

Racism: Against Jorge Garcia's Moral and Psychological Monism¹

Luc Faucher

Philosophy Department, Université du Québec à Montréal

Edouard Machery

Department of History and Philosophy of Science, University of Pittsburgh

(Forthcoming, as of Fall 2008, in *Philosophy of the Social Sciences*)

With very few exceptions (Corlett, 2003; Levine, 2004; Kelly and Roedder, 2008; Kelly et al., in press; Blum, 2004b), philosophers of race have not manifested much interest in what psychology reveals about the way we think about race or about the phenomena of racism. Sure enough, some philosophers have recognized the importance of understanding the mechanisms underlying racial thinking or racism. For example, Naomi Zack writes in *Racial Paradigms* (1999, p. 316; our emphasis):

“Given the history of ideas about race and the present knowledge that race does not have the biological foundation that the lay public continues to think it does, philosophers addressing race at this time would seem to have a professional obligation to think through the implications for related topics of the biological non-existence of race. For instance, [...] *what can we say about the psychological and moral aspects of racism in the absence of the biological foundation?*”

But if philosophers have delved into the project of understanding the moral problems associated with racism, Zack's edict did not convince them to consult the psychological literature about racial cognition.

We suspect that this lack of interest in psychology is in part due to how philosophers of race conceive of their work: For them, the job of a philosopher consists of defining what race and

¹ We are grateful to Dan Kelly, three anonymous reviewers, and to the editors of the special issue of *Philosophy of Social Science* for their comments and suggestions. This article was presented at the Philosophy of Social Science Roundtable in March 2008 and at the conference New Perspectives in Race Theory organized by Ron Sundstrom in April 2008. Earlier versions were presented at the Université Paris V, the Université de Montréal, and the California Round Table on Philosophy and Race in 2007.

racism are, and of understanding the moral issues these concepts raise. There is thus a division of labor between philosophers and psychologists: The former are interested in necessary truths and practice conceptual analysis; the latter are interested in contingent truths and use the empirical methods that are characteristic of science. Within this framework, a philosopher has nothing to expect from psychology. As Jorge Garcia puts it (1996/2003, p. 288): “Psychocultural explanation is unlikely to reveal (logically) necessary truths about the nature of racism.” Anthony Appiah also illustrates this point of view, when he identifies a form of resistance to belief revision as one component of racism (1990, p. 7-8): “My job [...] is not with the psychological or social process by which these forms of ideological resistance operate.” Rather, Appiah sees his job as consisting of defining what racism is, a task which in his view requires only a mundane knowledge of the phenomenon (see also Dummett, 2004).

In this article, we will show that it can be fruitful for philosophers interested in the nature and moral significance of racism to pay more attention to psychology. To clarify, our goal in this article is *not* to argue that philosophers interested in the nature and moral significance of racism *should* pay attention to psychology, because they cannot understand the nature and moral significance of racism without it. Our goal is more modest: We want to *illustrate* how psychology *can* be relevant to the debates about the nature and moral significance of racism. We do this by showing that psychology provides new arguments against Garcia's views about the nature and moral significance of racism.

To accomplish this, we will first focus on the debate between Jorge Garcia and Lawrence Blum concerning the psychological and moral characterization of racism (Section 1). In the remainder of the article, we will show how some scientific studies of racial cognition bring considerable support to Blum's two key arguments against Garcia: Garcia disregards (1) the rich affective texture of racism and (2) the diversity of what makes racial ills morally wrong. In Section 2, we will turn our attention toward the psychological literature on the emotions associated with racial prejudice, and we will show that contrary to what Garcia asserts, racism is linked to a wide variety of emotions; therefore, Garcia's psychological monism is incorrect. Sections 3 and 4 focus on Blum's second critique of Garcia's views. We will review the psychological research on implicit biases, and we will show that these biases can have a dramatic influence on people's behavior (Section 3). We will conclude that the psychological findings about implicit biases

clearly demonstrate the diversity of what makes racial ills morally wrong (Section 4). Therefore, Garcia's moral monism is also incorrect.

1 The Characterization of Racism

As Blum noted in a recent text (2004a), for some, racism has disappeared (at least from the public sphere), whilst for others, racism is everywhere. The controversy surrounding Hergé's illustrated comic storybook *Tintin in the Congo* deftly displays this disagreement: Some see the storybook as an offensive form of racism and demand that the book be sold with a warning²; others see it as a reflection of the bygone colonial era, finding it a source of harmless mild comedy.

Such disagreement is partly due to the fact that the meaning of the word "racism" is no longer clear (if it ever was). As Appiah noted in his critical review of Georges Fredrickson's book *Racism: A Short History* (Appiah, 2002, p. 11; our emphasis),

"Racist laws, pronouncements and actions, we agree, marred much of American history ... But how racist is the United States today? And who are the racists? Try getting agreement about that. One reason for the disagreement is just that people have different hunches about what's going on in other people's head But another reason we cannot agree about racism is conceptual: *people understand the term rather variously.*"

Added to the problem of the meaning of the word "racism" is the problem of the moral characterization of racism, because, as Appiah also noted (2002, p. 11; our emphasis), if (almost) everybody agrees that racism is morally wrong,

"[...] it is far from clear that most of us share a view about why it is wrong. *Is it that racists have hatred or contempt for people of other races? Or that they have irrational beliefs about them? Or that they tend to treat them badly?* Is it, as some opponents of affirmative action allege, that it is wrong to take into account a person's race at all?"

² "Tintin book is crude, racist and must be banned, watchdog says" titled the *Times Online* (July 12, 2007).

Appiah's quotation draws attention to three ways of characterizing what racism is and in what way it is morally wrong: the doxastic model, the behavioral model, and the affective model.³ Although we will focus on the third model in this article, we first present these three models in order to highlight, by way of contrast, the specificities of the affective model.⁴

1.1 The Doxastic Model

According to the proponents of the doxastic model such as, e.g., Appiah, racism essentially involves believing in a particular set of propositions together with specific dispositions with respect to these beliefs (see also Lengbeyer, 2004). To be racist, people must believe that races exist (a belief sometimes called "racialism") and that they are morally significant (either because races are correlated with morally significant properties, or because races are intrinsically morally significant). As Appiah noted (1990, p. 8; our emphasis), the dispositions themselves consist of

"(...) assent[ing] to false propositions, both moral and theoretical, about races—propositions that support policies or beliefs that are to the disadvantage of some race (or races) as opposed to others (...) and *to do so even in face of evidence and argument that should appropriately lead to giving those propositions up.*"

Racism would thus be a combination of false beliefs (for example, in the inferiority of one group compared to another) and of a cognitive resistance to rational revision of those beliefs in light of

³ Philosophers have also developed hybrid versions of these models (e.g., Corlett, 2003; Dummett, 2004).

⁴ These models typically do a double duty: They define what racism is and explain why it is wrong. [For the sake of space, we will not attempt to disentangle what racism is and what makes it wrong. We merely note that it might be](#) that these two tasks cannot be separated: [It might be that](#) one cannot understand what racism is without understanding why it is wrong, in the same way as one cannot understand what injustice is without understanding why it is wrong (for a different view, see Shelby, 2002).

countervailing evidence.⁵

1.2 The Behavioral Model

The behavioral model insists that racism essentially involves behaving or being disposed to behave in a way that is harmful to the members of a racial group. Michael Philips has proposed a version of this model in which the expression “racist” is used in “a logically primary sense” to qualify actions (what he calls “Basic Racist Acts”). Thus, he wrote (1984, p. 77):

“... P performs a Basic Racist Act by doing A when: (a) P does A in order to harm Q because Q is a member of a certain ethnic group; or (b) (regardless of P’s intentions or purposes) P’s doing A can reasonably be expected to mistreat Q as a consequence of Q’s being a member of a certain ethnic group.”

And he continued:

“Note that, on this account, P’s motives, beliefs, feeling, or intentions need not be taken into account in determining that P performed a racist act.”

In his book *Racism: A Short History*, Fredrickson has defended another version of the behavioral model (2002, p. 170):

“Racism exists only when one ethnic group or historical collectivity dominates, excludes or seeks to eliminate another on the basis of differences that it [the first group] believes are hereditary and unalterable.”

In Fredrickson’s version of the behavioral model as in the doxastic model, racism involves believing in the existence of racial differences (again racialism); however, in contrast to the doxastic model, this belief must be entertained by an ethnic group or by a collectivity that is able and disposed to act against another group. Behavior is thus essential to racism: Without it, one cannot truly talk of racism; there is only racialism. Commenting on and endorsing Appiah’s distinction between racialist and racist, Fredrickson wrote (2002, p. 154):

⁵ In a recent paper (2002), Appiah has partly revised this doxastic model, by emphasizing the affective aspects of racism.

“Racialists do not become racists until they make such convictions the basis for claiming special privileges for members of what they consider to be their own race, and for disparaging and doing harm to those deemed racially Other.”⁶

1.3 The Affective Model

While the behavioral model emphasizes people’s behavior and the doxastic model people’s beliefs, the affective model proposes that someone is racist if and only if she has some specific negative emotion or emotions (or some disposition to have this emotion or these emotions) toward specific races and that an action or a thought is racist if and only if it results from some such emotion(s) (or emotional disposition(s)).⁷ In a series of articles, Garcia (1996/2003, 1997, 1999, 2004) has elaborated a particularly clear and influential version of the affective model.⁸ He has proposed that a behavior (an institution, a thought, etc.) is racist if and only if it results from a vicious attitude, namely hate or malevolence toward (or lack of benevolence for) the members of a particular race, and that a person is racist if and only if she feeds on this hate and malevolence. As he has put it (1996/2003, p. 259), “in its central and most vicious form, it [racism] is a hatred, ill-will, directed against a person or persons on account of their assigned race.” What underlies those racial judgments and actions deemed to be racist is malevolence, a “deformation of character” (2004, p. 41) that is opposed to benevolence and justice (as he sometimes puts it, opposed to “benevolent love”).

Locating racism in people’s heart has many consequences. First, contrary to what the doxastic model asserts, for Garcia, it does not appear to be necessary to believe in the existence of races or in their inequality to be racist (1996/2003, p. 263). Someone can believe that races do not exist or

⁶ It is helpful to distinguish between structural (or institutional) and individualist variants of this model of racism. In the structural variant, racism may exist even when people are not racist, provided that some social institutions discriminate along racial lines (Haslanger, 2004). In the individualist variant, people are racists, depending on how they behave.

⁷ Kim, 1999; Levine, 2004; Arthur, 2007.

⁸ Garcia (1996) has been reprinted in many anthologies and is a *locus classicus* for race theorists who discuss the affective model.

that all races are equal, and yet still be racist if she has a malevolent attitude toward a particular racial group (2004, p. 44).

Moreover, someone can be racist even if she is powerless and cannot harm the members of another race, provided that an impure heart motivates her beliefs or intentions. For instance, someone could live in an isolated area, without any contact with the members of a specific race, yet still be racist if she has a vicious attitude toward them (1996/2003, p. 268).⁹

Contrary to what Frederickson claims, the existence of racism does not depend on the existence of a group acting or wishing to act against another group. An individual may be racist in a non-racist society (2004, p. 45). Racism does not even need to be directed toward another group; it can be directed toward the group to which the individual belongs.

An institution can also produce injustices following racial lines without being racist. Universities' admission policies that privilege students from specific high schools and that, as a side-effect, harm students from racial minorities are not racist if they were adopted via a process involving no malevolent motives. An institution would be racist only if its establishment was intended to fulfill some malevolent goals against some racial group or if it was currently organized so as to fulfill these goals (1996/2003, p. 266). In fact, such an institution would still be racist today even if it no longer affected the targeted group the way initially intended (1996/2003, p. 291-292).

Finally, at the individual level, a joke may be offensive to the members of a racial group without being racist if it is not motivated by an impure heart. Similarly, preferring one's own racial group (for example, creating academic fellowships for students belonging to one's own racial group) is not a form of racism if this preference is not caused by hatred toward other racial groups or by indifference toward them (1996/2003, p. 277-279).

One practical implication of Garcia's affective model is that to eradicate racism, it is not sufficient to change people's beliefs and behaviors toward the members of other races; one needs to eradicate the affective source of racism in people's character. As Garcia wrote (1999, p. 19):

⁹ As Garcia notes, stating that a powerless individual and an individual who actually discriminates because of her racist motivations are racist for the same reason—both have the same racial hatred—does not erase the moral differences between them. One can still say that the former individual is morally less condemnable than the latter.

“It means that to eliminate racism it is not enough that they change some laws or hiring practices; they themselves must change, and change in their depths—their hearts and minds.”

Garcia thus thinks that for racism to disappear, it is necessary to change the affective structure of the racist's psyche.

Our goal in this article is *not* to adjudicate between these three models of racism. While we think that each of these models could be improved by the consideration of the psychological literature, in what follows, we will focus only on the affective model. Indeed, we will focus on Garcia's version of the affective model (because it is particularly clear and influential) and argue that psychology undermines it. Thereby, we will illustrate how psychology can be relevant to the debates about the nature and moral significance of racism.

We have two main reasons for focusing on the affective model. First, some strong arguments have been developed against the doxastic and behavioral models. It has been argued against the behavioral model that being able and disposed to act against a racial group is not necessary for being racist or for behaving in a racist manner. Intuitively, someone who knows that she is powerless and has thus formed no intention to harm the members of another race can still be racist (Appiah, 2002; Blum, 2004a; Taylor, 2004; Glasgow, ms). Furthermore, being able and disposed to act against a racial group is not sufficient for being racist or for behaving in a racist manner. A behavior that harms exclusively the members of a particular racial group is racist only if the agent has intended to harm the members of this racial group qua members of this group (Haslanger, 2004; Levine, 2004). Against the doxastic model, it has been argued that beliefs are not necessary for racism. Someone who has some negative emotions (for instance, fear) toward a certain group without having any particular set of beliefs about this group is a racist (Todorov, 2000). Furthermore, beliefs might not be sufficient for an individual to be a racist (Dummett, 2004). Someone can sincerely believe in the inferiority of a racial group without being racist, if his belief does not lead to discrimination but to helpful actions aimed at remedying the alleged inferiority of this group.

Second, the affective model is a natural ally for those who hold that psychology can bear on the debates about racism.¹⁰ In pure (i.e., non hybrid) versions of the behavioral model, racism is not characterized by reference to people's psychological states but by reference to their behavior, while the doxastic model limits the psychological states relevant for characterizing racism to a single kind of psychological state—viz. people's beliefs. By contrast, although the affective model also characterizes racism by reference to a single kind of state—viz. emotions—it highlights the effect of these states on our beliefs, judgments, moods, and decisions—in brief, on our whole mind. It is thus natural to focus on this model in order to show that psychology can contribute to the debates about the nature and moral significance of racism.

1.4 Blum's Critiques

As one would expect, Garcia's affective model of racism has been challenged (e.g., Shelby, 2002; Mills, 2003; Glasgow, ms). We will concentrate here on two critiques developed by Blum (2002, 2004a, 2006), which question Garcia's monistic approach of racial phenomena. In both critiques, Blum advocates a form of pluralism.

First, we focus on Garcia's moral monism—the view that all racial ills are wrong for the same reason. Blum (2002, 2004a) has argued that Garcia's model cannot explain what makes numerous racial phenomena morally wrong. There are—to use his expression—some “racial ills” that are not caused by malevolence or by an absence of benevolence.¹¹

Blum gives a few commonsensical examples: word-of-mouth recruiting, discomfort with the members of other races, not acknowledging a person's racial identity, racial insensitivity, racial injustice, and undeserved racial privileges. Although these racial phenomena do not necessarily

¹⁰ We are grateful to a reviewer for this suggestion.

¹¹ Garcia would perhaps reply that he did not intend to characterize *all* the moral ills that are related to races, but *only* those that are due to racism. There are at least two problems with this reply. First, many of the racial ills discussed in the remainder of this article are plausibly manifestations of some form of racism, although they are not due to any malevolence. Second, whether or not these racial ills are manifestations of some form of racism, Garcia would have to acknowledge that his moral analysis is incomplete.

result from any malevolent intention, they can nonetheless be morally wrong. To use again one of Blum's examples, a schoolboard that decides to include dairy products in breakfasts offered to underprivileged students, but that is not aware of the fact that many students are lactose-intolerant due to their ancestral background harms Black children; but its action is not due to any form of malevolence.

Blum neither assumes nor looks for a fundamental unity underlying all racial ills and therefore opposes Garcia's moral monism. He proposes instead that an analysis of what is morally wrong with racial phenomena should take into account the “diversity of racial phenomena that constitute moral ills and a careful delineation of the moral character of each” (2004a, p. 77).

We now turn our attention to Garcia's psychological monism. Blum (2006) has recently argued that Garcia's affective model is not based on a sufficiently rich description of the racist's psychology. For instance, Garcia does not explain the link between the racist's motivational structure and her emotions—such as hate or disgust; nor does he characterize how the racist perceives the members of other races. Let's illustrate these two failures. For Garcia, the racism toward Blacks is psychologically similar to, and is wrong for the same reasons as, the racism toward Jews or the prejudices against homosexuals (1996/2003, p. 262). However, these three prejudices might rather involve distinct perceptions: Blacks might be seen as inferior (and as a result might be despised), Jews as being superior (and as result might be the object of jealousy), and gays as immoral (and as a result might be found disgusting). Because racists might perceive different racial groups differently, they might have different emotions toward them; as a result, they might be motivated to behave differently toward different racial groups: Racists might want to avoid those racial groups that disgust them, but they might want to harm those racial groups that are perceived as having more than they deserve or as taking advantage of their (i.e., the racists') group.

We agree with Blum that there are racial ills that do not result from malevolence. Thus, like Blum, we favor a fragmented explanation of what makes some racial phenomena morally wrong. We also believe that if one endorses the affective model, one cannot understand racism without a richer psychological description that casts some light on the complexity of the affective structure

of the racist mind.¹² In brief, we agree with Blum in rejecting Garcia's moral and psychological monism.

Nonetheless, we contend that in order to rebut the monistic approach to racial phenomena, it is useful to call upon psychology. Psychology is needed in order to correctly describe the numerous emotions that make up the racist mind and it helps us discover some racial ills that may otherwise remain unnoticed. In what follows, we will turn our attention to recent work in psychology so as to fulfill these two goals.

2 Psychological Pluralism

The research in social psychology that we will consider in this section in fact indicates that racism can result from many different emotions: The members of a given racial group may harbor distinct combinations of emotions toward different racial groups. We will support this claim by looking successively at psychologists Catherine Cottrell and Steven Neuberg's work (Neuberg and Cottrell, 2002; Cottrell and Neuberg, 2005) and at psychologist Susan Fiske's Stereotype Content Model (Fiske et al., 2007).

Cottrell and Neuberg argue that the traditional characterizations of prejudice and racism in psychology have neglected the rich texture of the emotions that individuals feel toward various social and racial groups (for a similar remark, see also Smith and Mackie, 2005; Fiske et al., 2007).¹³ Because Cottrell and Neuberg endorse a sociofunctional approach to emotions, we first describe this approach.

¹² In fact, even if one accepted Garcia's thesis concerning the central role of hate in racism, a more nuanced analysis would be necessary, since according to Sternberg (2003, p. 306), "[t]ypically hate is thought of as a single emotion [...] [b]ut there is reason to believe that hate has multiple components that can manifest themselves in different ways on different occasions". For Sternberg, hate is composed of the following elements: disgust, repulsion, anger, fear, and dislike.

¹³ To take only two examples, Allport describes a prejudice as being "[a]n antipathy based upon a faulty and inflexible generalization" (1954, p. 10), while for Fishbein, it is "[a]n unreasonable negative attitude toward others because of their membership to a particular group" (1996, p. 6).

According to the functional approach to emotions, each emotion (fear, disgust, anger, embarrassment, etc.) is a specific response (or a set of coordinated responses) to a specific problem. For example, psycho-physiologist Robert Levenson characterized emotions as follows (1994, p. 123; see also Keltner and Haidt, 1999, 2001):

"Emotions are short-lived psychological-physiological phenomena that represent efficient modes of adaptation to changing environmental demands. Psychologically, emotions alter attention, shift certain behaviors upward in response hierarchies, and activate relevant associative networks in memory. Physiologically, emotions rapidly organize the responses of disparate biological systems including facial expression, somatic muscular tonus, voice tone, autonomic nervous system activity, and endocrine activity to produce a bodily milieu that is optimal for effective response. Emotions serve to establish our position vis-à-vis our environment, pulling us toward certain people, objects, actions and ideas, and pushing us away from others."

The socio-functional approach to emotions is a version of the functional approach. In contrast to other functional approaches, which focus more on the problems posed by the physical environment (e.g., avoiding toxic food), it highlights the problems that we encounter in our social life. The problems posed by the members of other groups to our own group are amongst these problems. Some emotions constitute an answer to those problems by signaling that some action is needed to solve these problems and by motivating the form of appropriate action. The emotions evoked by a particular racial group should thus correspond to the problems that that group is seen as posing. Because a given racial group might be seen as posing several problems, it might evoke a combination of emotions. It is also plausible that different racial groups can be seen by the very same group as posing different types of problems (contamination, physical danger, economic danger, etc.). Thus, such distinct racial groups should evoke different emotions. Further still, sub-groups within these groups may be seen as posing different problems and may thus themselves evoke different emotions. For instance, in the United States, it is plausible that young Black men evoke fear, while Black women evoke a different emotion (unrelated to fear).¹⁴

¹⁴ We specify "in the United States" because the problems that a group is seen as posing may well vary from place to place.

These predictions are validated by the experiments run by Cottrell and Neuberg with White subjects in Vancouver. When White subjects were asked what types of emotions African Americans, Asians, and members of the First Nations evoke, it was found that African Americans evoke mostly fear (but also disgust and anger), members of the First Nations pity (and some anger), and Asians envy. Notably, each group evoked an equivalent measure of prejudice (as measured by questionnaires designed to probe explicit attitudes toward members of other groups); thus, prejudices that are superficially similar may in fact conceal a striking diversity of emotions. Cottrell and Neuberg also examined how subjects perceived the problems posed by these racial groups. They found that African Americans were seen as posing problems for property, health, reciprocity, social coordination, and security. Other groups (such as Asians) were seen as posing problems for the economy and the dominant values, but not for social coordination or reciprocity.

Other theoretical approaches agree with the sociofunctional approach to emotions that different racial groups evoke different emotions. For instance, Fiske and colleagues' (2007) Stereotype Content Model links the emotional reaction evoked by a social group to its perceived competence (whether group members are stereotypically judged intelligent and skillful) and warmth (whether group members are judged friendly, helpful, and trustworthy). Groups that are perceived as warm and competent (in the USA, White American middle-class males) evoke pride, while groups that are seen as cold and competent (in the USA, Jews or Asians) evoke envy. Groups that are seen as warm but incompetent (e.g., seniors) evoke pity, while groups that are seen as cold and incompetent evoke contempt or disgust. Like Cottrell and Neuberg's model, the Stereotype Content Model also predicts that subgroups within a given ethnic or racial group (e.g., male vs. female African Americans) may evoke different emotions (Fiske et al., 2007, p. 81).

The findings and models we have reviewed suggest that the emotions felt toward a racial group depend on the perception of the problems it is seen as posing and on the stereotypical characteristics associated with the members of this group. Different racial groups can thus have distinct emotions toward a given racial group, if they do not view this group as posing the same problem or if they associate different stereotypical characteristics with it. A given racial group can also be seen as posing more than one problem (for example, contamination or economic threat) and it can evoke several distinct emotions.

These findings are relevant for the proponents of the affective model of racism. Recall that the affective model proposes to define racism by reference to the emotions people might have toward racial groups. We have seen however that a range of emotions can underlie people's negative and prejudiced attitude toward others. These findings are clearly at odds with Garcia's exclusive focus on hate and malevolence. In order to preserve his version of the affective model, Garcia would perhaps reply that a prejudiced attitude toward a racial group that results from fear, disgust, or jealousy rather than from malevolence is not racist. The trouble with such a reply is that it is unprincipled: We see no reason not to count a prejudiced attitude caused by fear, disgust, or jealousy as racist. Garcia would perhaps also reply that fear, disgust, or jealousy toward a racial group are likely to be interwoven with some form of malevolence: People probably hate what they are afraid or jealous of. He would then add that what makes this attitude racist is the malevolence interwoven with fear, disgust, and jealousy. In reply, we note that although fear, disgust, or jealousy can certainly be interwoven with malevolence, they are not necessarily so interwoven. A negative and prejudiced attitude toward a racial group caused by disgust can exist without malevolence.

We conclude that Garcia's psychological monism is incorrect. To understand racism from the perspective of the affective model, it is crucial to study the functioning of several emotions (such as fear, disgust, anger, and envy), the cultural representations of the problems racial groups are seen as posing, and the stereotypes associated with particular racial groups.

3 Implicit Racial Biases

We now turn to Garcia's monistic moral characterization of racism. We argue that a large body of research in psychology—the research on people's implicit racial biases—undermines his moral monism. In this section, we review this body of research, and we show in the next section how it bears on Garcia's moral views.

3.1 Stereotypes and Evaluations

Contemporary social psychology distinguishes between explicit and implicit racial biases. Explicit biases are those biases that people are aware of having and that they can express (of

course, they may also decide to stifle these biases for, e.g., the sake of political correctness). By contrast, other biases are implicit in the sense that people may be unaware of having them. Social psychologists have repeatedly demonstrated that people may have no explicit racial biases while at the same time harboring some implicit racial biases toward the members of some racial group (for summary and discussion, see Kelly and Roedder, 2008; Kelly et al., in press; Faucher, in press).

Contrary to explicit biases (which may be measured by means of questionnaires such as the Modern Racism Scale), implicit biases must be assessed by means of indirect measures. One of these measures is a version of the Stroop task.¹⁵ White subjects are shown the picture of the face of a White man or of a Black man followed by a word with either a positive (e.g., “birth”) or negative (e.g., “death”) connotation. Subjects have to decide whether the word is negative or positive by pressing one of two keys. It is thought that this decision will be faster if the word is evaluatively congruent with the face than if it is not congruent. It has been repeatedly found that that after having seen a picture of the face of a Black man, White subjects tend to recognize a word with a negative connotation more quickly than a word with a positive connotation. This tendency is found even when subjects do not have any explicit bias (as measured by the Modern Racism Scale).

According to Amodio and Devine (2006), one must distinguish at least two types of implicit biases, namely implicit negative evaluations and implicit stereotypes. Implicit evaluations are negative affective responses directed at the members of specific racial groups, while implicit stereotypes are culturally shared beliefs about the members of these racial groups. Amodio and Devine show that implicit evaluations and implicit stereotypes have different effects. Implicit evaluations are good predictors of negative behaviors. For example, they predict the frequency of unfriendly attitudes toward the members of other races; they also predict the frequency of uncomfortable interactions, which may manifest themselves through hesitations, increased

¹⁵ In a Stroop task, subjects are asked to name the ink color of a written word. This word is the name of a color that is either congruent with the ink color (for example, “black” is written in black ink) or incongruent (“black” is written in blue ink). Subjects are slower to respond when they must identify colors that are incongruent with the words written: There is an interference between the color and the word.

blinking in a conversation with someone belonging to another race, or a tendency to sit further away from the members of other races. On the other hand, stereotypes predict cognitive effects, such as memory biases (for example, remembering more easily that a violent event was perpetrated by a Black person than by a White person) or perceptible biases (for example, finding the same behavior more violent when perpetrated by a Black person than by a White person). It would also appear that evaluations and stereotypes depend on different cerebral structures, the amygdala for the former and evolutionary more recent cortical zones for stereotypes.

To briefly summarize what we have just presented, social psychology posits the existence of two types of biases, viz. implicit and explicit biases. Both types of implicit biases can exist even when people have no explicit biases, and even when they have consciously and sincerely decided to be unbiased. There are two types of implicit biases—viz. evaluations and stereotypes—that may have different cerebral bases.

3.2 The Effects of Implicit Biases

Why be preoccupied with implicit biases? One reason is that they occasionally affect our behavior. In a frightening situation, access to the stereotypes about the cause of our fear reaction is facilitated. Attention is then directed toward those characteristics that are stereotypical of what has triggered our reaction (e.g., the objects that are stereotypically associated with what has frightened us). This lowers the detection threshold for detecting these characteristics. That is, when people are afraid, they need to be less certain that they see these characteristics to conclude that they really see them than when they are not afraid. Frightened people are thus more likely to imagine seeing these characteristics.¹⁶ Recall now that implicit biases involve both affective evaluations and stereotypes. Because implicit biases involve affective evaluations, other races should more often evoke some negative affective reactions when people have some implicit biases than when they don't; in addition, because implicit biases involve stereotypes, the consequences of these affective reactions should be more far-reaching when people have implicit biases than when they don't.

¹⁶ Detection thresholds are also lowered when we have to make a decision under time constraints or when we are tired or intoxicated.

Payne (2006) has thus shown that when a racial group is stereotypically associated with danger, people have a lower threshold for judging that a dangerous object is present. In pictures of members of the racial group in question, people more quickly, and more mistakenly, identify items as guns. In addition, people are faster to decide to shoot.¹⁷ These effects are stronger in implicitly biased people. In one experiment, Payne showed a male face (Black or White), just before an object appeared on the screen. He asked subjects to perform a visual discrimination task by identifying whether the object was a gun or a tool. Subjects were instructed to ignore the face that appeared before the object. In one condition, subjects had a great deal of time to respond, while in another condition, subjects had only half a second to respond. The results are striking (Payne, 2006, p. 287-288):

“In the snap-judgment condition, race shaped people’s mistakes. They falsely claimed to see a gun more often when the face was black than when it was white. Under the pressure of split-second decision, the readiness to see a weapon became an actual false claim of seeing a weapon.”

It should be noted that these effects are found even when subjects are explicitly instructed *not* to pay attention to the race of the faces on the screen, because it may influence them. It thus seems likely that implicit biases can influence one’s behavior even when one sincerely and actively attempts to be non-biased.

Other experiments confirm these results. Correll and colleagues (2002) asked White subjects to watch a screen on which appeared (either White or Black) men holding a revolver, a soda pop

¹⁷ Payne states that two factors are implicated in the decision to shoot: implicit biases and voluntary control. When voluntary control is impossible (due to time constraints or to impaired control skills following intoxication or fatigue), implicit biases strongly influence the decision to shoot. It is currently unknown whether the effects he describes are caused by implicit stereotypes or by implicit evaluative biases (although Payne thinks that fear is probably involved in the weapon bias) (Payne, personal communication, October 22, 2007). As he writes, “But it is unknown under what conditions one or the other is likely to be influential. Do emotional and semantic responses act in identical ways, or do they have different consequences? And do the mechanisms of control differ for emotional versus semantic responses?” (2006, p. 290)

can, a mobile telephone, or a wallet. Subjects were told to shoot if the character on the screen was armed, and to refrain from shooting if he was unarmed. Finally, subjects were also asked to pay attention to the character’s face, as there would be a recognition test at a later time. The results are similar to Payne’s findings: People shoot armed Black men faster than they shoot armed White men. People also decide not to shoot a White unarmed man faster than a Black unarmed man. Finally, people also decide more often not to shoot a White armed man than a Black armed man. Eberhart and colleagues (2004) also showed that seeing an African American’s face before seeing a series of blurry images lowered the threshold for detecting objects associated with danger and crime (like guns).

We have just seen that in addition to explicit biases, people also harbor implicit biases, which include implicit evaluations and implicit stereotypes. Typically, people have these biases in spite of having decided to be unbiased and in spite of their sincere efforts to be unbiased. Implicit biases can influence judgments and behaviors in situations of stress, fatigue, or intoxication, or under time constraints (Payne, 2006).

4 Racial Ills without Malevolence

These psychological findings about racial cognition, which go much beyond an intuitive and non-scientific understanding of racial phenomena, strongly reinforce Blum’s first critique of Garcia. Recall that for Garcia, racism is essentially affective and that racist behaviors (judgments, etc.) are morally wrong because they result from malevolence or from a lack of benevolence toward the members of a given race. Without rejecting the idea that malevolence and lack of benevolence explain why some racial ills are morally wrong, Blum notes that many racial ills are not the result of malevolence or of a lack of benevolence.

The existence of implicit biases clearly establishes this point. There is no doubt that many actions caused by people’s implicit biases (and perhaps all of them) are morally wrong. Because of these biases, the members of some racial groups (e.g., those that are perceived as threatening individual security or property) have a higher risk of being victims of racial profiling; they are also more likely to be viewed as displaying suspect behaviors; they may be more likely to be surveyed by the police or by the security services, and so on and so forth. Such discriminations can feed a feeling of injustice, which may have repercussions in an individual’s personal life. Furthermore,

independently of the actions they may cause, implicit biases in and of themselves are arguably morally wrong (for discussion, see Kelly and Roedder, 2008; Blum, 2004b).

For present purposes, the crucial point is that typically, people's implicit biases result from neither malevolence nor a lack of benevolence (nor do they necessarily result in such attitudes).¹⁸ This point becomes clear when one considers the fact that implicit biases may exist in individuals who are sincerely benevolent with respect to the members of other races. One can roughly define benevolence as the tendency to help others and to behave so as to increase the likelihood that others will reach what is good for them. It is typically accompanied by a disposition to sympathize with others' suffering and misfortune. Such behavioral dispositions may well be present in an individual who harbors some implicit biases. These biases might be dormant, because they are never applied in interracial interactions. An individual who harbors them might never classify others as belonging to a race; rather, she might treat people's skin color (and other racial features) as she treats people's hair or eye color (she is therefore race blind) and as a result, she might be entirely benevolent toward the members of other races.¹⁹ Alternatively, in addition to her implicit biases, she might possess some non-stereotypical beliefs and some positive evaluations about races (for example, the beliefs that races have no biological meaning, that race is social in nature, that there is no correlation between race and personality, etc.) and rely on these beliefs and evaluations in racial interactions. These beliefs and evaluations might then block the expression of her implicit biases.

So, some genuinely benevolent person can have implicit biases. Thus, these biases do not necessarily result from malevolence or from a lack of benevolence. However, they may be a source of moral ills—a fact demonstrated by Payne's research and also, unfortunately, by

¹⁸ To clarify further, our argument is not that the mere existence of implicit biases undermines Garcia's moral monism. Our argument is that these biases, which result in morally condemnable behaviors and which themselves might be morally wrong, *result neither from malevolence nor from a lack of benevolence*. Thus, not every racial ill results from malevolence or from a lack of benevolence.

¹⁹ Wheeler and Fiske (2005) indicate that when one does not identify others as members of racial categories, some effects that are typical of racial cognition disappear (this is also true of other social categories).

numerous news items—and they may themselves be moral ills. Hence, as Blum argued against Garcia, some moral ills are caused neither by malevolence nor by the absence of benevolence.

It is instructive to compare the racial ills we focused on in light of the psychology of racial prejudice and the racial ills mentioned by Blum to criticize Garcia. As shown in Section 1, Blum emphasizes that some actions—for example, recommending the distribution of dairy products to schoolchildren in African-American neighborhoods—may cause harm to racial groups. These racial ills do not result from the agents' emotions. Rather, they result from their ignorance: For instance, the individuals who decided to distribute dairy products did not know that an incapacity to digest lactose was more common among African Americans than in the general population. This ignorance itself was not necessarily due to malevolence or to a lack of benevolence; it may simply have been due to the fact that this particular piece of information was not available in the agents' environment. By contrast, when one considers the actions caused by implicit biases or the implicit biases themselves, one finds some racial ills that result from the affects of the agents.²⁰ These racial ills are more deeply anchored in people's psyche than the mere ignorance of some relevant fact is. Thus, psychology reveals at least one kind of racial ills that does not result from any kind of malevolence, but that results from something deeper than a mere form of ignorance.

We emphasize that it was only through empirical psychological research, and not armchair reflexion or conceptual analysis, that this deeper layer of psychological phenomena could be unearthed. These phenomena may not fit into any commonsense conception of the mind, yet they are relevant to the moral issues surrounding race.²¹

Garcia could perhaps respond that the implicit biases are a form of malevolence or, at the very least, that they show a lack of benevolence. After all, those who harbor some implicit stereotypes toward the members of a particular group or who negatively evaluate them cannot possibly be as benevolent as those who lack those stereotypes or who do not evaluate these members negatively.

To see why this reply is misguided, one needs only consider the fact that in some cases, *but not in others*, implicit biases are malevolent (or show a lack of benevolence). Thus, in and of themselves, they neither are malevolent nor show a lack of benevolence. Someone who dislikes African Americans might only watch racist television stations or only read right-wing racist

²⁰ At least for those implicit biases that consist of negative evaluations.

²¹ We are especially grateful to Dan Kelly for having insisted on this point.

newspapers. Her implicit biases and her resulting behavior would then be interwoven with her malevolence toward African Americans. By contrast, someone might have some implicit biases merely by watching television or reading the newspapers; but she might not harbor any ill-will toward the members of other races. Her implicit biases would not be malevolent.

Conclusion

It is a mistake to view racism as a simple phenomenon, motivated by a single type of emotion. Rather, psychology suggests that people's attitudes toward a particular racial group typically result from a combination of emotions that depend on the problems that this group is seen as posing and on the stereotypical characteristics associated with this group. Different groups are also likely to evoke different combinations of emotions. Different subgroups within a given racial group can even evoke different emotions. Consequently, we reject any form of psychological monism in order to endorse a kind of psychological pluralism. To understand racism from the perspective of the affective model, it is thus crucial to acknowledge the role of numerous emotions.

Psychology was also useful to rebut Garcia's moral monism: Some racial ills result neither from malevolence nor from a lack of benevolence. Even if people's explicit racist beliefs, their racial malevolence, and the resulting behaviors could be eliminated, some racial ills might still stem from people's heart—viz. from their implicit biases. In this sense, Garcia was correct: "Danger is lurking in the heart of the subjects." But he was wrong to assert that the racial ills that stem from people's hearts are all the result of malevolence: Malevolence is not the only danger to be lurking in people's hearts.

Psychology has played a key role in our two-pronged critique of Garcia's views about racism. Rather than merely appealing to our commonsense, mundane understanding of the racist mind, we have relied on robust, experimentally established evidence that racism results from a range of emotions. Second, and more strikingly, research on implicit biases has revealed a kind of racial ills that stems from the heart, but not from hatred. Our critique clearly illustrates that psychology can fuel new arguments relevant to the debates about the nature and moral significance of racism. More generally, we think that psychology can provide the following benefits to philosophers interested in these debates. When some philosophical arguments rely on claims about the racist

mind, psychology can keep our folk, unreliable, understanding of the mind in check. It can also reveal new, non-obvious, features of the racist mind that are relevant for these arguments. We hope to have shown that limiting oneself to a mundane understanding of the racist mind in effect means depriving oneself of original insights and arguments about the nature and moral significance of racism.

Bibliography

- Allport, G. 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley Publishing Co.
- Amodio, D. M. and P. G. Devine. 2006. "Stereotyping and Evaluation in Implicit Race Bias: Evidence for Independent Constructs and Unique Effects on Behavior". *Journal of Personality and Social Psychology*, vol. 91, p. 652-661.
- Appiah, K. A. 1990. "Racisms". In D. T. Goldberg (ed.), *Anatomy of Racism*, Minneapolis: University of Minnesota Press, p. 3-17.
- Appiah, K. A. 2002. "Racism: History of Hatred: A Review of George Frederickson's *Racism: A Short History*". *New York Times*, 4 August, Book Review Section, p. 11-12.
- Arthur, J. 2007. *Race, Equality, and the Burden of History*. Cambridge: Cambridge University Press.
- Blum, L. 2002. *I am not a Racist But ...": The Moral Quandry of Race*. Ithaca: Cornell University Press.
- Blum, L. 2004a. "What Do Accounts of 'Racism' Do?". In M. P. Levine and T. Pataki (eds.), *Racism in Mind*, Ithaca: Cornell University Press, p. 56-77.
- Blum, L. 2004b. "Stereotypes and Stereotyping: A Moral Analysis". *Philosophical Papers*, vol. 33, p. 251-289.
- Blum, L. 2006. "Racial Virtues". In R. Walker and P. J. Ivanhoe (eds.), *Working Virtues: Virtue Ethics and Contemporary Moral problems*, New York: Oxford University Press, p. 225-250.
- Corlett, J. A. 2003. *Race, Racism and Reparations*. Ithaca: Cornell University Press.

Correll, J., B. Park, C. M. Judd, and B. Wittenbrink. 2002. "The Police Officer's Dilemma: Using Race to Disambiguate Potentially Threatening Individuals". *Journal of Personality and Social Psychology*, vol. 83, p. 1314-1329.

Cottrell, C. A., and S. L. Neuberg. 2005. "Different Emotional Reactions to Different Groups: A Sociofunctional Threat-Based Approach to 'Prejudice'". *Journal of Personality and Social Psychology*, vol. 88, p. 770-789.

Dummett, M. 2004. "The Nature of Racism". In M. P. Levine and T. Pataki (Eds.), *Racism in Mind*, Ithaca: Cornell University Press, p. 27-34.

Eberhart, J., P. A. Goff, V. J. Purdie, and P. G. Davis. 2004. "Seeing Black: Race, Crime, and Visual Processing". *Journal of personality and Social Psychology*, vol. 87, p. 876-893.

Faucher, L. In press. "Unity of Science and Pluralism: Cognitive Neurosciences of Racial Prejudice as a Case Study". In S. Rahman, J. Symons, M. Torres and O. Pombo (eds.), *In Unity of Science: Non-Traditional Approaches*, Dordrecht : Springer.

Fishbein, H. D. 1996. *Peer Prejudice and Discrimination: Evolutionary, Cultural, and Developmental Dynamics*. Boulder, Colorado: WestView co.

Fiske, S. T., A. J. C. Cuddy, and P. Glick. 2007. "Universal Dimensions of Social Cognition: Warmth and Competence". *Trends in Cognitive Sciences*, vol. 11, p. 77-83.

Frederickson, G. 2002. *Racism: A Short History*. Princeton: Princeton University Press.

Garcia, J. L. A. 1996/2003. "The Heart of Racism". *Journal of Social Philosophy*, vol. 27, no. 1; reprinted in B. Boxil (Ed.), *Race and Racism*, Oxford: Oxford University Press, p. 257-296.

Garcia, J. L. A. 1997. "Current Conceptions of Racism: A Critical Examination of Some Recent Social Philosophy". *Journal of Social Philosophy*, vol. 28, p. 5-42.

Garcia, J. L. A. 1999. "Philosophical Analysis and the Moral Concept of Racism". *Philosophy and Social Criticism*, vol. 25, p. 1-32.

Garcia, J. L. A. 2004. "Three Sites for Racism: Social Structures, Valuing, and Vice". In M. P. Levine and T. Pataki, *Racism in Mind*, Ithaca: Cornell University Press, p. 35-55.

Glasgow, J. Ms. "Racism as Disrespect".

Haslanger, S. 2004. "Oppressions: Racial and Other". In M. P. Levine and T. Pataki (Eds.), *Racism in Mind*, Ithaca: Cornell University Press, p. 97-123.

Kelly, D., E. Machery, and R. Mallon. In press. "Racial Cognition and Normative Racial Theory". In J. Doris (Ed.), *Handbook of Moral Psychology*, Oxford: Oxford University Press.

Kelly, D., and E. Roedder. 2008. "Racial Cognition and The Ethics of Implicit Bias". *Philosophy Compass*, vol. 3, p. 522-540

Keltner, D., and J. Haidt. 1999. "Social Functions of Emotions at Four Levels of Analysis". *Cognition and Emotion*, vol. 13, p. 505-521.

Keltner, D., and J. Haidt. 2001. "Social Functions of Emotions". In T. Mayne and G. A. Bonanno (Eds.), *Emotions: Current Issues and Future Directions*, New York: Guilford Press, p. 192-213.

Kim, D. H. 1999. "Contempt and Ordinary Inequality". In S. E. Babbitt and S. Campbell (Eds.), *Racism and Philosophy*, Ithaca: Cornell University Press, p. 108-123.

Lengbeyer, L. A. 2004. "Racism and Impure Hearts". In M. P. Levine and T. Pataki (Eds.), *Racism in Mind*, Ithaca: Cornell University Press, p. 158-178.

Levenson, R. W. 1999. "The Intrapersonal Functions of Emotion". *Cognition and Emotion*, vol. 13, p. 481-504.

Levine, M. P. 2004. "Philosophy and Racism". In M. P. Levine and T. Pataki (Eds.), *Racism in Mind*, Ithaca: Cornell University Press, p. 78-96.

Mills, C. 2003. "Heart's Attack: A Critique of Jorge Garcia's Volitional Conception of Racism". *The Journal of Ethics*, vol. 7, p. 29-62.

Neuberg, S. L., and C. A. Cottrell. 2002. "Intergroup Emotions: A Biocultural Approach". In D. M. Mackie and E. R. Smith (Eds.), *From Prejudice to Intergroup Emotions: Differentiated Reactions to Social Groups*, New York: Psychology Press, p. 265-283.

Payne, K. 2006. "Weapon Bias: Split-Second Decisions and Unintended Stereotyping". *Current Directions in Psychological Science*, vol. 15, p. 287-291.

- Philips, M. 1984. "Racist Acts and Racist Humor". *Canadian Journal of Philosophy*, vol. 14, p. 75-96.
- Shelby, T. 2002. "Is Racism in the Heart?". *Journal of Social Philosophy*, vol. 33, p. 411-420.
- Smith, E. R., and D. M. Mackie. 2005. "Aggression, Hatred, and Other Emotions". J. F. Dovidio, P. Glick, and L. A. Budman (Eds.), *On the Nature of Prejudice: Fifty Years after Allport*, Malden, MA: Blackwell, p. 361-376.
- Sternberg, R. J. 2003. "A Duplex Theory of Hate: Development and Application to Terrorism, Massacres, and Genocide". *Review of General Psychology*, vol. 7, p. 299-328.
- Taylor, P. C. 2004. *Race: A Philosophical Introduction*. Malden, MA: Polity Press.
- Todorov, T. 2000. "Race and Racism". In L. Back and J. Solomos (Eds.), *Theories of Race and Racism*, London: Routledge, p. 64-70.
- Wheeler, M. E., and S. Fiske. 2005. "Controlling Racial Prejudice: Social-Cognitive Goals Affect Amygdala and Stereotype Activation". *Psychological Science*, vol. 16, p. 56-63
- Zack, N. 1999. "Philosophy and Racial Paradigms". *The Journal of Value Inquiry*, vol. 33, p. 299-317.