

# Putting Racism Back in the Head

Jordan Scott

Rutgers University, New Brunswick, New Jersey, USA

**Correspondence:** Jordan Scott ([jss455@philosophy.rutgers.edu](mailto:jss455@philosophy.rutgers.edu))

**Received:** 27 February 2025 | **Revised:** 1 April 2025 | **Accepted:** 3 April 2025

**Keywords:** analysis of racism | bias | individual racism | philosophy of race | prejudice

## ABSTRACT

Personal racism used to be widely considered a kind of cognitive defect, with racists being people with biased, irrational racial attitudes. This kind of epistemic “racism-in-the-head” view has fallen largely out of favor in recent decades. Few philosophers have defended it, with many turning toward moral or socio-political rival accounts. This paper offers a robust defense of the epistemic view. It advances a new, broader version, claiming: Personal racism is determined solely by the number and significance of one’s biasing attitudes about racial groups. This bias centered view of racism avoids key problems with previous epistemic views. It also improves upon dominant moral and socio-political views in accounting for a broader range of racist attitudes.

## 1 | Introduction

It used to be common to think that racism, or at least that personal kind of racism that determines whether someone is a racist, was largely an epistemic matter; a “thing in the head”, as it were. While epistemic views of racism differed on what exactly the distinctly racist-making failing was, they agreed that it was some kind of distortion from reality on matters of race. Most often, they argued that racism was rooted in false racial beliefs, or an irrational resistance to evidence.<sup>1</sup> This guiding thought has stuck around, with many still remarking on how orthodox it is to think of racists as inherently epistemically compromised.<sup>2</sup> However, such views have rarely received robust defense in the literature. They have had no shortage of attackers. Fanon, for example, pulls no punches when he says:

The habit of considering racism as a mental quirk, as a psychological flaw, must be abandoned ([1], page 77)

The epistemic views advanced so far hold one (or both) of the following claims: First, that in order to be racist, someone must have a kind of racist belief. Second, that the specific defect in our racist attitudes is a matter of epistemic irrationality (typically, being formed/maintained against the evidence). Both claims

have been considered implausible and overly restrictive, feeding a general suspicion of the epistemic view of racism.

The shift away from epistemic views happened in two directions. Some argued we should see racism, not as something in the head, but as something in the heart [2]; the kind of thing to be found in an agent’s willful states and moral character. Others argued it was neither, but something out in the world; broad, complex socio-political structures, world-systems, and the like [3, 4].

Yet these rival clusters, moral on one side, socio-political on the other, face serious problems of their own. As such, we have seen a hybrid turn, with new analyses stitching together conditions for racism from across different camps. With political views we saw the trend toward ideology accounts [5, 6], which bring back epistemic conditions with a focus on false beliefs, and with “prejudice plus power” views [7, 8]. With moral views, we see it with Blum [9] and Glasgow [10] who, compared to previous moral views, place a greater focus on antipathic, inferiorising, or disrespectful beliefs about racial groups. This hybrid turn, bolting together certain epistemic, moral, and political conditions, leaves accounts of racism increasingly narrow. But what this trend does suggest is that the old “racism in the head” accounts

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Philosophy & Public Affairs* published by Wiley Periodicals LLC.

were onto something: That racism might be better seen as a matter of distortion from truth than a matter of good and evil or justice and injustice.

The epistemic view of racism, increasingly scorned of late but still regularly taken to be a kind of naïve/default view, deserves a proper defense. I offer a new, broader version of the view, proposing that:

Personal racism is determined solely by the number and significance of one's biasing attitudes about racial groups.

On this view, the distinctive feature of racist attitudes is that they act as biases, distorting the way we think, feel, and thus act with regards to members of racialised groups.<sup>3</sup> This view avoids the failings of previous accounts in the broadly epistemic camp (and also compares favorably to moral and socio-political rivals). It allows a broader range of attitudes than mere beliefs to be racist-making. And, rather than focus on the aetiology of the attitude (e.g., whether it was formed rationally), my account focuses on what the racist attitude does.

## 2 | Racist People and Racist Attitudes

### 2.1 | Preliminaries: Personal Racism

This paper takes up the question of how to analyze personal racism, the quality that makes people “racists” or “more/less racist.” The view I’ll offer is meant as a general account of personal racism. There will be specific varieties that may come with extra conditions. Anti-Black racism, for example, will be a subcategory that requires Black people as its target and likely some extra condition about negative charge.

There are other things than people we sometimes call racist, and the conditions seem to vary with the type of racist entity. To tell if a person is racist, we typically look to their attitudes.<sup>4</sup> With an algorithm or functional object, we look to its outcomes. With symbols, we look at their history and associations. Some entities seem to have multiple pathways to racism. Institutions can be racist through discriminatory procedures and policies or by having a large number of racist persons within them.<sup>5</sup>

Some doubt all racism can be captured in a unified analysis,<sup>6</sup> which would make it worth theorizing about different domains separately. If taken in this spirit, my account would only disagree with, say, sociopolitical analyses of racism inasmuch as they also claim to determine the conditions of personal racism. For those who do want a unifying analysis, getting the best theory of personal racism is still crucial. Many unified accounts treat personal racism as primary, with other entities becoming racist through their relationship to racist persons. Consider the “infection model” championed by Garcia [2, 11] and Peebles [12], where personal racism “infects” things like racist institutions. Though these debates are beyond the scope of this paper, in the concluding remarks I’ll highlight reasons to think that my bias-centered view can move us closer to a unified account of racism.

While the default tendency has been to seek descriptive accounts of racism, some recent accounts are revisionary.<sup>7</sup> Descriptivists seek an account of racism that conforms to our ordinary usage. Revisionists instead aim to reengineer “racism” to apply to some new concept, typically to serve some pragmatic aim such as helping to fight racism or racial injustice [13].

This project is largely descriptive; engaging deeply with revisionist arguments is beyond its scope. When discussing existing accounts of racism, I include those with revisionist aims but examine them through a descriptive lens. That arguments rage so passionately over which things are rightly/wrongly described as racist (within and outside academia) highlights that getting a plausible, descriptive account is a project that means a great deal to many people. While some doubt whether it’s possible to give satisfactory descriptive accounts, this worry primarily concerns the abstract concept “racism.” I agree with Zack [14] that the same arguments do not apply as well to the quality “racist”, which we take to be instantiated in the world. There are also common worries (which I largely share) about conceptual engineering’s feasibility and effectiveness [15–18], and additional worries about “topic-changing” in conceptual engineering projects. There’s much debate over how pressing these worries are [16, 19, 20], but they are nonetheless pervasive concerns.

Of course, even the revisionist cannot escape their descriptive duties entirely. An effective revisionist definition must be one that many people can be persuaded to adopt. Revised concepts that stray too far from the current usage will be unlikely to be taken up widely enough to fulfill their pragmatic aims. So, even revisionists benefit from understanding the most plausible descriptive analyses.

### 2.2 | Actions and Attitudes

Personal racism is typically thought to require racist attitudes; the kinds of attitude that can make someone racist, or more racist. I shall sometimes refer to these as racist-making attitudes. I’ll say more about what these are in Section 3, but typical offerings include beliefs, emotions, and desires. Racist attitudes have two key features. First, they must be in some way about race. Second, they must be in some way criticizable.

There are two other kinds of attitude we sometimes call “racist” that must be distinguished from racist-making attitudes: ones which just manifest racism, and racist tells. Attitudes which manifest racism have been brought about by one’s racism, but needn’t be racist-making themselves. Someone might, through the racist stereotypes they hold about Black people, come to view a particular Black man as stupid. If we heard him expressing this view, we might angrily reply: “That’s just racist!” What we mean is that the attitude comes from (or manifests) his racism, not that it’s inherently racist-making. One could come to think a particular person is stupid for all kinds of reasons that wouldn’t be racist.

Racist attitudes should also be distinguished from attitudes that are merely strong evidence of racism. I call these racist tells. If we could only detect racist attitudes when directly expressed, we’d be in trouble. Instead, we must make assumptions based

on racist tells, like certain actions or uses of language that are typical of racists.

But must personal racism require attitudes? One might follow Philips [21], who sees racism as primarily concerning actions. However, even he thinks the actions must be backed up with some racially directed intention to harm, or disregard for foreseeable harm, and so this view also requires racist attitudes.

To further highlight the necessity of racist attitudes, imagine a nervous cyclist with poor balance who regularly swerves onto the pavement and collides with pedestrians while giving cars a wide berth. The cyclist is White, but lives in a majority Black neighborhood, so everyone he's hit has been Black. He has no racist attitudes; were he in a White neighborhood he'd hit the same number of White people. The cyclist may have some blameworthy general attitude, for example, recklessness. But clearly, if nothing about his recklessness is sensitive to the race of the people around him, it doesn't make him racist.<sup>8</sup> Consider the difference if the cyclist were intentionally targeting Black pedestrians, or even if some underlying racial attitude led him to be less cautious in the Black neighborhood than he would be in a White one.

There is a further question of how "about race" racist attitudes need to be. It arises for almost any theory of racism, and may be impossible to answer fully without agreement on what, if anything, races are. Plausibly, sometimes the attitude in question can be wilfully ignoring race, for example, with colourblind racism. There might also be attitudes that are about race through proxies, latching onto physical features, cultural or social practices, and so on. Someone who says: "I don't have anything against Indians, just people who eat strong-smelling curry, worship too many gods, and speak with an accent," is not someone we should let off the hook.

Of course, this may be a case of strong racist tells. The natural reaction, when confronted with such a man, would be to call him a liar. If someone really had three completely unconnected judgments about food, religion, and accent that neither originated from, led to, or were accompanied by any racist attitude, then perhaps they would not be racist, even though they'd share a suspicious degree of resemblance with racists (and could still be criticisable for their non-racial prejudices). If we decouple the attitude from race completely, intuitions change. Imagine the cyclist was deliberately targeting the pedestrians, but doing so on the basis of some other shared feature than their blackness. Perhaps each was carrying a Starbucks coffee and the cyclist owns a small artisanal coffee shop verging on bankruptcy. There may be some prejudicial "ist" it would be fitting to label the cyclist with, but it wouldn't be "racist."

So, being racist requires having racial attitudes. Might actions also be necessary? This seems implausible. A person who hates Black people is not exonerated because they haven't ever acted on this (e.g., because Black people knew to avoid them).<sup>9</sup> Any descriptive definition of "racist" should allow that a person can be racist through attitudes alone.

While people become racist through holding racist attitudes, this needn't mean any racist attitude is enough to make someone

a racist. Empirical data suggests we treat personal racism as a gradable property [22], and being "a racist" requires meeting a threshold of racism. After all, it's perfectly coherent to say of someone: "Sure, they're a little bit racist, but they're not a racist."

So, for a person to be a racist, their racist attitudes must pass some threshold of significance. I see two clear ways in which this can happen: number and severity. One can be a racist by having sufficiently many racist attitudes even if none would be enough individually. One might believe a single ideological, harmful stereotype about a group without becoming a racist, but if they believe enough of them, attaining a completely warped view of that group, they might thereby become a racist. A person could also become racist through having sufficiently severe racist attitudes, even if they have very few of them. A single, visceral hatred for a particular racial group is clearly sufficient. Blum [9] has noted similar things, and talks about the "moral asymmetry" of racism. He offers an intuitive framework of several conditions that can affect the severity of racism, including historical legacy, positional inferiority, patterns and prevalence, and contribution to racial injustice. As I'll highlight in Section 3.2, most accounts will need to appeal to some kind of significance threshold determining when a person becomes a "full-blown racist."

### 3 | Racism and Bias

#### 3.1 | Racially Biasing Attitudes

In the previous section, I argued that personal racism is determined by the number and significance of one's racist attitudes. Of course, most accounts of personal racism are attitudinal. The major divide is over the question "what is the particular, necessary failing that makes a racial attitude racist?"<sup>10</sup> Broadly, the views in the literature can be divided into three: Epistemic views think that the failing is a deviation from truth or rationality. Moral views think racist attitudes are immoral to hold. Socio-political views think the failing comes from a relation to broad political entities for example, ideologies or oppressive social systems. As I highlighted in the introduction, the epistemic view has largely fallen out of favor. In this section, I'll outline and defend a new, bias-centered version of the epistemic view, showing how it avoids the problems that turned people away from previous versions.

At a minimum, bias involves deviation from a norm.<sup>11</sup> Following Kelly [23], we can distinguish two families of norms that people can be biased relative to: the norms that govern our behavior and the norms that govern our mental lives (I'll call these practical norms and cognitive norms<sup>12</sup>). For example, a football referee who is being blackmailed into ignoring the fouls of one team would be exhibiting a practical bias, but not a cognitive bias (assuming they're fully aware of what they're doing). An alternative referee whose strident patriotism made them irrationally disbelieve their country's team was committing fouls would be exhibiting a cognitive bias.

We think of "biased people" as made so by their various "biases." While there can be racial biases that are cognitive or practical, the ones that become racist biases (i.e., the biasing attitudes that

can make people racist) must be cognitive biases. To illustrate, consider someone pretending to be racist because they have a dangerous, racist family whom they fear. In shunning and avoiding people of a certain race, their behavioral outcomes may be just as biased as a real racist. But this need not make them racist, whereas someone who behaves in the same way sincerely (motivated by genuine prejudice) would be.

Like many, I view biases (cognitive ones in particular) as attitudes; representational mental constructs which perform a biasing role in that they are disposed to cause us to systematically deviate from our genuine norms.<sup>13</sup> Certainly, with so-called “explicit biases”, these are taken to be representational attitudes: beliefs, desires, and the like. Things get a little more contentious with implicit biases. Much debate concerns whether these attitudes, if it makes sense to talk about them as a unified set,<sup>14</sup> are beliefs (or other representational attitudes). They are typically taken to be at least belief-like in many respects: They’re at least partly sensitive to evidence, and they’re used in inference.<sup>15</sup> That said, some think that there can be non-representational implicit biases. These could include “pure associations” [24], or “truly-implicit-biases” [25].

I’m certainly open to there being biasing states and constructs that are non-representational. However, the racist-making attitudes must be representational. After all, as I’ve argued, personal racism requires racial attitudes: attitudes which are about race. That certain biases are ruled out from being able to make people into racists is a consequence of most dominant philosophical theories of racism (as Levy [26] argues persuasively). That said, my view doesn’t need to take any controversial positions on the metaphysics of these attitudes or the nature of representation. Even on Johnson’s [25] functional picture, which allows for “truly-implicit-biases” these biases still encode stereotype-like content. However, this content is merely an emergent feature rather than an intrinsic one. So, a bias might become “about race” only in this emergent sense. Such biases would still be “quasi-representational” which is enough for my view (see Section 3.1.2).

On the view I’m defending, racist attitudes are a kind of racially biasing attitude. More specifically:

A racist attitude is a partly cognitive attitude about racial groups, which plays a truth-distorting role in forming other representational attitudes about racial groups.

This definition involves two key components: That racist attitudes need only be partly cognitive racial attitudes, and that racist attitudes must be “truth-distorting” with respect to race. I’ll motivate each in turn, doing so largely by contrast as I show how each part represents a much-needed departure from previous epistemic views.

### 3.1.1 | Partly Cognitive

Viewing racist attitudes as representational and at least partly cognitive is hardly novel. Epistemic theories of racism (and even

certain non-epistemic theories of racism) are also committed to this.<sup>16</sup> However, previous epistemic views have taken racist attitudes to be a kind of defective belief. Appiah [27, 28] and Banton [29] view them as evidence-resistant beliefs in the claim that there are heritable racial essences that play a role in determining the moral worth of different racial groups. Shelby [6] views them as the false beliefs that form and reinforce racist ideological belief systems.<sup>17</sup> But belief-centered views are often taken to be too restrictive, ruling out far too many attitudes.

The problems come from both directions. On one hand, some attitudes more minimal than full-blown belief can be racist-making attitudes. Implicit biases, for example, are typically taken to fall short of outright belief, perhaps being a kind of patchy endorsement or minimally-evidence-sensitive state [30]. Similarly, we might think “seemings” can be racist. Someone who sincerely asserts: “I don’t think I’d say I believe Black people are violent, but they do kind of seem to be” should not get off the hook.<sup>18</sup> Credences may not be beliefs, but are representational and can certainly be racist. Someone with credence 0.7 in the claim “Whites are the superior race” would still be a White-supremacist even if they didn’t outright believe it. Additionally, there are various emotional attitudes (presumably only object-directed emotions) that, though part cognitive, aren’t beliefs. Someone with a suspicion of Black people, or a fear of them, may not have an explicit belief like “Black people are dangerous,” but nonetheless seem to be mentally representing them as dangerous (See D’Arms and Jacobson [31] for persuasive arguments to this effect regarding object-directed emotions). And of course, on top of all this, there is the problem of desires and volitional states, which Garcia [2, 11, 32, 33] has long argued are paradigm racist attitudes, but can’t be fully explained in terms of belief.

On the other hand, attitudes more complex than particular beliefs seem to be racist attitudes, too. Racist attitudes might involve beliefs, but where none of the beliefs are themselves racist. Consider a set of beliefs like the following: “Group X tend to have some phenotypical characteristic C”, “People with C tend to have property Y”, and “Property Y is objectionable.” Taken together, we could get a racist attitude, but none of the individual beliefs is a racist attitude.<sup>19</sup> We can also stipulate that the person has never directly considered the conjunction of all of these individual beliefs, to avoid any worries about the real racist attitude just being a long, conjunctive belief.<sup>20</sup> What makes them racist is that they have a more complex attitude involving these beliefs and their interplay. Peebles [12] has recently offered a somewhat similar argument against views that focus on beliefs. He considers two people who share the same set of beliefs, but who differ on the way in which they emphasize certain beliefs. He argues that this emphasizing basis could make one racist and the other not, and so the racism of the complex attitude wouldn’t reduce to the beliefs alone.

So, a sole focus on beliefs is a bad-making feature for a theory of racism; it risks a pernicious kind of psychological monism (to borrow a term from [34]). While one might quibble over certain states mentioned above (e.g., perhaps “seemings” really are beliefs), there are enough plausible racist-making, non-belief states that the previous epistemic views seem too narrow. My account, by contrast, only requires that racist attitudes are “partly cognitive” attitudes about racial groups. This requires that they have

some cognitive, representational element (an anti-Black racist attitude must be about Black people). But it does not require that racist attitudes be beliefs. Racial hatred, seemings, credences, implicit biases, desires, and more can be racist attitudes on my view.

### 3.1.2 | Truth-Distortion

The second problem was that previous epistemic accounts often treated irrationality as the necessary failing of a racist attitude. Arthur's [35] epistemic view explicitly states that racist attitudes must be irrational to hold. Appiah [27] claims that racist beliefs are "evidence-resistant" (i.e., maintained irrationally).

But racist attitudes need not be formed/maintained irrationally. A person might become racist through being stuck in a bad-evidential environment, bombarded with evidence that justifies racist beliefs. Consider an insular community, where one's family and authority figures are racist, and counter-evidence is made near-impossible to come by. This could easily result in racist beliefs being rationally acquired (perhaps even required). In fact, Begby [36] spends large portions of his book going through the various ways that prejudicial beliefs can be formed and maintained while adhering to the standards of epistemic rationality/justification (internalist or externalist). One particularly noteworthy point is that, once prejudicial beliefs are acquired, it can be very difficult to gain enough counter-evidence to make us rationally drop them. For example, as racist stereotypes often take a generic form, merely observing counter-instances (group members who don't adhere to the stereotype) does not give us the necessary epistemic reasons to drop the stereotype.

Furthermore, there's a significant literature in the ethics of belief based around cases of people with purportedly racist yet rational beliefs. These typically involve people who believe in racial stereotypes (sometimes then generalizing from those to individuals) but hold these beliefs on the basis of statistical evidence [37–40]. The particular cases (e.g., waiters who believe a Black customer will tip badly, but because they're in possession of statistics showing a robust correlation) are controversial both in and outside of the ethics of belief literature,<sup>21</sup> but it further highlights that the notion there can be racist attitudes with rational aetiologies is widespread.

So, an irrationality requirement is implausible.<sup>22</sup> Nonetheless, I still think that the kind of failing present in racist attitudes is a "broadly" epistemic one; a matter of distortion from norms of truth or accurate representation. On Shelby's account, racism has an epistemic dimension as the ideological beliefs that make people racist function to mislead, playing a truth-distorting role in the service of social injustice [6]. While I don't think racist attitudes need to be false beliefs, an important insight of this view is that it focuses on what racist attitudes do, rather than how they come about. I take my view to build on this insight. Racist attitudes are disposed to distort people's racial reasoning in significant ways.

Two more features of the bias-view are worth highlighting. First, just as it's not only beliefs that can be racist attitudes, it's also not only beliefs that racist attitudes need act upon. Racial

attitudes might distort our beliefs, desire-formation (and thus the intentional-actions carried out on their basis), our object-directed emotions, and even our perceptions. Second, it's worth emphasizing that the biasing effect of racist attitudes, their truth-distorting function, must cause a racial bias. Racial attitudes might bias us with regards to other matters without thereby becoming racist. Beliefs about whether "there exist races" could bias us with respect to ontology, distorting our judgments about what things really exist, without making someone racist. The bias must create a differential cognitive treatment for people of different racial/(racialised) groups. If some attitude causes a person to fear people from every race, that is, all people, then this is not a racial bias, just good old-fashioned paranoia.

The fact that our racial biases can distort states like desires/emotions might raise a worry: How, then, is the bias a matter of distortion from truth or accuracy?

The assumption I'm making here (a rather common one), is that all of these states have fittingness conditions that can be represented in terms of accuracy/inaccuracy (or even truth/falsity).<sup>23</sup> This is clearly true with regards to beliefs, perceptions, credences and the like. But it's a little less obvious with regard to things like desires and emotions. Nonetheless, it's common to think that these states are representational on some level: That the person who experiences fear of something is having it in some important sense represented to them as dangerous. This makes sense of a lot of our talk about the rationality of emotions and desires. If we fear something patently un-dangerous, this is an irrational fear. The thing is being represented falsely as dangerous.

In this regard, something that is disposed to bias our object-directed emotions, for example, disposed to make us fear unfear-worthy things, can be seen as falling afoul of norms of accurate representation (see [31] for a persuasive defense of this view). The same goes for anger. Srinivasan [41] thinks that a necessary condition of being angry at something is that the anger's object is presented as a moral violation and highlights how racism in society can cause us to mislabel the anger of people from marginalized groups as inapt. Cherry [42] agrees, claiming that judgments about injustice, wrongdoing, or offense are an essential element of anger. A person who gets inappropriately angry (say flying into a rage at some minor irritation) is getting something wrong. And so, a bias disposed to make people experience inappropriate anger toward people of a certain race can count as a racist bias.

Similarly, it's common to think of desires as representational. Various different theories of desire agree to this much, though they disagree on what exactly it is that desires represent. Guise of the good theorists [43–45] hold that desires represent their objects as being "good" or "desire-worthy." Others think desires represent us as having "reasons for" their objects [46, 47]. In fact, some even think that desire is just a species of normative belief [48].

Now, none of this is to say that the agent's all-things-considered position must align with the desire/emotion. One could fear a spider while knowing full well that it's not dangerous. Someone allergic to chocolate could still desire a

Toblerone. So, our different faculties can have conflicting representations. In fact, even on views where all these attitudes reduce to beliefs, we could account for this as a matter of fragmentation [49].

Also, while I'm sympathetic to these common views on which desires, emotions, and so on are intrinsically representational attitudes,<sup>24</sup> it's worth stressing that my account doesn't depend on this. To show why, let's briefly turn back to beliefs. While representationalism about belief is the orthodoxy [50], there are scholars who think even beliefs aren't intrinsically representational (e.g., dispositionalists like Schwitzgebel [51]). But most will still want their accounts to capture the representation-like features of beliefs: Its evaluability in terms of truth/accuracy and rationality. Belief, even if not intrinsically representational, will still be quasi-representational (judged "as if it were" representational).<sup>25</sup>

The same goes for other attitudes. Even those moved to reject intrinsic representationalism about emotions/desires and so on, will want to vindicate the ways in which they can be judged in terms of rationality, accuracy, and aptness.<sup>26</sup> Even if they are only quasi-representational attitudes, that is enough for my view, as their fittingness conditions can still be represented in terms of truth or accuracy. As Hazlett [44] points out, a desire can have representational "correctness conditions" in common with an evaluative belief, without this implying that desire simply is an evaluative belief. Questions about the metaphysics of these states come apart from questions about their normativity. So, regardless of whether emotions/desires are intrinsically representational, they can still be judged as apt (/correct/fitting), and can still be distorted by our various biases.

### 3.2 | The Significance of Racist Attitudes

On the view I've been defending, people are made racist through their racist biases, with how racist they are determined by the number and severity of these biases. While one axis of severity is level of truth-distortion, severity is not solely an epistemic matter. Biases, racial and otherwise, can be more severe based on the effects they're disposed to have, as well as the extent of their distortion. Someone can become more racist by having attitudes disposed to lead people further astray, or through attitudes disposed to lead people astray in more impactful ways.

Importantly, what matters is the significance of the attitude itself, not the significance of some person's holding that attitude. Deep hatred of Asians is a significant racist attitude, disposed to have damaging, distorting effects. Mild suspicion of Asians is less significant. While either might be significant enough to make the holder racist, the first clearly makes one more racist than the second. Of course, a particular person's holding of the second could be more significant than their holding of the first. Someone with racial hatred while living in a place without any Asians could have less impact than a person in an area with plenty of Asians who only has the mild suspicion. Nevertheless, the former is still more racist. Similarly, a violent racist does not become less racist in virtue of breaking their leg, even if it renders their racism less impactful.

People can attend to their racist attitudes in various ways, but whether this makes them less racist will depend on how they do it. Someone aware of their disposition to fear people of a certain race might either try to reduce this fear, and thus reduce their racism, or merely try to mask its obviousness to make people of that race less uncomfortable. The latter may not make them less racist, but would still be a good thing to do.

One might worry that, even with the significance condition, my account is too broad: Significance coming from number or severity might seem to generate strange results for people with lots of barely significant biasing attitudes. Imagine someone who gains a lot of beliefs about people of a particular race that are all false, but not by much. Perhaps through studying from a dodgy, outdated textbook, they've gained many false statistical beliefs about Black people. Could they thereby become a full-blown racist?

Not necessarily. If the beliefs only fall slightly short of perfect accuracy, then the kinds of inference we're disposed to draw on their basis will be largely the same as for the accurate beliefs, and so none of the individual attitudes would be significantly distorting. The same goes for beliefs about niche, isolated topics that typically have little bearing on one's racial web of representations; these might have next-to-no distorting effect (as I noted in Section 3.1.2, racist attitudes must distort in significant ways). Of course, we could easily imagine enough small attitudes together adding up to cause a larger racial distortion. In this case, we can always say that the set of attitudes (or some emergent attitude) is the distorting, racist attitude. And this strikes me as exactly the result we want. As I highlighted earlier, we can get racist attitudes from belief sets even when no particular belief is particularly (if at all) racist itself.

A second potential worry about the significance condition is: Is this an ad hoc aspect of the view?

On the contrary, I think that there are strong reasons independent of my view to think that the concept of racism builds in a significance threshold. Many other accounts of racism, implicitly or explicitly, appeal to a kind of significance condition.

On most views, the kinds of failing distinctive of racism are gradable failings.<sup>27</sup> There is considerable variance in the degree to which one can, say, have racial disregard or reinforce a racist system. There'll be cases with levels so minimal that proponents of the relevant theories would deny that one becomes a racist in virtue of it. For example, paying your taxes may play some tiny contributory role in reinforcing racist systems if there are any powerful racists in (or supported by) one's government. Nonetheless, clearly, paying them would not make the taxpayer a racist person. Even if what makes a person racist is reinforcing a racist system, a sensible version of the view will have to impose some significance threshold on when the reinforcement is significant enough. Similarly, an epistemic view that centered on irrationality would face the same problem; having any slightly irrational attitude in some racial proposition shouldn't make one a racist. Let's say the rational credence in some particular proposition in the racist ideology is 0.1. While it would seem very plausible for a person with credence 0.8 in that proposition to be a racist, we

wouldn't want to say the same of the person with 0.2 (or 0.15) credence in it. After all, they'd still be very inclined against the claim.<sup>28</sup>

So, it's implicit in these views that there will be some sort of significance threshold for the characteristic failing to be racist-making. Of course, views rarely spell out the threshold. Nor do they need to. Most evaluative terms are threshold-concepts with vague thresholds, and it would be wrong to hold "racist" to a higher standard. That said, there are some views where a little more is said about the threshold, even if not in these terms. Moral views, in particular, build the threshold into their moral concept. For example, take Garcia's and Peebles' view that the central racist-making attitude is morally vicious disregard. Here "moral vice" is itself acting as the significance threshold. There will surely be levels of disregard, even racial disregard, so minute and insignificant that it wouldn't rise to the level of moral vice, and so one wouldn't be immoral, or racist, in virtue of having it. In fact, as various philosophers have argued, one might even have certain kinds of racial prejudice without full-blown racism ([2, 52], chap. 4).<sup>29</sup>

## 4 | Rival Views

### 4.1 | Moral Accounts

In this section, I highlight longstanding problems for moral and socio-political accounts, and show how the bias-centered view avoids them. However, while they aren't necessary conditions on my view, moral/political factors can still help determine who is and isn't racist. They do this through increasing the significance of one's racist biases.

Moral accounts propose a moral necessary condition on being racist. Garcia [32] thinks racism involves immoral racial disregard. Glasgow [10] identifies it with immoral disrespect; Blum [9], as antipathy or inferiorisation, where both must constitute a significant moral failing. A longstanding issue with moral accounts is that they struggle to account for people being racist through belief alone. Garcia's view has come under particular attack for this (see Shelby [53] and Mills [54]). Even moral accounts that do think beliefs alone can be immoral [10, 55] seem to fall short (as Shelby [6] argues persuasively). For example, Blum allows that there can be beliefs that are racist in their content (e.g., "Group X is morally inferior").<sup>30</sup> However, he does not allow that people can be racist merely in virtue of believing them, especially if they believe it in a certain kind of blameless ignorance and would drop it if they reflected on the belief ([9], 21).

As this highlights, even views where beliefs can be morally criticisable may require that these beliefs are held irrationally, and so run into the same problems with people in bad-evidential environments and purportedly rational racists as previous epistemic views. It's common to think that the only authoritative standards on which beliefs can be judged are evidential [56–59]. Moral theorists typically want "racist" to be a significant moral condemnation.<sup>31</sup> If "racist" provides robust moral criticism of beliefs, it suggests one should have believed otherwise. Under evidentialism, it's harder to make this claim unless the holder also went wrong epistemically.<sup>32</sup> A lack of doxastic voluntarism

also invites the worry: if people cannot control their beliefs, they cannot be held morally responsible for them [60].

This aligns with a broad (though not universal) consensus about a kind of epistemic encroachment on the moral ([36], chap. 10), with moral responsibility significantly (perhaps entirely) influenced by what we can be held to account for knowing [61, 62].<sup>33</sup> So, accounting for racism through belief alone has been a long-standing difficulty for moral theories. Even some of the most recent defenses of the moral approach accept that moral theories can't allow racism through belief alone. For example, Peebles concedes to Shelby that beliefs alone can't be immoral, and thus can't be racist ([12], 2513).

As well as the difficulties with belief-based racism, moral views struggle with cases where racism is not negatively charged. Consider a racist boxing coach who believes that Black men are disposed to violence. However, the coach also finds a propensity for violence to be a deeply admirable trait, in no way negative. Indeed, because of this, he treats any Black teenagers that come to his gym with particular warmth and respect. The coach is intuitively a racist, but their racism is not based in a negatively charged attitude toward Blacks. They don't wish them ill, or fail to have good-will toward them. They don't view Black people with antipathy or as inferior, and they have the utmost respect for them (even if for the wrong reasons).<sup>34</sup> Indeed, they seem racist because they've "gone wrong" with respect to race in a significant way; because Black people are not disposed to violence.<sup>35</sup>

For a milder, yet more common example, consider White people who are overly desperate to make Black friends, and so tend to hold them in high regard. This can seem racist (albeit mildly so), but not because they have a general disregard, or even disrespect, for Black people. They simply have a racial attitude that biases the way they think and feel about Black people. Of course, whether such a bias is significant enough to make someone properly racist is an open question. In that regard, my view can still make sense of the intuitions of people that do not think mild racial preferences are racist attitudes.

So, moral accounts face some serious explanatory pressures. Glasgow's [10] account may be somewhat of an outlier here. It offers a particularly weak moral condition (immoral disrespect), which doesn't require a negatively charged attitude.<sup>36</sup> That said, it probably won't satisfy those with strong moral intuitions, who want racism to be a robust moral violation [2, 9]. I also find Glasgow's use of "immoral disrespect" rather strained. He claims (2009, 83) that actions can be disrespectful (and thus presumably racist) absent any racist attitude, using the example of an ignorant child who, copying their brother, raises their middle finger at someone without knowing what the sign means. But the sense in which this is "disrespectful" strikes me as more a violation of etiquette than morality. Also, applied to racism, it would run afoul of my arguments in Section 2.2 that there is no personal racism without racist attitudes.

Glasgow also concedes to Mills [54] that there can be "benevolent racists", and wants to allow that believing racist things may be immorally racially disrespectful even if the belief was rationally required or the agent was utterly alienated from their belief. However, as he repeatedly acknowledges [10, 63], this requires

taking a highly controversial position on moral responsibility.<sup>37</sup> Of course, this is not a refutation of Glasgow's approach, but my aim is not to refute moral theories. It is to show that they face significant pressures that my epistemic account avoids.

## 4.2 | Socio-Political Accounts

I noted that my view takes some inspiration from the ideology views of Shelby [5, 6, 53] and others. Particularly instructive in these views is that what makes certain beliefs racist is not how they came about, but what they do. However, socio-political accounts, like the ideology view, have further commitments; one's beliefs or other racist attitudes only count as racist-making if they align with certain socio-political structures. Beliefs that don't conform to or support racist ideologies, or that target groups in positions of relative or equal power (on prejudice-plus-power views) are unable to be racist. This essentially imposes a kind of one-way condition on racism.

One consequence of these conditions is that they leave political accounts unable to handle unorthodox racists, or racists removed from socio-political structures, due to the pervasiveness requirement. If White supremacy had never spread beyond a small group of people, they'd still intuitively be racists, and we'd want to say the same about new racist views that could spring up. In these isolated-racist cases, ideological accounts miss the mark by basing the criticisability of the racist on their link to the socio-political environment ([35], 17; [10, 11]). Similarly, it's important to account for people with novel racist attitudes. Someone may have a racist view of a particular group which doesn't conform to the typical stereotypes. For example, someone may believe that the racist stereotypes typical of Indians are actually more apt for Jews, and vice versa, or may have come up with new ones entirely. Socio-political conditions that require conformity with broader social trends and structures will fail to account for many of these kinds of racist attitudes. The bias-centered view avoids these problems, as it requires no link to socio-political structures.<sup>38</sup>

While this would be a bad-making feature for a descriptive view, these more restrictive, one-way socio-political accounts are typically motivated on a pragmatic, rather than descriptive basis; the claim being that they shift focus onto institutional racism or power structures. I said a little in Section 2.1 to motivate my focus on descriptive theories, and it's beyond the scope of this paper to try to refute revisionism. However, it's worth saying a little more about why I think we should be suspicious of these pragmatic, socio-political views. One reason is that the one-way restrictions they employ seem largely pragmatically redundant. I argued earlier in the paper that most or all theories of racism will need to build in a significance condition. As I've argued elsewhere [8], one-way conditions on racism face a redundancy problem: The pragmatic benefit of a one-way condition comes largely from highlighting that certain forms of racial prejudice are more significant. However, if theories of racism must already have significance conditions, theories without one-way conditions already have the resources to do this.

Additionally, there are reasons to be skeptical of whether attempts at reengineering the term racism for pragmatic, group-specific

racial-justice ends are particularly effective. As well as the concerns I noted in Section 2.1, recent work by Darby [64] suggests that so-called "small-tent" and "race-flag-first" approaches to racial justice are generally unlikely to succeed. Those interested in racial-justice should instead, Darby argues, favor "big-tent" policies that can unify people across racial groups. I largely agree with this sentiment and think that one-way theories of racism are paradigm cases of small-tent, race-flag-first approaches to racism. With broader, "big-tent" theories of racism like the bias-centered view, we better avoid backlash or dismissal from majority racial groups and yet can still push the most significant cases to the top of the societal agenda.

## 5 | Concluding Remarks: Beyond Personal Racism

In this paper, I've focused on offering a theory of personal racism. I think there is much to be gained from this approach. After all, some think that there is no unified theory to be had, and the best we can do is give different accounts for different domains of racism [65–67]. Even among those who do want a unified account, many [2, 12] hold that personal racism is primary, with institutions, social structures, and so on. only becoming racist when infected by racist persons.

Of course, many find these positions unsatisfactory and want a truly unified concept of racism: One that offers a single set of racist-making conditions. The same sort of thing that makes the institution racist ought to make the algorithm, or the person, or the system racist, too. However, this seems hard to square with the fact that, with people, we seem to look to their attitudes for racism and largely ignore mere behavioral outcomes (though we use the latter to provide a window to the former). One more theoretical upshot of a bias-centered view is that it centers a property that can be possessed by other things than just people. Institutions can be racially biased, algorithms can be racially biased, as can anything else capable of treating people of one race differently than another.

Furthermore, seeing racism as a matter of bias uniquely allows us to explain the intuitive divide between personal and non-personal kinds of racism. The difference simply follows from our assessment of the relevant norms governing persons as opposed to non-persons. With things like technology, algorithms, social systems, and institutions,<sup>39</sup> we care about their outcomes, what they bring about when carrying out their function. If an algorithm, when used, produces racially biased effects, this is all that need interest us.

But people, on the other hand, are first and foremost thinking, feeling things. And as such, the kind of biases we care about are the norms that govern thought and feeling. The reason we do not judge people as racist by the standards of functional objects is that doing so would come uncomfortably close to treating people as mere means to ends.

## Acknowledgments

Thanks to Derrick Darby, Jorge Garcia, Susanna Schellenberg, Jessie Munton, Alex Guerrero, audiences at Rutgers University, and two anonymous reviewers for a great deal of insightful feedback.

## Conflicts of Interest

The author declares no conflicts of interest.

## Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

## Endnotes

<sup>1</sup> See, for example, ([68], 9; [27, 35, 69]).

<sup>2</sup> For example, in [36, 37].

<sup>3</sup> When I talk about “races” or “racial-groups”, this is not meant to imply a stance on the metaphysics of such groups, nor whether there really are races. It might well be that there are no races, only racialised groups. While I’ll continue to talk of “race” for ease, note that strictly speaking I mean “races or racialised groups.”

<sup>4</sup> At least for what determines their racism. We may well look to behaviors to find evidence of racism.

<sup>5</sup> Or perhaps by being controlled by powerful racists, though this may reduce to the institution’s behavior.

<sup>6</sup> See Blum [65], Headley [66], and a critical discussion in Glasgow [10].

<sup>7</sup> I take the following to be descriptivist: [2, 10, 27, 35, 70]. I take the following to be revisionist: [6, 9, 71, 72].

<sup>8</sup> While actions aren’t needed for personal racism, they might still be needed for racism in other domains, for example, institutional racism [73].

<sup>9</sup> See ([2], 13; [35], 17).

<sup>10</sup> One might wonder: Why not have a disjunctive view where one becomes racist through either an epistemic, moral, or political failing. I’m unaware of any theories that take this path, and see two immediate reasons why. First, methodologically, adding extra disjunctive conditions from other theories in order to accommodate counterexamples seems ad-hoc and risks over-fitting. Second, many views of racism face over-generation worries (take Cabezas’s [74] worry that Shelby’s ideology-reinforcement condition over-generates, for example). Disjunctively chaining together conditions from different theories of racism risks compounding the overgeneration worries of these theories.

<sup>11</sup> See [23, 75].

<sup>12</sup> It’s worth noting that I’m using “cognitive” in a broad sense. Some use it to concern only beliefs.

<sup>13</sup> The term “attitude” is not used in a perfectly uniform way across disciplines. The way I’m using it aligns more with the broader, so-called “philosopher’s sense” (which includes propositional, doxastic attitudes) than the “psychologist’s sense” (they often use “attitude” to mean mere likings/dislikings) [76, 77]. Regarding the “genuine norms” I mean the norms that govern our thoughts, desires, and other representational attitudes; norms of rationality and fittingness.

<sup>14</sup> See [78] for arguments that implicit biases are heterogenous, though I agree with Levy [26] that the case made against heterogeneity by [79] is more convincing.

<sup>15</sup> Mandelbaum [77] provides a compelling case that implicit biases are at least belief-like, being somewhat evidence responsive and used in inference. I’m less convinced that they are themselves beliefs, but minimalist accounts of belief may think that being “somewhat evidence responsive” is sufficient [80].

<sup>16</sup> Even moral, virtue-based accounts of racism like Peebles’ [12] hold that racist attitudes have a doxastic component.

<sup>17</sup> Dror [81] thinks ideological beliefs needn’t be false, but must still have some aetiological flaw like being unjustified or based on misleading evidence.

<sup>18</sup> It’s controversial exactly what seemings are, but many think they aren’t reducible to beliefs [82].

<sup>19</sup> Perhaps property Y is something trivial, and the aversion to it is irrational.

<sup>20</sup> There’d likely be additional problems for an epistemic view that took this line (particularly if racist beliefs must be irrational too). After all, inferring a conjunction from its conjuncts is epistemically blameless, so it may come out impossible to get a racist conjunction from non-racist conjuncts.

<sup>21</sup> For example, it’s not clear these cases do not involve some epistemic fault. They typically involve using statistics to form beliefs about individuals. Some recent work [83, 84] argues this may be epistemically illegitimate. One might go wrong in ignoring better evidence, or be mistaken regarding the modal profile of their generalizations.

<sup>22</sup> That said, irrationality in our racist attitudes might be even more pervasive than it would at first seem. Inasmuch as racial attitudes are formed through our navigating and interpreting the social-world (depending on all sorts of low-level social inferences) there’s plenty of opportunity for errors to creep into the aetiology. Whether this is enough to leave our racial attitudes “epistemically compromised” or irrational in a significant sense is unclear. After all, this will be true of all our social attitudes, and we should be wary of falling into a kind of social skepticism.

<sup>23</sup> People use a variety of different terms for the relevant conditions; “fittingness conditions”, “correctness conditions”, “aptness” to name but a few. They may not be perfectly theoretically neutral, each carrying some meta-normative baggage, but for our purposes there’s rather little difference between them. What’s important is that they are representational conditions, analogous to accuracy conditions on belief.

<sup>24</sup> As Clark [85] aptly notes: “The notion of intrinsic content is not, I think, one of the clearer and more distinct products of mid twentieth-century philosophizing.” Nonetheless, I largely follow his characterization and view intrinsic representation of content as when content is carried in an entirely non-conventional way (not depending on any history of use or social practice). One might think, for example, that emotions do not intrinsically carry evaluative content, and that we merely “treat them as if they did” (giving them content in a conventional, and perhaps socially-determined, sense).

<sup>25</sup> Non-representational theories will have plenty of ways of explaining these quasi-representational standards. For example, they can hold that we judge the normativity of belief in a representational way, even if they think that the metaphysics of belief is a dispositional matter. Alternatively, they can just say that norms which look an awful lot like the representational norms are still entailed by a non-representational picture.

<sup>26</sup> Aptness, of course, is often taken to entail accuracy [86].

<sup>27</sup> And rightly so, given there’s empirical data highlighting we treat racism as a gradable property [22].

<sup>28</sup> And depending on the proposition, the same would plausibly hold for slightly positive, if still broadly agonistic, credences. It would be an implausible stretch to say that the person with 0.51 credence in a racial proposition where the evidence only supports it to 0.47 is a full-blown racist on this basis.

<sup>29</sup> Though this is controversial, and the term “prejudice” is also used in different ways. Some use it in a strict “pre-judging” sense, on which it is a matter of beliefs [2]. Others use it more loosely to mean a kind of negatively charged attitude ([52], chap. 4; [9], chap. 2). Begby [36] offers an account of prejudice in terms of stereotypes, though it’s unclear how faithful this account is to ordinary usage.

<sup>30</sup> Which Shelby [6] argues is in tension with a strong moral requirement.

<sup>31</sup> For example, Urquidez ([72], sec. 1.2.3) claims it is impossible to know what “racism” means without knowing it is used to condemn.

<sup>32</sup> Though certain radical moral-encroachers would disagree [38, 39]. Philosophers willing to accept these radical views in epistemology

would have an easier time defending a moral view from this objection, but they'd still struggle to explain the moral fault in having racist credences.

<sup>33</sup> Alternatively, one might argue people can have a kind of moral responsibility for epistemically blameless racist attitudes even if they can't be morally blameworthy for them ([36], chap. 10). Kelly and Roedder [87] suggest a similar move regarding implicit bias, but highlight its contentiousness and are reluctant to embrace it.

<sup>34</sup> One might wonder if the coach could be disrespecting Black people by falsely homogenizing them. But that doesn't seem sufficient for immoral disrespect. A person who thinks "all Black people have faced overt direct racism" (when presumably at least some haven't) falsely homogenizes, but this doesn't seem disrespectful or racist.

<sup>35</sup> Perhaps there's a kind of moral violation that supervenes on the epistemic one. Harman [88] thinks there's a moral obligation to believe the moral truths relevant to our actions. Of course, as I argued earlier, actions need not be necessary for racism at all.

<sup>36</sup> Though Glasgow may be able to capture these cases, we don't know for sure. he doesn't provide an account of this immoral disrespect, or give us the resources to determine what does/doesn't make something disrespectful.

<sup>37</sup> Ought-implies-can violations are not to be taken lightly in a moral theory. That said, though I'm skeptical of it, the view of context-sensitive-variantism he espouses in [63] is interesting and well worth reading.

<sup>38</sup> That said, there may be other senses in which racism is an inherently social phenomenon. After all, racial attitudes are not raw sensations, but a kind of social attitude; a response to, and way of carving up, the social world. Thanks to an anonymous reviewer for noting this.

<sup>39</sup> Though institutions are an interesting case, as they might become racist either through producing racist outcomes, or through simply containing enough racist persons.

## References

1. F. Fanon, *Black Skin, White Masks* (Grove Press, 1967).
2. J. L. A. Garcia, "The Heart of Racism," *Journal of Social Philosophy* 27, no. 1 (1996): 5–46.
3. E. Bonilla-Silva, "Rethinking Racism: Toward a Structural Interpretation," *American Sociological Review* 62, no. 3 (1997): 465–480, <https://doi.org/10.2307/2657316>.
4. R. Grosfoguel, "What Is Racism?," *Journal of World-Systems Research* 22, no. 1 (2016): 9–15.
5. T. Shelby, "Ideology, Racism, and Critical Social Theory," *Philosophical Forum* 34, no. 2 (2003): 153–188.
6. T. Shelby, "Racism, Moralism, and Social Criticism," *Du Bois Review: Social Science Research on Race* 11, no. 1 (2014): 57–74.
7. J. R. Barndt, "Understanding and Dismantling Racism," in *Understanding & Dismantling Racism: The Twenty-First Century Challenge to White America* (Fortress Press, 2007).
8. J. Scott, "Does Racism Equal Prejudice Plus Power?," *Analysis* 82 (2022): 455–463.
9. L. Blum, *I'm Not a Racist, but: The Moral Quandary of Race* (Cornell University Press, 2002a).
10. J. Glasgow, "Racism as Disrespect," *Ethics* 120, no. 1 (2009): 64–93.
11. J. L. A. Garcia, "Virtue Ethics in Social Theory: Defending a Volitional Analysis of Racism From Tommie Shelby's Challenges," *American Philosophical Quarterly* 60, no. 4 (2023): 329–340.
12. I. S. Peebles, "Toward a Virtue-Based Account of Racism," *Philosophical Studies* 181 (2024): 2499–2523.
13. B. Mitchell-Yellin, "A View of Racism: 2016 and America's Original Sin," *Journal of Ethics and Social Philosophy* 13, no. 1 (2018): 53–72.
14. N. Zack, "The Concept of Racism and the Adjective Racist," *Ethical Theory and Moral Practice* 24, no. 3 (2021): 673–677.
15. J. Andow, "Conceptual Engineering Is Extremely Unlikely to Work. So What?," *Inquiry* 64, no. 1–2 (2021): 212–226.
16. H. Cappelen, *Fixing Language: An Essay on Conceptual Engineering* (Oxford University Press, 2018).
17. E. Fischer, "Conceptual Control: On the Feasibility of Conceptual Engineering," *Inquiry* (2020): 1–29.
18. A. F. Gibbons, "Rational Conceptual Conflict and the Implementation Problem," *Inquiry* 67, no. 9 (2023): 3355–3381.
19. C. Dutilh Novaes, "Carnapian Explication and Ameliorative Analysis: A Systematic Comparison," *Synthese* 197, no. 3 (2020): 1011–1034.
20. T. Sundell, "Changing the Subject," *Canadian Journal of Philosophy* 50, no. 5 (2020): 580–593.
21. M. Philips, "Racist Acts and Racist Humor," *Canadian Journal of Philosophy* 14, no. 1 (1984): 75–96.
22. S. Liao and N. Hansen, "'Extremely Racist' and 'Incredibly Sexist': An Empirical Response to the Charge of Conceptual Inflation," *Journal of the American Philosophical Association* 9, no. 1 (2023): 72–94.
23. T. Kelly, *Bias: A Philosophical Study* (Oxford University Press, 2022).
24. B. A. Nosek, F. L. Smyth, J. J. Hansen, et al., "Pervasiveness and Correlates of Implicit Attitudes and Stereotypes," *European Review of Social Psychology* 18, no. 1 (2007): 36–88.
25. G. M. Johnson, "The Structure of Bias," *Mind* 129, no. 516 (2020): 1193–1236.
26. N. Levy, "Am I a Racist? Implicit Bias and the Ascription of Racism," *Philosophical Quarterly* 67, no. 268 (2016): pqw070, <https://doi.org/10.1093/pqw/pqw070>.
27. K. A. Appiah, "Racisms," in *Anatomy of Racism*, ed. D. Goldberg (University of Minnesota Press, 1990), 3–17.
28. K. A. Appiah, "Understanding Racism," *Inquiry* 67, no. 8 (2024): 2229–2242.
29. M. Banton, "The Concept of Racism," in *Race and Racism*, 1st ed., ed. S. Zubaida (Routledge, 1970), 17–34.
30. N. Levy, "Neither Fish nor Fowl: Implicit Attitudes as Patchy Endorsements," *Noûs* 49, no. 4 (2015): 800–823.
31. J. D'Arms and D. Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," *Philosophy and Phenomenological Research* 61, no. 1 (2000): 65.
32. J. L. A. Garcia, "Current Conceptions of Racism: A Critical Examination of Some Recent Social Philosophy," *Journal of Social Philosophy* 28, no. 2 (1997): 5–42.
33. J. L. A. Garcia, "Philosophical Analysis and the Moral Concept of Racism," *Philosophy & Social Criticism* 25, no. 5 (1999): 1–32.
34. L. Faucher and E. Machery, "Racism: Against Jorge Garcia's Moral and Psychological Monism," *Philosophy of the Social Sciences* 39, no. 1 (2009): 41–62.
35. J. Arthur, *Race, Equality, and the Burdens of History* (Cambridge University Press, 2007).
36. E. Begby, *Prejudice: A Study in Non-Ideal Epistemology* (Oxford University Press, 2021).
37. R. Basu, "The Wrongs of Racist Beliefs," *Philosophical Studies* 176, no. 9 (2019a): 2497–2515.
38. R. Basu, "Radical Moral Encroachment: The Moral Stakes of Racist Beliefs," *Philosophical Issues* 29, no. 1 (2019b): 9–23.

39. R. Basu and M. Schroeder, "Doxastic Wronging," in *Pragmatic Encroachment in Epistemology*, ed. B. Kim and M. McGrath (Routledge, 2019).

40. R. J. Bolinger, "Varieties of Moral Encroachment," *Philosophical Perspectives* 34, no. 1 (2020): 5–26.

41. A. Srinivasan, "The Aptness of Anger," *Journal of Political Philosophy* 26, no. 2 (2018): 123–144.

42. M. Cherry, "Political Anger," *Philosophy Compass* 17, no. 2 (2022): e12811.

43. G. E. M. Anscombe and E. M. Gertrude, *Intention. Intention* (Harvard University Press, 1963).

44. A. Hazlett, "Desire and Goodness," *Philosophy and Phenomenological Research* 105, no. 1 (2022): 160–180.

45. J. Raz, "On the Guise of the Good," in *Desire, Practical Reason, and the Good*, ed. S. Tenenbaum (Oxford University Press, 2010).

46. A. Gregory, "The Guise of Reasons," *American Philosophical Quarterly* 50, no. 1 (2013): 63–72.

47. M. Milona and M. Schroeder, "Desiring Under the Proper Guise," in *Oxford Studies in Metaethics*, ed. R. Shafer-Landau, vol. 14 (Oxford University Press, 2019).

48. A. Gregory, *Desire as Belief: A Study of Desire, Motivation, and Rationality* (Oxford University Press, Incorporated, 2021).

49. A. Egan, "Seeing and Believing: Perception, Belief Formation and the Divided Mind," *Philosophical Studies* 140, no. 1 (2008): 47–63.

50. J. Quilty-Dunn and E. Mandelbaum, "Against Dispositionalism: Belief in Cognitive Science," *Philosophical Studies* 175, no. 9 (2018): 2353–2372.

51. E. Schwitzgebel, "A Dispositional Approach to Attitudes: Thinking Outside of the Belief Box BT - New Essays on Belief: Constitution, Content and Structure," in, ed. N. Nottelmann (Palgrave Macmillan UK, 2013), 75–99.

52. J. A. Corlett, *Race, Racism, and Reparations* (Cornell University Press, 2003).

53. T. Shelby, "Is Racism in the 'Heart'?", *Journal of Social Philosophy* 33, no. 3 (2002): 411–420.

54. C. W. Mills, "'Heart' Attack: A Critique of Jorge Garcia's Volitional Conception of Racism," *Journal of Ethics* 7, no. 1 (2003): 29–62.

55. L. Blum, "Racism: What It Is and What It Isn't," *Studies in Philosophy and Education* 21, no. 3 (2002b): 203–218.

56. W. K. Clifford, *The Ethics of Belief and Other Essays* (Prometheus Books, 1877).

57. R. Feldman and E. Conee, "Evidentialism," in *Evidentialism* (Oxford University Press, 2004).

58. N. Shah, "A New Argument for Evidentialism," *Philosophical Quarterly* 56, no. 225 (2006): 481–498.

59. J. Way, "Two Arguments for Evidentialism," *Philosophical Quarterly* 66, no. 265 (2016): 805–818.

60. E. Begby, "Doxastic Morality: A Moderately Skeptical Perspective," *Philosophical Topics* 46, no. 1 (2018): 155–172.

61. M. Fricker, *The Epistemic Life of Groups* (Oxford University Press, 2016).

62. G. Rosen, "Culpability and Ignorance," *Proceedings of the Aristotelian Society* 103 (2003): 61–84.

63. J. Glasgow, "Alienation and Responsibility," in *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, ed. M. Brownstein and J. Saul (Oxford University Press, 2016).

64. D. Darby, *A Realistic Blacktopia: Why we Must Unite to Fight* (Oxford University Press, 2023).

65. L. Blum, *What do Accounts of 'Racism' Do?* (Cornell University Press, 2018), 56–77.

66. C. Headley, "Philosophical Approaches to Racism: A Critique of the Individualistic Perspective," *Journal of Social Philosophy* 31, no. 2 (2000): 223–257.

67. L. Faucher, "Racism," in *The Routledge Companion to Philosophy of Race*, 1st ed., ed. P. C. Taylor, L. M. Alcoff, and L. Anderson (Routledge, 2018), 405–422.

68. G. W. Allport, *The Nature of Prejudice* (Addison-Wesley Publishing Co, 1954).

69. H. Schuman and J. Harding, "Prejudice and the Norm of Rationality," *Sociometry* 27, no. 3 (1964): 353–371.

70. J. A. Corlett, "Race, Racism, and Reparations," *Journal of Social Philosophy* 36, no. 4 (2005): 568–585.

71. S. Haslanger, "Racism, Ideology, and Social Movements," *Res Philosophica* 94, no. 1 (2017): 1–22.

72. A. G. Urquidez, *(Re-)defining Racism: A Philosophical Analysis* (Palgrave Macmillan, 2020).

73. S. Carmichael and C. V. Hamilton, *Black Power: The Politics of Liberation in America* (Random House, 1967).

74. C. Cabezas, "Is Affirmative Action Racist? Reflections Toward a Theory of Institutional Racism," *Journal of Social Philosophy* 54, no. 2 (2023): 218–235.

75. T. Kelly, "Bias, Norms, Introspection, and the Bias Blind Spot," *Philosophy and Phenomenological Research* 108, no. 1 (2024): 81–105.

76. M. Brownstein, "Implicit Bias," in *Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University, 2025).

77. E. Mandelbaum, "Attitude, Inference, Association: On the Propositional Structure of Implicit Bias," *Noûs* 50, no. 3 (2016): 629–658.

78. J. Holroyd and J. Sweetman, "The Heterogeneity of Implicit Bias," in *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, ed. M. Brownstein and J. Saul (Oxford University Press, 2016), 80–103.

79. A. Madva and M. Brownstein, "Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind," *Noûs* 52, no. 3 (2018): 611–644.

80. G. Helton, "If You Can't Change What You Believe, You Don't Believe It," *Noûs* 54, no. 3 (2020): 501–526.

81. L. Dror, "The Epistemic Defects of Ideology and False Consciousness," *Journal of Philosophy* (forthcoming).

82. A. Cullison, "What Are Seemings?", *Ratio* 23, no. 3 (2010): 260–274.

83. G. Gardiner, "Evidentialism and Moral Encroachment," in *Believing in Accordance With the Evidence*, ed. K. McCain (Springer Verlag, 2018).

84. J. Munton, "Beyond Accuracy: Epistemic Flaws With Statistical Generalizations," *Nous-Supplement: Philosophical:Issues* 29, no. 1 (2019): 228–240.

85. A. Clark, "Intrinsic Content, Active Memory and the Extended Mind," *Analysis* 65, no. 1 (2005): 1–11.

86. E. Sosa, *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I* (Oxford University Press, 2007).

87. D. Kelly and E. Roedder, "Racial Cognition and the Ethics of Implicit Bias," *Philosophy Compass* 3, no. 3 (2008): 522–540.

88. E. Harman, "Does Moral Ignorance Exculpate?", *Ratio* 24, no. 4 (2011): 443–468.