we must situate it as part of a larger system of race-based oppression. The first causal-explanatory claim follows from the task of identifying the mechanisms that cause the institution to have a negative racial impact (e.g. word-of-mouth hiring coupled with network effects). Once we pinpoint these causal mechanisms, we are entitled to claim that the institution partially explains race-based advantage and disadvantage. The second causal-explanatory claim goes beyond the identification of causal pathways. It argues that our understanding of these causal pathways will improve if we situate them within a larger system of racial oppression made up of individual actions, ideologies, institutions, and other social processes. Once they are situated in this way, the causal pathways no longer appear as isolated and accidental features of social life, but as interconnected elements of a system of racial oppression.<sup>89</sup> This second claim is what grants the term “racism” its distinctive causal-explanatory power.

Moreover, the fact that racist institutions reliably reproduce racial disadvantage is not accidental. While the racist impact of these institutions may not be actively intended by their members, the reliability of this impact is (partially) explained by the legitimizing function of racist ideology. The output-based model of racism aims to explain the persistence of systems of racial oppression. This causal-explanatory project relies on

clarifying the causal connections between key components of the system—individual racism, ideological racism, and institutional racism. Ideological racism plays a key role in the maintenance of the system by motivating individuals to engage in functionally racist behavior and participate in racist institutions. Racist ideology shapes individual behavior via psychological mechanisms, such as beliefs, associations, dispositions, and habits, as well as cultural mechanisms, such as the cultural schemas that filter and shape our experiential and interpretive interaction with the world and others. Another function of ideological racism is to legitimize systems of racial oppression by distorting agent’s perception of the system and its component parts. This second feature of ideology is important because it explains why individuals who do not participate in racist institutions may still fail to recognize those institutions as morally problematic. Once racist ideologies like colorblindness become embedded in a society’s cultural schemas, institutions that have a disparate racial impact but no trace of racial ill-will are perceived as morally legitimate by the majority of the population. This allows these racist institutions to remain unchallenged and keep contributing to the reproduction of racial oppression.

Pace Miles, the output-based model of institutional racism offers an insightful causal explanation of the persistence of systemic racial oppression. The output-based model explains the psychological and cultural mechanisms through which racist ideology motivates individual participation in racist institutions and legitimizes those institutions in the eyes of the general public. Miles is correct that the output-based model does not explain the causal origin of racist institutions solely in terms of racial ill-will. However, this is because it focuses on institutions with negative racial impact that are not infected by racial ill-will. Agents that participate in those institutions may be acting out of ignorance, habit, or pursuing their own (non-race-related) ends. To return to the example of Doodle, its adoption of word-of-mouth hiring is not motivated by racial ill-will. Instead, Doodle is guided by the race-neutral goal to increase the efficiency of the hiring process. Given the race-neutral motivation of Doodle’s word-of-mouth hiring, the relevant causal explanation is not whether race-based ill-will influenced Doodle’s adoption of the practice. Rather, we need to explain what

allows the practice to remain unchallenged and to continue to have a negative impact on non-white job seekers. It is not that the output-based model of institutional racism fails to pay attention to the causal determinants of racist institutions. Rather, the output-based model pays attention to the kind of causal factor that is relevant to its causal-explanatory project. In the output-based model, the focus is on the various causal mechanisms that explain the institution's reliable racist impact, and by extension, its contribution to the reproduction of systemic racial oppression. In the input-based model, race-based ill-will is the defining explanatory factor. A wide-scope conception of institutional racism can take up either of these causal-explanatory projects depending on the matter at hand.

I have responded to each of Miles' criticisms of the output-based model of institutional racism as a causal-explanatory concept. The causal explanations that the output-based model cannot handle can be covered by the input-based model. This is why I endorse a wide-scope theory of institutional racism that encompasses input-based and output-based cases. This wide-scope theory of institutional racism contributes to the development of more fine-tuned intervention strategies aimed at overhauling or abolishing racist institutions, along with their contribution to racial oppression. As I made clear at the outset of this section, my goal here was not to empirically substantiate the causal-explanatory power of the output-based model. Rather, my main concern was to defend the conceptual viability of this model, which has become prominent among anti-racist activists, social scientists, and policymakers.

### Conclusion

In this paper, I showed that conceptual inflation is not a problem for a wide-scope theory of institutional racism that accommodates the input-based and output-based models. Since conceptual inflation critics already recognize the input-based model, I focused my efforts on defending the output-based model. Critics argue that the output-based model of institutional racism contributes to bad conceptual inflation because it obscures our moral understanding of racial phenomena, hinders the pragmatic goal of ending or ameliorating racial oppression, and undermines the project of causally explaining the persistence of racial

disadvantage. I showed that the output-based model of institutional racism is immune to all three versions of the conceptual inflation challenge. The output-based model of institutional racism offers moral and causal explanations of racial disadvantage at the institutional level, which complement rather than replace the insights from the input-based model. Insofar as a richer causal explanation of the reproduction of racial disadvantage may contribute to the development of better intervention strategies, the output-based model also advances the pragmatic goal of eliminating or ameliorating racial oppression. Of course, the causal-explanatory power of the model—as well as the success of the intervention strategies inspired by it—must be constantly tested by social scientists, policymakers, and activists, who will in turn revise the model in light of their findings. Part of what it means to treat “racism” as not only a moral-explanatory concept, but also a pragmatic and causal-explanatory concept is that these empirical matters will bear on our normative judgments of what counts as a good (or bad) case of conceptual inflation.

## Endnotes

\* I would like to thank Eric Bayruns García, Daniel Brinkerhoff Young, Robert Gooding-Williams, Yarran Hominh, Annette Martín, Alberto Urquidez, Philip Yaure, and my colleagues in the Temple University philosophy department for their feedback at various stages of the project. I am also grateful to the Ethics reviewers and editors who provided substantial comments.

---

<sup>1</sup> Michael Brownstein and Jennifer Saul, eds., *Implicit Bias and Philosophy* (Oxford: Oxford University Press, 2016); Neil Levy, “Am I a Racist? Implicit Bias and the Ascription of Racism,” *Philosophical Quarterly* 67 (2017): 534–51. It is worth noting that there is disagreement among philosophers as to whether implicit bias should be categorized as racist, as can be seen in the differing answers to this question by the contributors of *Implicit Bias and Philosophy*. In his standalone article, Levy concludes that explicit egalitarians who harbor implicit bias are somewhat racist, but not racists in any full-blown sense.

<sup>2</sup> Lauren Freeman and Jeanine Weekes Schroer, eds., *Microaggressions and Philosophy* (New York: Routledge, 2020).

<sup>3</sup> Here I follow Hardimon’s distinction between narrow-scope and wide-scope conceptions of racism in Michael Hardimon, “Should We Narrow the Scope of ‘Racism’ to Accommodate White Sensitivities?,” *Critical Philosophy of Race* 7 (2019): 223–46. See also Tommie Shelby, “Racism, Moralism, and Social Criticism,” *Du Bois Review: Social Science Research on Race* 11 (2014): 57–74. This move towards a wide-scope conception of racism in lay discourse has not been uniform or total. Many still use “racism” primarily to refer to racial bigotry and ill-will. However, the wide-scope conception has become dominant in the discourse of anti-racist movements and their allies, which in turn is highly influential in contemporary lay race talk.

<sup>4</sup> J. L. A. Garcia, “Current Conceptions of Racism: A Critical Examination of Some Recent Social Philosophy,” *Journal of Social Philosophy* 28 (1997): 5–42; J. L. A. Garcia, “Philosophical Analysis and the Moral Concept of Racism,” *Philosophy & Social Criticism* 25 (1999): 1–32; Lawrence Blum, “*I’m Not a Racist, But...*”: *The Moral Quandary of Race* (Ithaca: Cornell University Press, 2002); Lawrence Blum, “What Do Accounts of ‘Racism’ Do?” in *Racism in Mind*, ed. Michael Levine and Tamas Pataki (Ithaca: Cornell University Press, 2004). Joshua Glasgow, “Racism as Disrespect,” *Ethics* 120 (2009): 64–93. Elizabeth Anderson, *The Imperative of Integration* (Princeton: Princeton University Press, 2010).

<sup>5</sup> Shen-yi Liao and Nat Hansen, “‘Extremely Racist’ and ‘Incredibly Sexist’: An Empirical Response to the Charge of Conceptual Inflation,” *Journal of the American Philosophical Association* 9 (2023): 72–94; Alberto Urquidez, “White Individualism and the Problem of White Co-Optation of the Term ‘Racism,’” *Radical Philosophy Review* 25 (2022): 161–190; Shelby, “Racism, Moralism, and Social Criticism”; Hardimon, “Should We Narrow the Scope?”

<sup>6</sup> Robert Miles, *Racism, 2<sup>nd</sup> edition* (London: Routledge, 2003).

---

<sup>7</sup> Hardimon, “Should We Narrow the Scope?,” 226.

<sup>8</sup> Garcia, “The Heart of Racism,” 11.

<sup>9</sup> My aim in this section is to develop an internally coherent account of the output-based model of racism. To do so, I start from the contributions of anti-racist theorists and activists. However, I also revise and discard some of their claims in order to keep the account internally coherent. When needed, I make unstated claims and assumptions explicit, fill in theoretical gaps, and re-interpret features of the output-based model so that it can withstand the conceptual inflation challenge in section 3.

<sup>10</sup> Charles W. Mills, “White Supremacy,” in *The Routledge Companion to Philosophy of Race* (Routledge, 2017). Eduardo Bonilla-Silva, “Rethinking Racism: Toward a Structural Interpretation,” *American Sociological Review* 62 (1997): 465–80. Sally Haslanger, “Racism, Ideology, and Social Movements,” *Res Philosophica* 94 (2017): 1–22.

<sup>11</sup> I take this example from Lawrence Bobo et al., “Laissez-Faire Racism: The Crystallization of a Kinder, Gentler, Antiblack Ideology Search within Citing Articles,” in *Racial Attitudes in the 1990s: Continuity and Change*, by Steven A. Tuch and Jack K. Martin (Greenwood Publishing Group, 1997). The idea of racist systems as dynamically stable comes from Haslanger, “Racism, Ideology, and Social Movements,” 17.

<sup>12</sup> For simplicity of exposition, in the last few sentences I referred to the U.S. racist system as if it were an agent capable of adapting, distributing resources and justifying its racial impact. I want to clarify that I am not committed to an ontology of social systems as group agents. To take List and Pettit’s account of agency, I do not claim that social systems have representational states, motivational states, or a capacity to process them and act on their basis, within feasible limits. In my view, to the extent that social systems “act”, they only do so through the actions (sometimes coordinated, sometimes not) of the individual agents who make up the system. See Christian List and Philip Pettit, *Group Agency: The Possibility, Design, and Status of Corporate Agents* (Oxford; New York: Oxford University Press, 2013), 20.

<sup>13</sup> Douglas Porpora, *Reconstructing Sociology: The Critical Realist Approach* (Cambridge: Cambridge University Press, 2015).

<sup>14</sup> John F. Dovidio et al., “Prejudice, Stereotyping and Discrimination,” in *SAGE Handbook of Prejudice, Stereotyping and Discrimination*, ed. John F. Dovidio et al. (London: SAGE Publications Ltd, 2010), 3-28.

<sup>15</sup> Edouard Machery, Luc Faucher, and Daniel Kelly, “On the Alleged Inadequacies of Psychological Explanations of Racism,” *The Monist* 93 (2010), 229.

---

<sup>16</sup> Kenneth Stickers, ““But I’m Not Racist”: Toward a Pragmatic Conception of Racism,” *The Pluralist* 9.3 (2014): 1-17. Clevis Headley, “Philosophical Approaches to Racism: A Critique of the Individualistic Perspective,” *Journal of Social Philosophy* 31.2 (2000): 223-257.

<sup>17</sup> Stickers, “But I’m Not Racist,” 5.

<sup>18</sup> As I show in section 2.1., this second set of actions would count as racist on the input-based model of racism. See also footnote 19.

<sup>19</sup> My output-based model of individual racism as functionally racist behavior departs from established accounts in the literature insofar as it de-emphasizes the question of intentionality. When theorizing individual racism, philosophers tend to adopt an input-based model that focuses on whether the individual’s actions are motivated by race-based ill-will or disregard, or whether these actions express race-based moral wrongs, such as antipathy, inferiorization, and disrespect. For these alternative accounts of individual racism, see Garcia, “Current Conceptions”; Blum, *I’m not a Racist, But...*; and Glasgow, “Racism as Disrespect”.

Following my commitment to a wide-scope conception of racism, I do not deny that it is appropriate in some contexts to describe individual agents (and their actions, beliefs, and desires) as “racist” if they are motivated by race-based ill-will or disregard (e.g. for the purposes of assigning moral responsibility). However, from the perspective of the output-based model, individual racism matters not because of its input (i.e. race-based ill-will or disregard) but because of its output (i.e. reinforcing a system of racial oppression).

<sup>20</sup> Tommie Shelby, “Ideology, Racism, and Critical Social Theory,” *The Philosophical Forum* 34 (Summer 2003): 153-188.

<sup>21</sup> Sally Haslanger, “What is a Social Practice?,” *Royal Institute of Philosophy Supplement* 82 (2018), 240.

<sup>22</sup> *Ibid.*, 238-9.

<sup>23</sup> Tommie Shelby, *Dark Ghettos: Injustice, Dissent, and Reform* (Cambridge, MA: Belknap Press, 2016), 25.

<sup>24</sup> Rahel Jaeggi, “Was ist eine (gute) Institution?” in Axel Honneth and Rainer Forst (eds.), *Sozialphilosophie Und Kritik* (Berlin: Suhrkamp, 2009).

<sup>25</sup> This definition is inspired by Shelby’s account of intrinsic institutional racism and Garcia’s volitional account of institutional racism. See Shelby, *Dark Ghettos*; and Garcia, “Current Conceptions”.

<sup>26</sup> Garcia, “Heart of Racism,” 6, 33.

<sup>27</sup> *Ibid.*, 11.



- 
- <sup>28</sup> Shelby, *Dark Ghettos*, 24-26. Here I repurpose Shelby's theory of intrinsic institutional racism in order to shed light on the input-based model of institutional racism.
- <sup>29</sup> Robert Gooding-Williams, "Revisiting the Ferguson Report: Antiblack Concepts and the Practice of Policing." *Critical Inquiry* 47 (2021): S132-7.
- <sup>30</sup> César Cabezas, "Is Affirmative Action Racist? Reflections toward a Theory of Institutional Racism," *Journal of Social Philosophy* 54 (2023): 218-35. For alternative accounts of institutional racism that follow the input-based model, see Garcia, "Current Conceptions"; Blum, *I'm not a Racist, But...*; and Glasgow, "Racism as Disrespect". For accounts of institutional racism friendly to the output-based model, see Shelby, *Dark Ghettos*; and Headley, "Philosophical Approaches to Racism."
- <sup>31</sup> Gertrude Ezorsky, *Racism and Justice: The Case for Affirmative Action* (Ithaca: Cornell University Press, 1991), 14-18.
- <sup>32</sup> On the role of segregation in perpetuating racial inequality, see Anderson, *The Imperative of Integration*.
- <sup>33</sup> There is ample evidence of this claim in empirical sociology. Two highly influential research projects are Eduardo Bonilla Silva's work on color-blind racist ideology and Lawrence Bobo's work on laissez-faire racist ideology. See Eduardo Bonilla-Silva, *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America*, Fourth Edition (Lanham: Rowman & Littlefield Publishers, 2013); and Bobo et al, "Laissez-Faire Racism".
- <sup>34</sup> Bonilla-Silva, *Racism without Racists*. See also Anderson, *The Imperative of Integration*, chapter 8.
- <sup>35</sup> George M. Fredrickson, *Racism: A Short History* (Princeton, N.J.: Princeton University Press, 2002), 158ff.
- <sup>36</sup> Magnus Hirschfeld, *Racism*, trans. Eden Paul and Cedar Paul (London : V. Gollancz, 1938).
- <sup>37</sup> Miles, *Racism*, 2<sup>nd</sup> edition, 60.
- <sup>38</sup> Of course, the existence and study of ideologies and systems of racial oppression predate the coining of the term "racism" in the early 20<sup>th</sup> century. However, to the best of my knowledge, pre-20<sup>th</sup> century scholars did not use the term "racism" to describe the ideologies, actions, institutions, and systems of racial oppression that were the subject of their studies. As a widely-used term for describing racially oppressive phenomena, "racism" is of relatively recent origin.
- <sup>39</sup> Fredrickson, *Racism*, 167. Fredrickson uses the terms prejudice and discrimination to refer to the narrow-scope conception.
- <sup>40</sup> These criteria for assessing conceptual inflation are specific to "racism" and its cognates. I thank an anonymous editor for encouraging me to clarify this point.
- <sup>41</sup> Garcia, "Current Conceptions of Racism," 6. Blum, "What Do Accounts of "Racism" Do?," 57.
- <sup>42</sup> Blum, *I'm Not a Racist, But...*, chapter 3.

---

<sup>43</sup> Ibid., 62.

<sup>44</sup> Blum, *I'm not a Racist, But...*, 9.

<sup>45</sup> Ibid., 22.

<sup>46</sup> Ibid., 24-5.

<sup>47</sup> César Cabezas, "Racism: A Moral or Explanatory Concept?," *Ethical Theory and Moral Practice* 24 (2021): 651–59.

<sup>48</sup> Haslanger, "Racism, Ideology, and Social Movements," 2. Shelby, "Racism, Moralism, and Social Criticism," 63.

<sup>49</sup> On the distinction between personal and political morality, see Shelby, "Racism, Moralism, and Social Criticism."

<sup>50</sup> Garcia, "The Heart of Racism," 10.

<sup>51</sup> Alberto Urquidez, *(Re-)Defining Racism: A Philosophical Analysis* (Cham: Palgrave Macmillan, 2020).

<sup>52</sup> Ibid., 23.

<sup>53</sup> On how the practice of seniority has a disproportionate negative impact on Black workers, which in turn reinforces negative stereotypes of Blacks as lazy and lacking personal motivation, see Ezorsky, *Racism & Justice*.

<sup>54</sup> This is not a matter of ideal theorizing. In "‘Extremely Racist’ and ‘Incredibly Sexist’", Liao and Hansen offer empirical evidence that the term "racist" in ordinary use is applied to varying degrees of racial wrongs and is compatible with nuanced forms of moral condemnation.

<sup>55</sup> Garcia, "Current Conceptions," 26.

<sup>56</sup> Here I heed Frye's call to adopt a macroscopic view in order to recognize how seemingly innocuous individual acts can be part of a larger system of oppression. See Marilyn Frye, "Oppression" in *The Politics of Reality: Essays in Feminist Theory* (Freedom: The Crossing Press).

<sup>57</sup> An anonymous referee encouraged me to discuss this example.

<sup>58</sup> The Associated Press, "More Than 9,000 Anti-Asian Incidents Have Been Reported Since The Pandemic Began," *NPR*, August 12, 2021, <https://www.npr.org/2021/08/12/1027236499/anti-asian-hate-crimes-assaults-pandemic-incidents-aapi>.

<sup>59</sup> Department of Justice. <https://www.justice.gov/hatecrimes/learn-about-hate-crimes>.

<sup>60</sup> I do not share this intuition, but it seems to be consistently held among defenders of the input-based model.

<sup>61</sup> Garcia, "Current Conceptions," 24.

<sup>62</sup> I respond to Garcia's reductio in detail in Cabezas, "Is Affirmative Action Racist?"

<sup>63</sup> Blum, *I'm not a Racist, But...*, viii.

- 
- <sup>64</sup> Blum, *I'm Not A Racist, But...*, 8.
- <sup>65</sup> Blum, "What Do Accounts of "Racism" Do?," 65.
- <sup>66</sup> Blum, *I'm not a Racist, But...*, 2.
- <sup>67</sup> Benjamin Mitchell-Yellin, "A View of Racism: 2016 and America's Original Sin," *Journal of Ethics and Social Philosophy* 1 (2018): 53–72.
- <sup>68</sup> Blum, *I'm not a Racist, But...*, 35.
- <sup>69</sup> For a critique of Blum along these lines, see Hardimon, "Should We Narrow the Scope".
- <sup>70</sup> Kwame Ture and Charles V. Hamilton, *Black Power: The Politics of Liberation* (New York: Vintage, 1992).
- <sup>71</sup> Ture and Hamilton, *Black Power*, chapter 3.
- <sup>72</sup> Miles, *Racism, 2<sup>nd</sup> Edition*, 16.
- <sup>73</sup> *Ibid.*, 73.
- <sup>74</sup> *Ibid.*, 103-109.
- <sup>75</sup> In Miles' framework, a practice (or institution) is racially exclusionary when it contributes to the reproduction of race-based inequality in a certain domain, such as the labor market. See *Ibid.*, 110-114.
- <sup>76</sup> *Ibid.*, 110-112.
- <sup>77</sup> *Ibid.*, 79.
- <sup>78</sup> *Ibid.*, 80.
- <sup>79</sup> *Ibid.*, 79-80.
- <sup>80</sup> Sally Haslanger, "Oppressions: Racial and Other," in *Racism in Mind*, ed. Michael Levine and Tamas Pataki (Ithaca: Cornell University Press, 2004).
- <sup>81</sup> Of course, there are non-white women, but "women" as a social group refers to a gender, rather than a racialized group.
- <sup>82</sup> Even though Jewish people have for the most part been assimilated into whiteness, the recent rise of overt antisemitism is a reminder of their fragile position within whiteness.
- <sup>83</sup> In fact, it may turn out that women of color are excluded from white social networks to a greater extent than men of color due to the compound effects of racism and sexism.
- <sup>84</sup> Cecilia L. Ridgeway, *Framed by Gender: How Gender Inequality Persists in the Modern World* (Oxford: Oxford University Press, 2011).
- <sup>85</sup> Miles, *Racism, 2<sup>nd</sup> Edition*, 71.

---

<sup>86</sup> Ibid., 79.

<sup>87</sup> Ibid., 16.

<sup>88</sup> I thank an anonymous referee for raising this question.

<sup>89</sup> To return to Frye, applying the label “racist” to the institution encourages us to adopt the macroscopic view of oppression. As Frye argues in “Oppression”, this perspective contributes to the causal-explanatory project by allowing us to recognize the institution as part of a larger network of systematically related forces and barriers that disadvantage non-whites, i.e. as part of a racist system.