

To appear in *The Oxford Handbook of Moral Psychology*

Please do not quote without permission of one of the authors; comments welcome at [drkelly@purdue.edu](mailto:drkelly@purdue.edu), [machery+@pitt.edu](mailto:machery+@pitt.edu), or [rmallon@philosophy.utah.edu](mailto:rmallon@philosophy.utah.edu)

Last Changes: 1/12/09

# Race

By Daniel Kelly, Edouard Machery, and Ron Mallon<sup>1</sup>

A core question of contemporary social morality concerns how we ought to handle racial categorization. By this we mean, for instance, classifying or thinking of a person as *Black*, *Korean*, *Latino*, *White*, etc.<sup>2</sup> While it is widely agreed that racial categorization played a crucial role in past racial oppression, there remains disagreement among philosophers and social theorists about the ideal role for racial categorization in future endeavors. At one extreme of this disagreement are short-term eliminativists who want to do away with racial categorization relatively quickly (e.g. Appiah 1995, D’Souza 1996, Muir 1993, Wasserstrom 2001 /1980, Webster 1992, Zack 1993, 2002), typically because they view it as mistaken and oppressive. At the far opposite end of the spectrum, long-term conservationists hold that racial identities and communities are beneficial, and that racial categorization – suitably reformed – is essential to fostering them (e.g. Outlaw

---

<sup>1</sup> We are grateful to the Moral Psychology Research Group for several useful discussions of this material, and are particularly thankful to John Doris, Tim Schroeder and Erica Roedder for their many insightful comments on earlier drafts of this paper. We would also like to thank Luc Faucher for his feedback on a previous version. Remaining mistakes are ours. Finally, we would like to thank Project Implicit (<http://www.projectimplicit.net/>) for permission to use their stimulus materials in this paper.

<sup>2</sup> In order to standardize terminology throughout the paper, we have elected to use “Black” and “White” rather than other options such as “African-American” or “White-American”.

1990, 1995, 1996). While extreme forms of conservatism have fewer proponents in academia than the most radical eliminativist positions, many theorists advocate more moderate positions. In between the two poles, there are many who believe that racial categorization is valuable (and perhaps necessary) given the continued existence of racial inequality and the lingering effects of past racism (e.g. Haslanger 2000; Mills 1998; Root 2000; Shelby 2002, 2005; Sundstrom 2002; Taylor 2004; Young 1989). Such authors agree on the short-term need for racial categorization in at least some domains, but they often differ with regard to its long-term value.

Our purpose here is not to delve into the nuances of this debate, nor is it to weigh in on one side or the other. Rather, we want to explore the intersection of these normative proposals with recent empirical work on the psychology of racial cognition. Race theorists often trade in normative arguments for conservationist or eliminativist agendas, and these normative arguments typically involve evaluations of the costs and benefits attached to those agendas (e.g., Boxill 1984; Appiah 1995; Muir 1993; D'Souza 1996; Outlaw 1990, 1995, 1996). For instance, these types of evaluations are present in Outlaw's discussions of the benefits of racial communities (1995), Appiah's (1996) weighing of the costs and benefits of racial identification, Sundstrom's (2002) insistence on the value of racial categorization in social science, and Taylor's (2004) exploration of the social and ethical dimensions of racial classification, which weighs the value of employing racial categories in different ways against the costs. Such evaluations invariably involve background assumptions regarding the feasibility of the proposals, and the ease with which racial categorization and racism can be eliminated or reformed.

Given how pervasive these assumptions about feasibility are, one might expect discussions regarding the role of human psychology in constraining or facilitating various reform proposals. Instead, contemporary race theory is nearly devoid of effort to engage the burgeoning literature from social psychology and cognitive science on racial categorization and racial prejudice. This is unfortunate, for as we show, the surprising psychological forces at work in racial cognition and related behavior often bear directly on the revisionist goals of conservationists and eliminativists. Our aim, then, is to demonstrate the need for normative racial philosophy to more closely engage the contemporary psychology of racial categorization and racial prejudice.

We begin section 1 by examining several positions within the philosophy of race in more detail, in the process pointing out where hitherto unappreciated facts about the psychology of race could have an impact upon the feasibility of reform proposals offered by philosophers. In Sections 2 and 3, we review two relatively separate sets of psychological literature: one from evolutionary cognitive psychology and the other from social psychology. Section 2 focuses on recent research on racial categorization, and argues that a large body of evidence shows that the content of racial thought is not a simple product of one's social environment, but is also shaped by the operation of certain evolved psychological mechanisms. Moreover, we show that this research has substantial implications for assessing the feasibility of eliminativist and conservationist proposals.

In Section 3, we turn to the question of racial evaluation, and consider recent studies of divergences between implicit and explicit racial cognition. This research program suggests that implicit racist biases can persist even in persons sincerely

professing tolerant or even anti-racist views; apparently implicit racial evaluations can be insulated in important ways from more explicitly held beliefs. We then argue, once again, that these findings bear on the feasibility of proposals made in the philosophical literature on race, and may be used to help shape novel proposals in the conservationist spirit. We conclude that though it has not received much discussion in the philosophy of race, the recent empirical work on racial cognition can have a direct impact upon the normative projects of race theory.

## **1. Race, Philosophy, and Psychological Research**

### **1.1. Thick Racialism and the Ontological Consensus**

The late nineteenth and early twentieth centuries were marked by the widespread endorsement of biologically rooted *racialist* doctrines - doctrines that divided human beings into putatively natural categories.<sup>3</sup> Such doctrines held that “natural” races exist, and that sorting people into racial groups on the basis of phenotypic features like skin color, hair type and body morphology also served to sort them according to a range of other underlying properties that expressed themselves in a variety of physical, cultural, moral, and emotional differences among the various races. We will call this view *thick racialism*. With the advent of modern genetics in the early twentieth century, it seemed obvious that the appropriate interpretation of such thick racialist claims was in terms of this emerging science of human heredity. In particular, it seemed that beliefs about the putative cultural, moral, and emotional differences between races would be vindicated by

---

<sup>3</sup> This is not to repeat the common claim that racialism was invented in the late nineteenth century (or at any other time, for that matter). See section 2.1.

the discovery of specific and systematic genetic differences between races. However, subsequent research in biology, anthropology, social theory, as well as cognitive, social, and evolutionary psychology has brought about a consensus that thick racialism is false. The reasons for this *ontological consensus* that thick racialism is false are many, but an increased understanding of human genetic variation played an important role in undermining the supposition that there are genetic characteristics shared by all and only members of a race.<sup>4</sup>

At the same time, there remains substantial debate about what could be called *thin racialism*, i.e., the idea that racial categorization might be useful in identifying *some* important genetic differences or other biological properties, -- for example, properties that might be useful for epidemiology, medicine, and forensic science.<sup>5</sup> Nevertheless, the important point for present purposes is that this ontological consensus against thick racialism is a point of agreement for all the authors we discuss below, and we will take it for granted it what follows.

## **1.2. Eliminativism, Conservatism, and Psychology**

We will call the normative philosophic position that recommends we do away with racial categories *eliminativism*. Eliminativists envisage a society in which there are no racial categorizations at all, typically because they believe that such categorizations are arbitrary and oppressive. For example, K. Anthony Appiah writes:

---

<sup>4</sup> Arguments referring to human genetic variation can found in, e.g., Andreasen 1998, p. 206; Appiah 1996, p. 68; D'Souza 1996; Kitcher 2000, p. 87-88; Zack 1993, p. 13-15. They are rooted in pioneering work done in human genetics in the 1970s by Nei and Roychoudhury (1972, 1974), Lewontin (1972), and others. For a recent review of human genetic variation, see Brown and Armelagos 2001.

<sup>5</sup> See, for example, the papers in *Nature Genetics* Nov. 2004 Supplement; Gannett 2005; Root 2005.

The truth is that there are no races: there is nothing in the world that can do all we ask "race" to do for us. The evil that is done is done by the concept and by easy – yet impossible – assumptions as to its application. (1995, 75)

Here Appiah articulates both of the ideas central to many contemporary eliminativist positions: the first being that thick racialism is false, the second that continued use of racial classification is oppressive.

In contrast, *conservationism* is the position that recommends we conserve racial categories, but do as much as we can to jettison their pernicious features.

Conservationists are best understood as offering proposals for (at least short term) rehabilitation of racial thinking, for conservationists typically advocate both the rejection of thick racialism and the eradication of *racism*, but hold that racial categories themselves should not be completely eliminated.<sup>6</sup> Outlaw, for example, agrees that "the invidious, socially unnecessary, divisive forms and consequences of thought and practice associated with race ought to be eliminated to whatever extent possible" (1995: 86), but thinks that "the continued existence of discernible racial/ethnic communities of meaning is highly desirable *even if, in the very next instant, racism and invidious ethnocentrism in every form and manifestation were to disappear forever*" (1995: 98). Conservationists like Outlaw appear to recommend a system composed of discernible racial groups, but one wherein those groups share equal social worth, as opposed to being hierarchically ranked.

Eliminativists and conservationists are best understood as revisionist: both suggest we reform current practices of racial categorization, but differ in whether it

---

<sup>6</sup> We follow the practice of using "racism" to involve both an endorsement of thick racialism *and* the evaluative ranking of races on the basis of the alleged natural distinctions between races.

would be best to eliminate or rehabilitate them. Commitment to either type of reform, however, appears to entail commitment to substantive, if often tacit, psychological assumptions as well.

Consider first eliminativism. What exactly would eliminativists like to eliminate? Politically conservative eliminativists (e.g. D'Souza 1996) are committed to the elimination of racial categorization *in public policy*. But many eliminativists (including a variety of liberal thinkers) have something much more sweeping in mind, and suggest reform extending from large-scale features of social organization all the way to *individual habits of thought and action*. In such normative proposals like this, that recommend altering individual's habits of thought and action, the psychological assumptions of eliminativism are fairly close to the surface. Consider, for example, a classic paper in which Richard Wasserstrom (2001 [1980]) writes:

A nonracist society would be one in which the race of an individual would be the functional equivalent of the eye color of individuals in our society today. In our society no basic political rights and obligations are determined on the basis of eye color. No important institutional benefits and burdens are connected with eye color. Indeed, except for the mildest sort of aesthetic preferences, a person would be thought odd who even made private, social decision by taking eye color into account. (2001, [1980], 323)<sup>7</sup>

Clearly, Wasserstrom's ideal involves a substantial re-ordering not only of contemporary social policies, but also of the patterns of categorization underwriting even personal behaviors and thoughts. Given this goal and the assumptions involved, work on the

---

<sup>7</sup> Note that Appiah's worry about the evil done by the *concept* of race suggests a similarly sweeping ideal.

psychology of racial categorization and racism is obviously relevant to assessing the ease with which (or the extent to which) such ideals can be realized. Moreover, if it turns out that certain ideals cannot be realized, that same psychological work will be useful in determining what sort of less-than-ideal goals are more attainable.

With conservatism, the connections with psychology are more complicated, but it seems clear that conservationists, like Outlaw above, are typically committed to retaining racial categorization while eliminating racism and thick racialism.<sup>8</sup> Indeed, to the extent that individuals or groups can reap the (supposed) benefits of racial identification and categorization while avoiding harmful and distorting implications of racism, conservatism enjoys considerable appeal. But is this division of racial categorization from racial evaluation really possible? Here too, there is strong reason to think information about human psychology is relevant to assessing conservationists' proposals.

In sum, both eliminativist and conservationist agendas include, often tacitly, goals of psychological reformation. In particular:

*Eliminativists' Goal:* A reduction of racial categorization in thought and behavior.

*Conservationists' Goal:* The retention of racial categorization together with a rejection of thick racialism and pernicious racial discrimination.

As we will go on to show, the extent to which these psychological aims can be achieved depends on the particular facts of racial cognition.

### **1.3. Normative Proposals, Feasibility, and the Disregard of Psychology**

---

<sup>8</sup> See Mallon 2004, Section 2 for a similar interpretation of Mills (1998).

Costs of normative proposals can be evaluated along various dimensions, including economic, legal, and social ones. We'll continue talking about these costs in terms of a proposal's "feasibility": the feasibility of a proposal is a function of the ease with which its goal can be reached. Neither "feasibility" nor "ease" is terribly precise, but we take the basic idea behind each to be clear enough to get our discussion going. Indeed, since we need some way to talk about different types of conditions that are relevant to assessing a proposal (economic, legal, social, psychological, etc.), insisting on greater precision would hinder the terms' usefulness. Naomi Zack (1998: 16), for example, considers whether completing the project of racial eliminativism is politically feasible given the protection the First Amendment provides to even mistaken thoughts and speech.

One dimension that is rarely considered in these assessments is their *psychological* feasibility, the ease with which eliminativist and conservationist goals can be reached given the psychological facts about human racial cognition. This is puzzling. As we have seen, both eliminativist and conservationist proposals depend in substantial ways on our ability to reform our practices of racial categorization, and these in turn depend in part on the character of the psychology that underwrites these practices. Why, then, is there almost no engagement with the psychology of racial categorization by philosophers of race? The question is not one that can be simply answered by reference to disciplinary boundaries, for philosophical racial theorists typically engage research from a variety of sources, including history, sociology, and anthropology. Yet these

same theorists make almost no effort to engage with psychological research, despite its obvious prima facie relevance.<sup>9</sup>

Rather than speculate on what motivates this *disregard of psychology*, we instead devote our efforts to showing how recent findings about racial cognition are indeed relevant to assessing the feasibility of both eliminativism and conservatism.

Realistically evaluating eliminativist and conservationist goals can only be accomplished if one takes into account some of the more robust and surprising results in current psychology. Below, we will describe two such areas of research and illustrate how they make the disregard of psychology in the normative racial literature untenable. Along the way, we also draw out some more detailed conclusions about how specific psychological results affect the feasibility of competing normative proposals.

## **2. Racial Categorization and Evolutionary Psychology**

Both eliminativists and conservationists want to modify our practices of racial categorization: eliminativists by eliminating them, conservationists by doing away with

---

<sup>9</sup> Paul Taylor (2004) is one of the few philosophers to offer an argument for the disregard of psychology.

Taylor defends his decision not to consider psychological causes of racism on the grounds that he has, little sympathy for the idea that racism derives from the operation of innate, race-specific mechanisms ...it's not clear to me why we need to appeal to some hard-wired mechanism that routinely cranks out organisms that indulge in racist exclusions. We'd still have to explain the peculiar forms of exclusion that the mechanism produces under specific cultural conditions, which seems to me to leave all of the interesting questions unanswered" (37-38).

Taylor's case for the importance of culture in forming particular racialist and racist practices and racism is compelling, but it is the burden of this paper to show that his exclusion of psychological factors is less so.

thick racialism and mitigating the more unsavory evaluations that may accompany the use of racial categories. In this section, we will review recent work in evolutionary cognitive psychology on racial categorization, and show how this work bears on the normative debates.

## **2.1 Racial Categorization and Specialized Cognitive Mechanisms**

Racial categorization presents a puzzle for evolutionary-minded psychologists and anthropologists (Hirschfeld 1996; Gil-White 1999, 2001; Kurzban et al. 2001; for a critical review, see Machery and Faucher 2005a). People classify themselves and others on the basis of physical, putatively racial properties, and seem to assume that these classifications group together people who share important biological properties (and perhaps also important psychological and moral properties as well). However, it is hard to account for this phenomenon with the explanatory resources favored by evolutionary theorists, namely by appeal to something like a “race module”—an evolved cognitive system devoted to race and racial membership. First, it is difficult to identify a selection pressure that would have driven early humans to pay attention to physical properties now associated with race and putative racial differences, like skin color, body shape, etc. Long-distance contacts were probably rare during most of the evolution of human cognition, and our ancestors would have had little direct contact with groups whose members had substantially different physical phenotypes from their own. Moreover, as pointed out in the first section, there is an ontological consensus amongst researchers from a variety of disciplines that whatever else they might be, racial categories do not systematically map onto any biological categories that support robust physical, social,

psychological, and behavioral generalizations.<sup>10</sup> Thus, even if contacts with people with substantially different phenotypical properties had been common during the evolution of humans, the adaptive benefit of classifying others on the basis of these properties would still be unclear.

Thus, rather than postulating a race module on the standard grounds, evolutionary psychologists instead propose that racial categorization is indeed subserved by a module, but that the module in question was initially selected for some other function (not related to race). Evolutionary psychologists theorize that this cognitive system contributes to our social cognition more broadly construed, and is a component of the collection of loosely affiliated cognitive systems that allow humans to navigate the social world. As we shall see below, much of the disagreement amongst evolutionary psychologists is over the nature and proper function of the cognitive system that now underlies racial thinking.

Some background will be useful in understanding these debates between proponents of the evolutionary-cognitive approach itself, for that approach stands in contrast to previous explanations of racial categorization that have been offered in psychology and the social sciences. These include socialization explanations, perceptual saliency explanations, and group prejudice explanations. Psychologists favoring explanations in terms of socialization have assumed that children are either explicitly taught to draw the distinctions used in racial categorization, or that they easily pick them up from the general social environment, even without anyone (e.g. parents, teachers,

---

<sup>10</sup> It is doubtful that racial categories can even be used to express true generalizations about morphological characteristics of members of the same putative race, as there is a tremendous amount of morphological variation within a given recognized race (consider for example Ethiopians and Africans from West-Africa). For discussion, see Diamond 2004.

peers) explicitly instructing them in the use of racial categories (e.g., Allport 1954). In contrast, evolutionary psychologists, while not denying that socialization plays some role, insist that it is not the whole story. Instead, they propose that our tendency to classify people racially is underwritten by an evolved cognitive system, whose development in children is to a large extent independent of teaching and socialization.

Another view at odds with the evolutionary approach holds that racial categorization results from the simple fact that people classify a wide variety of objects (animals, objects, etc.) into categories based on their *perceptually salient* features. The view just sees racial classification as a special case of this much more general tendency: since color is a salient visual property, skin colors trigger this domain-general categorization system, and as a result, people form and rely on racial categories (e.g., Taylor et al. 1978). In contrast, evolutionary psychologists reject the idea that racial categorization can be explained *merely* by the perceptual saliency of skin color, and they argue that racial categorization results from a cognitive system that has evolved to deal with a specific domain in the social world, rather than with categories or perceptual salience in general.

Finally, some social psychologists maintain that racial categorization and racism are to be accounted for by a general tendency to form *group prejudices* about social groups, be they women, races, or social classes (e.g., Crandall and Eshleman 2003). Evolutionary psychologists reject this idea on the grounds that not all social classifications and prejudices behave the same. They hold that not all classifications and prejudices are produced by the same cognitive system, and conclude that racial cognition should be distinguished from other forms of group related cognition.

Evolutionary psychologists offer a variety of considerations in support of their distinctive approach to racial categorization. Though they differ on the details, each of the evolutionarily informed positions we will consider sees racial categorization as a by-product of a fairly specialized cognitive system that evolved to deal with some specific aspect of the social environment. Before getting to the differences between the three positions within the evolutionary-cognitive camp, however, we will review five lines of argument that undermine the socialization, perceptual saliency, and group prejudice explanations just described.

First, and most controversially, evolutionary psychologists hold that people in *many* cultures and historical epochs have relied on skin color and other bodily features to classify their fellows, and have further believed that such classifications also group together people who share underlying biological commonalities. This is controversial because many social constructionist social scientists argue instead that racial categorization is the result of specific historical, political, or social circumstances in the recent history of Europe (see, e.g., Omi and Winant 1994; Fredrikson 2002). *Pace* social constructionists, however, there is evidence that across cultures and historical epochs—e.g., in Classical Greece and in the Roman Empire (Isaac 2004)—people have relied on classifications that are similar to modern racial categories in two central respects. First, these classifications are supposed to be based on phenotypic properties: members are supposed to belong to the same racial category because they share some phenotypic, i.e. morphological or behavioral, properties. Second, people assume or act as if racial categories map onto biological categories: members who share the relevant phenotypic properties are assumed to share some important and distinctive set of underlying

biological properties as well. This is not to deny that racial categorization varies across cultures and times in many respects, but rather to stress that these core elements of racial categorization are not a merely parochial cultural phenomenon.<sup>11</sup>

The presence of these common themes across different cultures is just what an evolutionary psychologist would expect, since evolutionary psychologists view racial cognition as a byproduct of a cognitive system shared by all normally developing humans. In contrast, because socialization accounts cannot explain why these core elements should recur across times and cultures, they are at best incomplete.

Additionally, despite having such beliefs about racial properties at an early age (see below), children do not acquire the tendency to classify racially from their familial environments. If children were explicitly taught by their parents, or if they merely picked up the classifications their parents used even without being explicitly instructed in their use, one would expect children's beliefs about races to be similar to their parents' beliefs. However, this is not the case (Branch and Newcombe 1986; Aboud and Doyle 1996). This dissociation between parents and their children constitutes a second type of evidence against socialization explanations of the disposition to categorize racially.<sup>12</sup>

---

<sup>11</sup> Further undermining the social constructionist view is that its proponents fail to agree on where, when and why racial categorization appeared. Some locate it at the end of the Middle Ages (Fredrickson 2002), others with the development of scientific biological classifications by Linnaeus and Blumenbach in the 18<sup>th</sup> century (Banton 1978), while still others hold European social ideology from the end of the 19<sup>th</sup> century ultimately responsible (Guillaumin 1980).

<sup>12</sup> Admittedly, the evidence discussed in this paragraph does not undermine every variant of the view that children are socialized into classifying racially. For instance, if children were taught to classify racially by their peers, rather than by their parents, the dissociation between their own beliefs and their parents' beliefs

Third, explanations of racial cognition that rely on perceptual saliency take for granted one of the very things they are supposed to be explaining, namely why people classify each other on the basis of phenotypic properties like skin color. Color is not always intrinsically salient, or an important feature for categorization purposes. For example, we often do not pay attention to the color of artifacts, and, when we do happen to take their color into account, we rarely treat it as a property that is important for classificatory purposes (see, e.g., Brown 1990; Keil et al. 1998). When children are trained to use a new tool of a particular color, they do not thereby show a preference for similar tools of the same color, but rather show a tendency to use tools that have a similar shape. Thus, in contrast to features such as their shape or rigidity, children do not treat color as an important property of tools or tool identity (Brown 1990). Examples like these undermine the tacit assumption that colors are salient and important for classification in general. Hence, in the case of perceptual saliency explanations of racial classification, the saliency and importance of *skin* color needs to be explained, not assumed.

Fourth, social psychologists' emphasis on group prejudice is unable to account for the differences between different types of social classification and the different types of prejudices associated with each. Stereotypes about social groups vary substantially from one type of group to the next. To take only one example, stereotypes about political groups, such as liberals and Republicans, do not seem to include the idea that these groups are biological kinds (Haslam et al. 2000). Races, on the other hand, *are* thought

---

would not be problematic. Children may also just pick up the tendency to classify racially from their peers or from the broader social environment (without being instructed to do so).

of as biological kinds (for some cross-cultural empirical evidence, see Machery and Faucher ms). If all prejudicial stereotypes were produced by a unique cognitive system, or were driven by a single, general tendency to form stereotypes about social groups—we should not find such differences.

Fifth and finally, Lawrence Hirschfeld has provided an important body of experimental evidence that is *prima facie* inconsistent with the non-evolutionary explanations of racial categorization considered above, but that is congenial to the evolutionary approach (Hirschfeld 1996). Hirschfeld amasses some striking evidence that three- to seven year-old preschoolers treat skin color differently from other properties. Unlike properties like body shape, for instance, preschoolers expect skin color to be constant over a lifetime and to be transmitted across generations. By contrast, they believe that body shape can change across a lifetime and is not necessarily transmitted across generations (Hirschfeld 1996, 97-101). These beliefs about racial properties reflect a kind of intuitive *essentialism*: racial properties are viewed as stable (racial properties do not change during one's lifetime), intrinsic (racial properties are thought to be caused by one's inner nature), innate (the development of racial properties does not depend much on one's rearing environment), and inherited (parents transmit their racial properties to their children). This sort of essentialism is also characteristic of children's and adults' folk biological reasoning (Gelman and Wellman 1991). Because it is plausible that not all prejudices involve this form of essentialism, this makes up

another form of evidence against the group prejudice explanation of racial categorization.<sup>13</sup>

Hirschfeld also provides some evidence that three- and four-year-old preschoolers pay attention to people's race when this information is presented verbally, but not when it is presented visually. On the one hand, when they are told a story involving various protagonists, children remember the race of these protagonists, even when they are not prompted to pay attention to it. However, when the story is presented by means of drawings, instead of verbally, children do not remember the race of the protagonists (Hirschfeld 1996, chap. 6). This raises obvious problems for the view that intuitive racial categorization can be completely accounted for by appeal to the perceptual saliency of skin color alone. Indeed, while Hirschfeld's experiments are not the final word on racial categorization, it is striking that his results would not be predicted by *any* of the three alternative approaches considered above.

In brief, evidence suggests the following. Racial categorization develops early and reliably across cultures; it does not depend entirely on social learning; it is, in some respects, similar to commonsense biological classification. Thus, racial categorization seems to be neither the product of socialization alone nor of the perceptual saliency of skin color alone. It does not appear to result from a general tendency toward group prejudice, either. Rather, this body of evidence is best explained by the hypothesis that racial categorization results from a specialized, species-typical cognitive system that,

---

<sup>13</sup> *Some* other kinds of stereotypes, such as sexist stereotypes, also involve some form of essentialism (e.g., Haslam et al. 2000). However, what matters for the present argument is the fact that not *all* stereotypes involve some form of essentialism.

even if it did not initially evolve to deal with racial categorization, has been recruited for this purpose.

Evolutionary psychologists also infer a few more specific properties of the system underlying racial thought. Since the operation of the cognitive system is constant across cultures and shielded from the influence of teaching, it is thought to be *canalized*: roughly speaking, a trait is environmentally canalized to the extent that its development is the same across different environments and environmental variables.<sup>14</sup> Given the specific properties of this capacity, namely the tendency to classify into races and the typical beliefs that accompany racial categorization, it appears to be driven by a cognitive system that is distinct from whatever cognitive system underlies stereotypes about other social categories. Finally, because it is species-typical, environmentally canalized, and functionally complex, this cognitive system is plausibly thought to be the product of evolution by natural selection.<sup>15</sup>

It is important to point out up front that without further argument, such an evolutionary account of racial categorization in no way implies that racial categorization cannot be eliminated or modified. Consider the human taste for sweetness, which is also arguably the product of evolution by natural selection. It too develops early, reliably, and cross-culturally. However, during development, several factors determine whether or not

---

<sup>14</sup> For a more nuanced discussion of the notion of canalization, see Griffiths and Machery (2008).

<sup>15</sup> It is worth emphasizing that the evolutionary psychological approach does not imply that the evolved cognitive system is the *unique* cause of racial categorization. For instance, Machery and Faucher (2005b) have proposed that people's disposition to classify racially results from the interaction of an evolved cognitive system and some form of social learning, which involves a disposition to imitate one's prestigious peers and a disposition to imitate the majority of one's peers (conformism).

and how much people will be attracted to sweet foods (Rozin 1982). Thus, although it is a canalized product of natural selection, a person's taste for sweetness is not inevitable or completely impervious to modification. Analogously, racial categorization may thus result from an evolved cognitive system without being inevitable or unalterable.

Understanding the possibilities for eliminating or modifying racial categorization, however, and discovering the most effective means of doing either, will depend on the specific empirical details of its development and operation.

## **2.2 Controversies within Evolutionary Psychology**

Against this backdrop of broad theoretic agreement, disputes have emerged about the specific character of our capacity to make racial classifications. Hirschfeld (1996, 2001), Kurzban and colleagues (2001), and Francisco Gil-White (1999, 2001) have proposed three different accounts of the cognitive system that is assumed to underlie racial categorization. The dust has not settled yet, but the resolution of their disagreements may have an impact upon the debate between eliminativism and conservatism. In what follows, we briefly review each of these three accounts.

First, according to Hirschfeld (1996, 2001), racial categorization results from the interaction of an innate, evolved capacity for *folk sociological* thinking, on the one hand, and the specific social structure in which it is operating, on the other. The evolved function of the posited folk sociological mechanism is to identify the social groups in the social environment. Given the importance of social life during the evolution of human beings (e.g., Dunbar 2003), the ability to map the social world was most likely selected for by evolutionary pressures. According to Hirschfeld, an important aspect of this hypothesized cognitive system is that it essentializes whatever groups are salient in a

given social environment: membership in these groups is associated with a set of immutable properties thought to be caused by some essence common to all group members. When societies are divided along racial lines, the folk sociological mechanism guides us in the identification and essentialization of these groups. In societies with a different social structure, of course, different social groups will be picked out and essentialized. In India, for instance, castes rather than races are the salient social groups, and Hirschfeld's view predicts that in such a social environment, Indians' folk psychological system will essentialize castes (for consistent evidence, see Mahalingam 2003).

Kurzban, Tooby and Cosmides (2001) offer a second account. Instead of positing a folk sociological mechanism that picks out the salient social groups in a given social environment, as Hirschfeld (2001) does, they argue that racial categorization results from a cognitive system whose function is to track *coalitions* (that is, groups of individuals that cooperate with each other) in a given social environment. Kurzban and colleagues assume that races are coalitions in many modern settings, including contemporary American society; since the posited cognitive system tracks coalitions in the social environment, it picks out races in those modern societies.

To support this claim, they provide some intriguing evidence that adults' encoding of skin color and racial membership is influenced by whether racial membership is a relevant cue to coalitional membership. In their experiment, participants were shown pictures of the members of two basketball teams, where each team is composed of some Black and some White players. Participants were also given a fictional verbal exchange between members of the teams. In the next stage of the

experiment, participants were presented with individual sentences from the exchange, and asked to remember who uttered them. The experimenters then looked at the mistakes made by participants, and checked whether, when they were in error about who uttered a sentence, they mistakenly ascribed it to a basketball player of the same *race*, or to one on the same *team*. The resulting patterns of mistaken ascriptions were taken to indicate how participants classified the basketball players. For instance, if participants had categorized the players involved in the verbal dispute according to race rather than team, then when they made mistakes, they should have been more likely to ascribe a statement made by a White player to another White player than to a Black player.

The results of this experiment were along the lines that Kurzban and his colleagues expected. When coalitional membership (viz. membership in each basketball team) was not emphasized, participants implicitly categorized the individuals involved in the verbal exchange according to race. However, when coalitional membership was emphasized—by giving a distinctively colored jersey to the members of each mixed race team—participants appeared to rely much less on race. Kurzban and colleagues concluded that in the absence of any obvious indicators of coalitional boundaries, racial membership is often taken to be a cue to coalitional membership. This hypothesis explains why, when other indications of coalitional membership are made particularly evident or social environments make coalitional boundaries more salient, people are less prone to classify into races. Based on this conclusion, Kurzban and colleagues further suggest that if skin color were not a reliable cue to coalition membership—if, for instance, the social environment were structured differently—people would tend to classify much less on the basis of skin color.

The third account is offered by Gil-White (1999, 2001), who argues that evolution has selected for an *ethnic cognitive system*, that is, for a cognitive system whose evolved function is to identify ethnic groups. In brief, at some point during the evolution of our species (around 100,000 years ago), our ancestors lived in groups called “ethnies,” which were made up of (at least) several hundred or thousand culturally homogenous members. Those ancestors displayed their membership to the group by means of specific ethnic markers, e.g. clothes, body paintings, etc. Gil-White maintains that it was important for our ancestors to map this dimension of the social world and argues that folk biology—the set of commonsense beliefs about animals and biological kinds together with the cognitive systems responsible for classifying and reasoning about animals and biological kinds—was recruited or “exapted” for this purpose (for further detail, see Gil-White 2001).<sup>16</sup> As a result, we have evolved to pay attention to possible ethnic markers and to classify social actors on their basis. Moreover, because the folk biological system essentializes the entities it classifies, we now tend to essentialize the groups we discern on the basis of these ethnic markers. Finally, according to Gil-White, racial categorization can be driven by this cognitive system, because skin color and other racial properties (such as body type) are often taken to be ethnic markers. Because of this, races can be *mistaken* for ethnies by the ethnic cognitive system, despite the fact that they are, in general, *not* ethnies.

To summarize, controversies remain even among those who agree on the basic evolutionary-cognitive approach. Particularly, disagreements center around details of

---

<sup>16</sup> A trait is said to be exapted when it is used for something different than for what it was originally selected.

how the cognitive system believed to now underlie racial categorization is structured, and what it initially evolved to do—track salient social groups, track coalitions, or track ethnies. The three accounts also suggest different reasons why skin color triggers this cognitive system.

### **2.3. Consequences for the Debate between Eliminativists and Conservationists**

While interesting in its own right, the research on racial categorization in evolutionary psychology shows that there are some specific obstacles to the feasibility of eliminativism and conservatism that have been ignored by race theorists. To begin, each of these three accounts of racial cognition leads to a similar conclusion about eliminativism: any eliminativist proposal is committed not just to a substantial amount of social reform, but, in light of the constraints imposed by the psychology of racial categorization, to social reform of a fairly specific sort. This should feature in any serious cost/benefit analysis for or against eliminativism. Consider Hirschfeld's account: during development, the cognitive system that underlies racialism is triggered by the use of race terms ("Black", "White", "Hispanic"... ) by parents, peers, etc., when parents, peers, etc., refer to social groups or characterize individuals. Children rely on such terms to identify the important social groups in their social environment, and they essentialize such groups. Race terms are mapped onto specific visual cues (skin color, body shape...) later in development (Hirschfeld 1996, 136). Obviously, this account leaves many aspects of the development of racial categorization unspecified. However, it suggests that the feasibility of eliminating racial categories in part turns on the importance of races in people's social environment, and perhaps the prominence of racial terms in their

vocabulary. If races are socially important, people will refer to them, and children are likely to develop a tendency to classify racially.

Kurzban, Tooby and Cosmides's hypothesis leads to a similar conclusion. They propose, remember, that the saliency of skin color depends on the coalitional status of races. People pay attention to races because races act as coalitions in many modern societies. Thus, if races continue to act – or seem to act – as coalitions, achieving the ideal of race blindness will be hindered by the fact that putative racial properties like skin color shared by putative coalition members will continue to be salient to our evolved coalitional cognitive system. Remarkably, however, Kurzban and colleagues conclude their (2001) article remarking on “how easy it was diminish the importance of race by manipulating coalition” and suggesting that “the prospects for reducing or even eliminating the widespread tendency to categorize persons by race may be very good indeed” (15391). We are skeptical of this conclusion. If they are right, the existence of racial categorization is linked to the existence of racially-based coalitions. These coalitions are reinforced by the economic and social divisions of contemporary societies, which are not themselves easily alterable. The prospects for eliminating racial categorization, on this story, are tied to the prospects for extensive economic and social reform, and may require putting an end to the sorts of economic, social, and even geographic segregation that continues to separate racial groups.

Kurzban et al.'s hypothesis also places interesting constraints on the type of programs that ought to be used to promote eliminativism. For example, programs where Blacks help other Blacks (e.g., programs which assign junior racial minority professionals to a senior minority mentor of the same race), for example, could tend to

reinforce racial categories, if Kurzban et al. are right. On the other hand, programs in which members of *other* races help Blacks (e.g., a classic affirmative action program in a predominantly White company) might not trigger coalitional thinking.

Although leading to a slightly different conclusion, Gil-White's views also suggest that eliminativism is committed to substantial and specific social reforms. According to him, as we saw, skin color and other phenotypic properties are often taken to be ethnic markers, that is, physical cues that indicate membership in ethnies. Nowadays, in most societies, social groups differ in many respects from the paleoanthropological ethnies in response to which ethnic cognition is supposed to have evolved. Nonetheless, like paleoanthropological ethnies, some modern groups may have substantial cultural homogeneity, in the sense that members of these groups endorse similar behavioral norms, and identify each other by similar markers. If for some historical reason, racial distinctions in a given society map onto such groups, the posited ethnic cognitive system will be triggered not only by skin color and other phenotypic properties, but also by other cues (names, accents, behaviors...). Arguably, this is the case of Blacks in contemporary America.<sup>17</sup> If Gil-White's account is correct, eliminativism might require modifying the cultural structure of society – weakening perceived cultural differences between racial groups (such as Blacks and Whites in the United States). Given that such cultural differences are sometimes claimed to be

---

<sup>17</sup> Although this was not the case when African slaves arrived in the US. They came from different cultures in Africa.

constitutive of individuals' identities, this is an important and potentially controversial cost for eliminativism.<sup>18</sup>

We also note that the reforms suggested by Gil-White's account are of a different sort than the changes that would be required if Kurzban and colleagues were right. Kurzban and colleagues' account of the nature of racial categorization suggests that to eliminate the tendency to classify racially, one should prevent the development of preferential cooperative links between members of the same race, and one should undermine these links if they already exist—that is, one should discourage Hispanics from preferentially helping Hispanics, Blacks from preferentially helping Blacks, and so on. By contrast, Gil-White's account suggests that to eliminate the tendency to classify racially, one should prevent members of the same race from developing shared cultural norms, or one should undermine such norms if they exist—that is, one should discourage Hispanics (or Asians, or Blacks, or Whites) from having a shared and distinctive accent, shared and distinctive behavioral norms, and so on.

Note that the need for such specific social reforms may not be an inescapable difficulty for eliminativism. Eliminativists are well aware that the thrust of their position is ambitious and calls for significant social change. Examples of reform with regards to race are not a mere or even distant possibilities, either: they are evident in actual societies, for instance in the form of affirmative action, school integration, and voting reform in American society. And, as we noted in Section 1, eliminativism can come in

---

<sup>18</sup> Of course, this cost is already explicitly considered in discussions of the value of racial identity in social theory (e.g. Outlaw 1995, 1996; Appiah 1996). What empirical models bring to the discussion are theories and evidence that bear on the question of whether culture and racial identity are, in fact, closely linked in folk racial thinking.

different strengths, or be targeted on different social domains. Nevertheless, it remains the case that evaluations of social reforms should include an assessment of the psychological feasibility of eliminativist proposals.

Conservationism, on the other hand, may not seem as affected by these consequences, since conservationists want to preserve racial distinctions. They do not have to change the cues that trigger the cognitive system that underlies racial categorization, and thus, do not have to reform the social or cultural structure of our societies. Additionally, conservativists do not appear committed to anything that may entail the weakening of cultural or racial identities. Conversely, the evolutionary psychology considered in this section suggests that eliminativists *are* committed to such projects.

Nevertheless, the feasibility of conservationist goals will also be directly affected by which psychological view turns out to be correct. Hirschfeld's and Gil-White's accounts tentatively suggest that racial categorization and essentialism—i.e., the belief that racial groups are biological groups, whose members share an underlying essence that explains their shared physical, behavioral and moral properties—are the product of the same cognitive system. Details and evidence are scarce at this point: particularly relevant is the fact that Hirschfeld does not adduce explicit evidence that moral properties are essentialized. Still, Hirshfeld's and Gil-White's accounts suggest that whenever people categorize racially (because races are salient social groups or because children take skin color and other physical properties to be ethnic markers), they essentialize the groups that are delineated. Thus, according to their accounts conserving racial categorization, while reforming its normative connotations, may be hindered by the nature of the evolved

cognitive system that underlies racial categorization. For example, an attempt to encourage people to adopt a nonessentialized metaphysics for race (of the sort suggested by, e.g. Omi and Winant 1994; Mills 1998; or Taylor 2004) may be defeated or at least complicated by the very structure of the system underlying racial cognition. Of course, none of this implies a nonessentialist conservatism is impossible. For, as illustrated above with the example of our taste for sweetness, the effects of an evolved and canalized cognitive system are not inevitable. But understanding the means and chances of achieving a nonessentialist conservatism in light of this psychological research is certainly an important factor in the cost/benefit analysis of any specific conservationist proposal.

The situation would be very different if Kurzban, Tooby and Cosmides's account turned out to be correct. For they propose that essentialism, on the one hand, and the salience of racial physical properties, on the other, stem from two different cognitive systems (Cosmides et al. 2003). Again, on this view, racial categorization is the product of human coalitional system. Essentialism comes from our folk biology. If this is right, the nature of human racial psychology does not prevent in any way the dissociation between racial categorization and its essentialist implications.

To summarize, recent evidence supports the idea that among the causes of racial categorization, one finds an evolved, canalized, and species-typical cognitive system. If true, these evolutionary hypotheses would reveal that there are some definite and significant problems for eliminativists and for conservationists alike. The three views considered here reinforce the thought that eliminativism is committed to some form of social reform. Moreover, as we saw, each view suggests that a distinct sort of social

reform is needed for eliminativism, and each raises specific and difficult normative questions about the way in which the cultural or coalitional unity of a group would have to be compromised in order to eliminate racial categorization. Additionally, Hirschfeld's and Gil-White's views suggest that dissociating racial categorization and essentialism, as is proposed by conservationists, may be hindered by the nature of the cognitive system that underlies racial categorization, while Kurzban and colleagues' view is congenial with such proposals. In either case, neglecting psychology amounts to neglecting specific obstacles that need to be addressed in order for eliminativist or conservationist proposals for reform to be viable.

### **3. Racial Evaluation and Implicit Social Cognition**

Racial categorization looks to raise problems both for eliminativists and conservationists. One might be tempted, however, to think those results weigh especially heavily against eliminativism, and tilt the balance of considerations toward conservationism. In this section, we suggest that the conservationist goal of reducing negative racial evaluation has problems of its own - problems that the disregard of psychology has kept from being addressed.

In social psychology, recent advances in experimental measurement techniques have allowed psychologists to explore the contours of our capacities for racial evaluation with great precision, and a set of unsettling results have emerged. Most relevant of these is a particular phenomenon that has been confirmed repeatedly: people who genuinely profess themselves to be tolerant, unbiased, and free of racial prejudice nonetheless often display signs of implicit racial bias on indirect experimental measures. These methods

were designed to bypass one's explicitly held views, i.e. those available via introspection and self-report, and instead systematically probe the less transparent workings of attitudes, associations, and processes linked to categorization and evaluation. After reviewing the relevant findings, we will go on to assess their implications for the normative debate between eliminativism and conservationism.

### **3.1 Indirect Measures and Implicit Cognition**

Consider how you could find out about someone else's mathematical prowess, or their ability to distinguish the subtleties of red wines. Perhaps the most obvious way would be to simply *ask* that person outright, "How good are you at math? Can you integrate a multi-variable equation?" or "How educated is your wine palette? Can you appreciate the difference between a California merlot and a French cabernet sauvignon?" Alternatively, you might take a more circuitous route, and proceed by giving the person a set of math problems or a wine taste test, and infer their mathematical abilities or wine sophistication from their performance on the respective tests. The first type of strategy depends for its reliability on the sincerity of the person's self-report, the absence of self-deception in their self-assessment, and their ability to introspectively access the relevant information. The second type, though less direct in some ways, has the advantage of bypassing all three of these obstacles.

For similar reasons, indirect strategies have become trusted instruments for investigating many cognitive capacities, and research on implicit social cognition is no

exception. We will call measures that rely on such strategies *indirect measures*.<sup>19</sup>

According to Nosek et al. (2007), most indirect measures are:

[M]easurement methods that (a) avoid requiring introspective access, (b) decrease the mental control available to produce the response, (c) reduce the role of conscious intention, and (d) reduce the role of self-reflective, deliberative processes. (2005, p. 4)<sup>20</sup>

This description isn't definitive, but it gets across the flavor of indirect measures, the most prominent of which will be described in more detail below.

First, though, some terminological stipulations will lend clarity to the discussion. The term 'implicit' is a source of potential confusion in this literature, as it is often applied to both the cognitive processes as well as the experimental measures used to probe them, and is treated as loosely synonymous with 'automatic', 'unconscious', and various other terms (Greenwald and Banaji 1995, Greenwald et al. 1998, Cunningham et al. 2001, Eberhardt 2005, Nosek et al. 2007). In what follows, we will use 'indirect' to describe measurement techniques, namely those that do not rely on introspection or self report, and reserve 'implicit' only for mental entities being measured. Moreover, we will follow Banaji et al. (2001) and use 'implicit' to describe those processes or mechanisms operating outside the subject's conscious awareness, and 'automatic' to denote those that operate without the subject's conscious control.

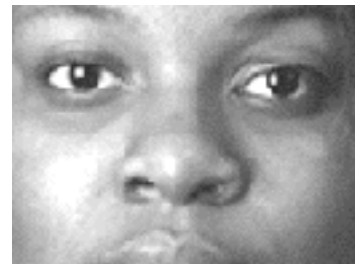
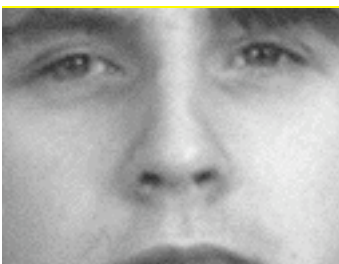
---

<sup>19</sup> Phelps et al. 2000 and Phelps et al. 2003 use this term to distinguish them from 'direct' measures that use techniques like interviews or questionnaires that rely on verbal and written self-report.

<sup>20</sup> Thus characterized, indirect testing is not a particularly recent development to psychology (see, for instance, Stroop 1935).

### *The Implicit Association Test (IAT)*

The IAT has been the most widely used indirect measure, and has been consequently subjected to the most scrutiny.<sup>21</sup> It was initially conceived of as “a method for indirectly measuring the strengths of associations,” designed to help “reveal associative information that people were either unwilling or unable to report” (Nosek et al. 2007, page 269). At its heart, the IAT is a sorting task. Most instances of the IAT involve four distinct categories, usually divided into two pairs of dichotomous categories. For instance, an IAT might involve the category pairs *Black* and *White* (called “target concepts”), on the one hand, and *good* and *bad* (called “attribute dimensions”) on the other. In one common case, the exemplars of the categories *Black* and *White* are pictures of Black and White faces, while exemplars of the other two categories are individual words, such as “wonderful”, “glorious”, and “joy” for *good*, “terrible”, “horrible”, and “nasty” for *bad*. During trials, exemplars are displayed one at a time, in random order, in the middle of a computer screen, and participants must sort them as fast as they can.



---

<sup>21</sup> The first presentation of the test itself, along with the initial results gathered using it, can be found in Greenwald et al. 1998. Greenwald and Nosek 2001 and Nosek et al. 2007 both present more recent reviews of research using IATs, as well as assessments of the methodological issues generated by use of the test and interpretation of results. It should also be noted that there are several variants of this basic paradigm (e.g., Cunningham et al. 2001).

Crucial to the logic of the test is the fact that participants are required to sort the exemplars from the *four* categories using only *two* response options. For instance, they are told to press ‘e’ when presented with any exemplar of *good* or any exemplar of *Black*, and press ‘i’ when presented with any exemplar of *bad* or any exemplar of *White*. Equally crucial to the logic of IATs is that they are *multi-stage* tests (often comprised of 5 stages), and the response options (the ‘e’ and ‘i’ keys) are assigned to different categories in different stages. So one stage might require the participant to respond to exemplars of *good* or *Black* with the ‘e’ response option and exemplars of *bad* or *White* with the ‘i’ response option, while the next stage assigns *bad* or *Black* to the ‘e’ response option and *good* or *White* to the ‘i’ response option. Paired categories such as *good* and *bad*, or *Black* and *White*, however, never get assigned to the same response options (each response option is assigned one “target concept” and one “attribute dimension”). When a participant makes a sorting error, it must be corrected as quickly as possible before he or she is allowed to move on to the next exemplar. Precise reaction times are measured by the computer on which the test is being taken, as is correction time and number of errors.<sup>22</sup>

Coarse-grained interpretation of performance is fairly straightforward. Generally speaking, the “logic of the IAT is that this sorting task should be easier when the two concepts that share a response are strongly associated than when they are weakly associated.” More specifically, “ease of sorting can be indexed both by the speed of

---

<sup>22</sup> See the citations in footnote 21 for a much more detailed and technically precise discussion of this technique. In order to get the feel of the test, however, one is much better off simply taking one; different versions of it are available at <https://implicit.harvard.edu/implicit/demo/>.

responding (faster indicating stronger associations) and the frequency of errors (fewer errors indicating stronger association)” (Nosek et al. 2007, page 270). The idea can be illustrated with our example case. If a participant is able to sort exemplars faster and more accurately when *good* and *White* share a response option than when *good* and *Black* share a response option, this fact is interpreted as an indirect measure of a stronger association between the two categories *good* and *White*, and hence an implicit preference for White, or, conversely, an implicit bias against Black. This is called the IAT effect. The size of the relative preference or bias is indicated by the disparity between the speed and accuracy of responses to the same stimuli using different response option pairings. Finally, the associations thus revealed are taken to be indicative of processes that function implicitly and automatically, because the responses must be made quickly, and thus without benefit of introspection or the potentially moderating influence of deliberation and conscious intention. While the details of the method can seem Byzantine, the basic idea behind the test remains rather simple: stronger associations between items will allow them to be grouped together more quickly and accurately; the sophisticated set up and computerization just allow fine-grained measurement of that speed and accuracy.

### *Modern Racism Scale (MRS)*

By way of contrast with indirect measures like the IAT, the MRS is a direct measure of racial attitudes, one that is often used in conjunction with the indirect measures. This is standard self-report questionnaire that was designed to probe for racial biases and prejudices (McConahay 1986). It poses statements explicitly about racial issues (e.g. “Over the past few years, Blacks have gotten more economically than they

deserve”; “It is easy to understand the anger of Black people in America”; “Blacks are getting too demanding in their push for equal rights”), and allows participants to react to each statement by selecting, at their leisure, one of the responses, which range from Strongly Disagree to Strongly Agree.

The use of direct measures *together* with indirect measures is important because it is the conjunction of the two that supports the inference to not just automatic but *implicit* processes and biases in the sense discussed earlier. Recall that implicit processes operate outside the introspective access and awareness of participants, while automatic processes are those that operate beyond conscious control. There is much overlap, but these two terms are not completely coextensive; disgust responses, for example, may be automatic, but they are rarely implicit. That participants can exhibit biases on indirect measures, despite the fact that they report having no such biases when asked directly, lends support to the conclusion that what manifests in the indirect tests is indeed the result of processes that are unavailable to introspection and self-report.

## **3.2 Evidence of Biases and their Effects**

### **3.2.1 Implicit Racial Bias**

These types of indirect measures have been used to probe and reveal a wide variety of implicit biases, including age biases (e.g. Levy and Banaji 2002), gender biases (e.g. Lemm and Banaji 1999), sexuality biases (e.g. Banse et al. 2001), weight biases (e.g. Schwartz et al. 2006), as well as religious and disability biases (see Lane et al. 2007 for a review). Some of the first and most consistently confirmed findings yielded by

these tests, however, center on racial biases.<sup>23</sup> Participants who profess tolerant or anti-racist views on direct tests often reveal racial biases on indirect tests. This result is quite robust; similar dissociations have been found using a wide variety of other indirect measures, including evaluative priming (Cunningham et al. 2001, Devine et al. 2002), the startle eyeblink test (Phelps et al. 2000, Amodio et al. 2003), and EMG measures (Vanman et al. 1997). In other words, it is psychologically possible to be, and many Americans actually are, *explicitly racially unbiased while being implicitly racially biased*.<sup>24</sup> Moreover, not only is it possible for two sets of opposing racial evaluations to coexist within a single agent, but, as we will see, when it comes to altering and controlling them, the different types of biases may be responsive to quite different methods.

---

<sup>23</sup> The first paper to showcase the IAT included the results from 3 separate experiments, one of which was a test for implicit racial biases in white American undergraduates (Greenwald et al. 1998). Results exhibited a now familiar, but still disturbing, pattern: while most (19 of 26) of the participants explicitly endorsed an egalitarian, or even pro-Black, position on the direct measures (including the MRS), all but one exhibited an IAT effect indicating implicit white preference. This was the first study using the IAT to investigate this phenomenon, but previous work using less sophisticated methods had revealed similar results (e.g. Devine 1989, Greenwald and Banaji 1995, Fazio et al. 1995). Since the initial 1998 paper, similar results from IATs have been reported so often and found so reliably that they have become a commonplace (Kim and Greenwald 1998, Banaji 2001, Ottaway et al. 2001).

<sup>24</sup> While the fact that implicit and explicit racial biases can be dissociated is no longer a subject of much controversy, the relationship between the two is still very much in question. While early discussions stressed the complete independence of subjects' performances on direct and indirect tasks (Greenwald et al. 1998), follow up work has shown that the two can be involved in complicated correlations (Greenwald et al. 2003, Nosek et al. 2007).

### 3.2.2 Implicit Racial Bias and Behavior

Perhaps a natural question to ask before going any farther is whether or not the biases revealed by indirect measurement techniques have any influence on judgments or ever lead to any actual prejudicial behavior, especially in real world situations.

Obviously, the question is important for a variety of reasons, not least of which is assessing the feasibility of revisionist proposals offered by philosophers of race. Racial theorists (and others) skeptical of the relevance of this psychological literature might be inclined to simply dismiss it on the grounds that tests like the IAT measure mere linguistic associations or inert mental representations that people neither endorse nor act upon in real world scenarios (see, for instance, Gehring et al. 2003). Others, who grant that the results of indirect tests (which usually turn on differences that are a matter of milliseconds) are of legitimate theoretic interest to psychologists<sup>25</sup>, might still remain skeptical that implicit biases, whatever they turn out to be, are powerful enough to make any practical difference in day to day human affairs.

We do not think that such skepticism is justified. First, we are impressed by mounting evidence that race and racial bias can still have measurable and important effects in real world situations. In a field study by Bertrand and Mullainathan (2003), researchers responded to help-wanted ads in Boston and Chicago newspapers with a variety of fabricated resumes. Each resume was constructed around either a very Black sounding name (e. g., “Lakisha Washington” or “Jamal Jones”) or a very White sounding

---

<sup>25</sup> For instance, some psychologists see problems with the quick inference from IAT results to the attribution of implicit prejudice (Blanton & Jaccard 2008, Arkes & Tetlock 2004).

name (e.g. “Emily Walsh” or “Greg Baker”). When the resumes were sent out to potential employers, those bearing White names received an astonishing 50 percent more callbacks for interviews. Moreover, those resumes with both White names and more qualified credential received 30 percent more callbacks, whereas those highly qualified Black resumes received a much smaller increase. The numbers involved are impressive, and the amount of discrimination was fairly consistent across occupations and industries; in Bertrand and Mullainathan’s own words:

“In total, we respond to over 1300 employment ads in the sales, administrative support, clerical and customer services job categories and send nearly 5000 resumes. The ads we respond to cover a large spectrum of job quality, from cashier work at retail establishments and clerical work in a mailroom to office and sales management positions.”

Interestingly, employers who explicitly listed “Equal Opportunity Employer” in their ad were found to discriminate as much as other employers.

Similar evidence of race and racial bias influencing real world situations comes from a recent statistical analysis of officiating in NBA games, which claims to find evidence of an “opposite race bias” (Price and Wolfers, manuscript). The study, which took into account data from the 12 seasons from 1991-2003, found evidence that White referees called slightly but significantly more fouls on Black players than White players, as well as evidence of the converse: Black referees called slightly but significantly more fouls on White players than on Black players.

The racial composition of teams and refereeing crews was revealed to have slight but systematic influence on other statistics as well, including players’ scoring, assists,

steals, and turnovers. The study found that players experience a decrease in scoring, assists and steals and an increase in turnovers when playing before officiating crews primarily composed of members of the opposite race. (For example, a Black player's performance will fall off slightly when at least two of the three referees are White. For the purposes of the study all referees and players were classified as either Black or not Black.) These findings are especially surprising considering the fact that referees are subject to constant and intense scrutiny by the NBA itself, so much so that they have repeatedly been called "the most ranked, rated, reviewed, statistically analyzed and mentored group of employees of any company in any place in the world" by commissioner David Stern (e.g. Schwartz and Rashbaum 2007).

While neither the IAT, nor any other indirect, controlled experimental technique was given to participants in either the NBA or the resume studies, explanations that invoke implicit biases look increasingly plausible in both cases. Indeed, the sorts of real world findings coming from these sorts of statistical analyses and field studies, on the one hand, and the types of automatic and implicit mental processes revealed by the likes of the IAT, on the other, appear to compliment each other quite nicely. Explicit racism on the part of NBA referees or the employees responsible for surveying resumes and deciding whom to contact for job interviews may account for some fraction of the results, but given the conditions in which the respective groups perform their jobs, we are skeptical that appeal to explicit racism alone can explain all of the results. Especially in the heat of an NBA game, referees must make split second judgments in high-pressure situations. These are exactly the type of situations where people's behaviors are likely to be influenced by automatic processes.

Moreover, researchers have begun to push beyond such plausible speculation and explicitly link indirect measures with behavior in controlled settings. These studies further confirm that when participants have to make instantaneous decisions and take quick action, racial biases affect what they do. Payne (2006) reviews a large body of evidence concerning participants who are asked to make snap discriminations between guns and a variety of harmless objects. Participants, both White and Black, are more apt to misidentify a harmless object as a gun if they are first shown a picture of a Black, rather than a picture of a White. This effect has become known as the “weapon bias.” Similar results are found with participants who explicitly try to avoid racial biases. Moreover, presence of a weapon bias correlates with performance on the racial IAT (Payne 2005). This suggests that implicit racial biases may indeed lie behind the weapon bias. (For more discussion and a wider range of cases that link implicit biases of all sorts to behavior, see Greenwald et al. in press.)

The real world relevance of such findings is increasingly difficult to deny. It could help explain familiar anecdotes of sincerely egalitarian people who are surprised when they are called out for racist behavior or biased decision making, especially when such accusations turn out to be legitimate. Another, more concrete example is provided by the highly publicized death of Amadou Diallo in 1999. He was shot and killed by New York police officers who thought he was drawing a gun, when in actuality he was just reaching for his wallet.

### **3.2.3 Mitigating the Effects of Implicit Racial Bias**

In addition to its direct real world relevance, this body of psychological research has implications relevant to normative racial theorists. Before discussing those implications, however, we wish to call attention to a relevant offshoot of this literature that investigates whether and how implicit biases can be brought under control, and whether their expression in behavior and judgment can be mitigated.<sup>26</sup> Preliminary evidence suggests that implicit biases and the downstream effects they typically give rise to can indeed be manipulated. Research is beginning to shed some light on the effectiveness, and lack thereof, of different methods for bringing them under control. We consider three different methods of mitigating the effects of implicit biases: manipulating the immediate environment, self-control, and blocking the development or acquisition of implicit bias.

First, some of these studies suggest that while implicit biases operate beyond the direct conscious control of the participants themselves, they can be rather dramatically influenced by manipulating aspects of a person's immediate *environment*, often their social environment. Dasgupta and Greenwald (2001) showed participants pictures of admired and disliked Black and White celebrities (Denzel Washington, Tom Hanks, Mike Tyson, Jeffrey Dahmer) and found that exposure to admired Blacks and disliked Whites weakened the pro-White IAT effect. They also found that the weakening of the implicit bias measured immediately after exposure to the pictures was still present 24 hours later, while the subjects' explicit attitudes remained unaffected. Lowery et al. (2001) found that the implicit biases of White Americans (as measured by the IAT)

---

<sup>26</sup> See the special issue of *Journal of Personality and Social Psychology* vol. 81, issue 5, in 2001, for an introductory overview and collection of articles devoted to this topic.

could be lessened merely by having the participants interact with a Black experimenter rather than a White experimenter. Richeson and Ambady (2003) showed situational differences can affect implicit biases: when White female participants were told they were going to engage in a role-playing scenario, either as a superior or a subordinate, immediately after they completed an IAT, those anticipating playing a subordinate role to a Black in a superior role showed less traces of implicit racial bias than those anticipating play a superior role to a Black in a subordinate role.

Other studies investigated the extent to which a participant can obliquely influence their own implicit biases by some form of *self-control*, either by actively suppressing their expression or indirectly affecting the implicit processes themselves. For instance, Blair et al. (2001) found that participants who generate and focus on counter-stereotypic mental imagery of the relevant exemplars can weaken their IAT effects. Richeson et al. (2003) present further brain-imaging and behavioral data suggesting that while so-called “executive” functions (in the right dorsolateral prefrontal cortex) can serve to partially inhibit the expression of racial biases on indirect tests, the act of suppressing them requires effort and (or perhaps in the form of) attention.

A different way to eliminate the pernicious effects of implicit biases might be to nip the problem in the bud, so to speak, and to keep people (young children, for instance) from acquiring or developing them in the first place. Research raises difficulties for this possibility, however. Preliminary evidence suggests that implicit biases are easier to acquire than their explicit counterparts. The same evidence suggests implicit biases are harder to alter once acquired, and are extremely difficult to eliminate. This is given a rather striking experimental demonstration by Gregg et al. (2006). Participants in this

study were told about two imaginary groups of people, the second of which was cast in a negative light in order to induce biases against its members. After given this initial information, however, participants were told that the damning description of the second group was incorrect, the mistaken result of a computer error. Gregg and his colleagues then gave participants both direct and indirect tests, and found that while their explicit biases had disappeared, their implicit biases, as measured by an IAT, remained. Work on acquisition and the development of the capacity for implicit social cognition in general is still in its infancy, but initial forays into the area suggest that the development of the capacity for implicit bias is rapid, independent of explicit teaching, and distinct from the development of explicit biases (see Dunham et al. 2008).

These findings make up the beginning of a promising research program centered not only on implicit racial cognition itself, but how the unwanted influence of implicit biases on judgment and behavior can be mitigated or brought under control. On the currently available evidence, it is not yet clear whether the most effective strategies act on the implicit biases themselves, or on ancillary processes that underlie their expression in behavior or judgments. The bulk of this work does suggest that, at the very least, the expression of implicit biases is not impossible to alter. Indeed, while they are inaccessible via direct introspection and appear not to require – indeed, can even *defy* – deliberation or conscious intention, these studies suggest that implicit biases are amenable to some methods. While blocking their development or acquisition may be an uphill battle, their expression can be restrained via strategic alterations of the social environment and specific forms of self-control.

### **3.3 Consequences for the Debate between Eliminativism and Conservationism**

While it is fascinating in its own right, this body of work in social psychology is clearly relevant to a variety of philosophical issues concerning race.<sup>27</sup> To be forthright, the psychological story is still far from complete, and in a number of ways:

- (a) the extent to which many of the results reported can be generalized from one culture to the next remains uncertain, as does the manner in which those results might be generalized;
- (b) whether and which results can be generalized to racial groups beyond Blacks and Whites within a single culture (to include other putative racial group such as Hispanics, Indians, Asians, etc.) is also uncertain (but see Devos et al. 2007);
- (c) there is little systematic data concerning the ontogenesis of implicit racial biases (but see Baron and Banaji 2006, Dunham et al. 2008);
- (d) a more detailed account of the cognitive architecture underlying these implicit biases is needed, preferably one that can shed light on the admittedly live issue of how and how often the evaluations measured by the indirect tests are also involved in causal processes that lead to actual judgment and action;
- (e) it is currently far from clear whether implicit biases of different types, for instance implicit racial biases, gender biases, age biases, disability biases, etc., all reflect the workings of the same set of cognitive mechanisms;
- (f) more fine-grained and theoretically-motivated distinctions are needed, since the term 'group' used to interpret much of the data is probably too ambiguous to be of much serious use - as alluded to in Section 2, different sorts of groups, for instances

---

<sup>27</sup> For an initial attempt to wrestle with the ethical implications of implicit racial biases, see Kelly and Roedder 2008, Faucher and Machery, forthcoming.

coalitions, ethnies, families, political parties, or even professions may be cognized differently by distinct systems in the human mind.

We list these points not as an indictment or criticism, but by way of emphasizing the richness of the research project, and the breadth of the issues it might eventually be able to shed light on. Moreover, the contours of the emerging picture are already discernible, and they have implications of their own. Since many of those implications crucially involve not just racial categorization but *evaluation*, we will here consider the impact they have on the conservationist position.

We noted at the outset that a typical conservationist position advocates retaining racial categorization while reducing or eliminating the belief that racial groups are biologically distinct, as well as racist evaluations that favor one group over another. In this way, the conservationist position is continuous with familiar social programs in the United States that attempt to diminish or redress racism and its effects while retaining racial categories. (For example, affirmative action is a program for which racial categories are indispensable). At first, proposals along these lines seem both sensible and realistically achievable. Indeed, as has been noted in a number of places (Biernett and Crandall 1999, Shuman et al. 1997, Phelps et al. 2000), the last couple of decades have shown a significant decrease in the expression of explicit racist attitudes, as measured by self-report. While this is surely a sign of progress, the results reported in the previous section suggest that the actual state of affairs is more complicated, and that achieving conservationist goals involves more than the reduction of explicit bias. That it is psychologically possible to be, and that many Americans indeed *are*, explicitly unbiased, but implicitly biased, suggests that maintaining racial categorization while at the same

time purging racial categories of all of their derogatory evaluative baggage is committed to addressing two different families of evaluative states instead of just one. While no one is under the illusion that racism will be easy to eradicate,<sup>28</sup> the work in social psychology can help shed light on the exact nature of the difficulties involved. In turn, by disregarding that work, and the fact that implicit biases appear to exist in many explicitly unbiased people, conservationists are at risk of ignoring some of the obstacles that stand in the way of their own proposals.

We take the empirical research to have established a number of claims. A large body of evidence clearly indicates that implicit racial biases exist, and are fairly prevalent in the population. They are different from, and can coexist with, their explicit counterparts. Statistical analyses like those provided in Price and Wolfer's paper on the NBA, and field studies like those described in Bertrand and Mullainathan's resume paper complement work done in controlled experimental settings, strongly suggesting that implicit biases indeed effect judgment and behavior, even in real world situations. For conservationists, the broadest conclusion to draw from this is that to the extent that implicit biases have not been systematically taken into account, the feasibility of achieving their professed ideals remains largely unknown.

Additionally, explicit prejudices have declined steadily over the last several decades while implicit biases remain prevalent and may be more robust (though we lack similar data tracking the level of implicit bias through same span of years). Whatever has been successful in bringing about the drop off of explicit racial bias does not appear to have eliminated implicit bias. This suggests that not all racial evaluations can be revised

---

<sup>28</sup> E.g. Outlaw 1995, Taylor 2004.

and altered by the same methods. Hence, assessing the feasibility of specific conservationist proposals for dealing with negative racial evaluations should take into account not just implicit biases themselves, but the costs and benefits of implementing the sorts of techniques most likely to effectively deal with them.

Conversationists may take different stances in light of the existence and character of implicit racial biases. On the one hand, they may maintain that the proper ideal to strive for remains the complete eradication of negative racial evaluations, both explicit and implicit alike. If future research vindicates the preliminary results, then once implicit biases are taken into account, achieving such an ideal may be even more difficult than previously thought. For, two ways that immediately come to mind of achieving the conservationist ideal are by blocking the acquisition or development of biases in younger generations, and by eradicating biases in those persons who are already harboring them. Recall, however, that initial findings indicate that implicit biases a) develop quite early, often without benefit of explicit teaching (Dunham et al. 2008), b) implicit biases are easier to acquire than their explicit counterparts, and that c) especially relative to their explicit counterparts, implicit racial biases appear difficult to eradicate (or reverse, i.e. flip from a negative to a positive valence) once acquired. As mentioned above, this is given a striking demonstration in Gregg et al. (2006), where participants had biases induced about a fictional group, only to be later told that the damning information used to induce the biases was incorrect; the participants' explicit biases against the group disappeared, but their implicit counterparts did not. Taking implicit biases into account raises serious challenges for both of the most obvious general strategies for doing away

with implicit evaluations, and these challenges should be reflected in assessing the feasibility and cost associated with specific proposals based on them.

Psychological research might point the way to other less explored options, too. Future research may still help conservationists who remain committed to the ideal of complete eradication of racist evaluation by discovering more effective ways to deal with them at early stages of ontogeny, before they are fully developed or entrenched. Current research may also be mined for inspiration as well. For example, some studies have linked IAT effects with emotions, suggesting that implicit biases are often affect laden (e.g. Phelps et al. 2000, Phelps and Thomas 2003). If this turns out to be the case, emotion-based techniques may provide more effective means by which conservationists can achieve their goals. One interesting possibility emerges from work by Rozin (1997) who describes how *moralization*, which crucially involves emotional elements, has had effects both in the promulgation of vegetarianism and in the decrease of the acceptability of smoking. As such, moralization might be successful in the mitigation and elimination of implicit racial biases as well. Previously developed methods of social influencing that appeal to emotions (and which may therefore fall under Rozin's concept of moralization) might also be successfully applied to implicit racial biases.<sup>29</sup> These might include, for instance, casting racist biases, judgments and behaviors as not just wrong but shameful and viscerally disgusting. More speculatively, other sorts of emotion-based methods of persuasion may be recruited from advertising and marketing or political campaigning.

---

<sup>29</sup> Though such proposals are certainly attractive, there are reasons to be cautious. For instance, Dan Fessler and his colleagues (Fessler et al. 2003) have argued that "moral" vegetarianism may have little to do with disgust-based moralization.

Such methods may more effectively speak to implicit racial biases than straightforward education, rational discussion or careful argumentation.

On the other hand, conservationists impressed by the psychological findings might abandon the idea of complete eradication of both implicit and explicit bias, and instead embrace a more pragmatic goal of eradication of explicit bias, together with some agenda of controlling or mitigating the expression of implicit biases (e.g. see Lengbeyer 2004, who argues for a similar approach). Proposals for achieving this goal may center on the promulgation of techniques that are most effective in suppressing or bring implicit biases under control. Such proposals, of course, need to be formulated in detail before they could be properly assessed, but they might be guided by the sort of research discussed in section 3.2.3, which showed how implicit biases are not immune to certain forms of influence. For example, if future research bears out the preliminary findings that altering the social environment in targeted ways can reduce the expression of implicit biases, then the most effective conservationist proposals to mitigate the expression of racial bias might include suggestions for structuring the social environment in ways that the psychological research suggests is most helpful.

Other proposals may be inspired by the research on self-control. Here the conservationist gets a mixed bag. For, preliminary research suggests that on the one hand individuals are able to suppress the expression of implicit racial biases in judgment and behavior. On the other hand, as indicated by the work of (Richeson et al. (2003), Richeson and Shelton (2003) and Govorun and Payne (2006), effort and attention is required to exert this kind of self control; indeed, Bartholow et al. (2006) have shown that alcohol consumption interferes with the capacity to intentionally control the expression of

these biases. This may be construed as a cost that attaches itself to proposals that center on self-control. For, implementing the widespread and consistent suppression of implicit biases may also require ensuring the vigilance and effort (and perhaps sobriety!) of those individuals who harbor them. Alternatively, future psychological research may help uncover additional techniques that can help enhance the effectiveness of self-control, as Blair and colleagues (2001) found of generating and focusing on counter-stereotypic mental imagery.

The main conclusion of this section is that the psychological work on implicit racial bias is directly relevant to the normative debate over race, and is especially important for conservationists. Individual proposals can be properly assessed only in light of the psychological research, and until implicit biases are systematically taken into account, the feasibility and costs associated with such proposals remains unclear. In addition to facilitating a more realistic assessment of extant proposals, the psychological work can also be a source of inspiration for novel positions and proposals in the conservationist spirit, and can also point the way towards more effective methods for achieving conservationists goals.

#### **4. Conclusion**

Our aim was not to weigh in on one side of the controversy between eliminativism and conservatism, but to point out an assumption apparently made by both sides of the debate, and show it to be untenable. That debate takes place against the backdrop of an acknowledged ontological consensus. United by the shared rejection of a biological basis of race, eliminativists and conservationists have proceeded to take the

fields of biology and genetics to be by and large irrelevant to the normative racial debate. We have asserted that the normative debate takes place against the backdrop of a somewhat analogous, though generally *unacknowledged* consensus that gives rise to the widespread disregard of psychology in that literature. In contrast to the attention paid to anthropological and historical factors, the philosophical literature on race fails to consider whether and how psychological factors could affect the feasibility of the various normative proposals that have been offered. We have argued that this disregard of psychology is unjustified, and have shown how empirical research on racial cognition is directly relevant to the goals held by normative racial theorists, and to the feasibility of the proposals made for achieving them.

## References

- Aboud, F.E. and Doyle, A.B. 1996. "Parental and Peer Influences on Children's Racial Attitudes." *International Journal of Intercultural Relations*, 20: 371-383.
- Allport, G. W. 1954. *The Nature of Prejudice*. Cambridge, MA: Addison-Wesley.
- Appiah, K. A. 1995. "The Uncompleted Argument: Du Bois and the Illusion of Race." In L. A. Bell and D. Blumenfeld (1995). pp. 59-78.
- Appiah, K. A. 1996. "Race, Culture, Identity: Misunderstood Connections." In K. A. Appiah and A. Guttman (eds.), *Color Conscious: The Political Morality of Race*. Princeton, NJ: Princeton University Press.
- Arkes, H. & Tetlock, P. E. (2004) Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?" *Psychological Inquiry*, 15(4), 257- 278.
- Banaji, M. R. 2001. "Implicit Attitudes Can Be Measured." In H. L. Roediger, III, J. S. Nairne, I. Neath, and A. Surprenant (eds.), *The Nature of Remembering: Essays in Honor of Robert G. Crowder*. Washington, DC: American Psychological Association. pp. 117-150.
- Banse,R., Seise,J., and Zerbes,N. 2001. "Implicit Attitudes Toward Homosexuality: Reliability, Validity, and Controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, 48: 145–160.
- Banton, M. 1978. *The Idea of Race*. Boulder, CO: Westview.
- Baron, A. S. and Banaji, M. R. 2006. "The Development of Implicit Attitudes: Evidence of Race Evaluations from Ages 6 to 10 and Adulthood." *Psychological Science*, 17: 53-58
- Bartholow, B.D., Dickter, C.L. and Sestir. M.A. 2006. "Stereotype Activation and Control of Race Bias: Cognitive Control of Inhibition and Its Impairment by Alcohol." *Journal of Personality and Social Psychology*, 90: 272-287.
- Bertrand, M. and Mullainathan, S. 2003. "Are Emily and Greg More Employable Than Lakisha and Jamal?: A Field Experiment on Labor Market and Discrimination." Poverty Action Lab Paper No. 3.  
[http://povertyactionlab.org/papers/bertrand\\_mullainathan.pdf](http://povertyactionlab.org/papers/bertrand_mullainathan.pdf)
- Biernat, M. and Crandall, C. 1999. "Racial Attitudes." In P. Robinson, D. Shaver, and L. Wrightsman (eds.), *Measures of Political Attitudes*. San Diego, CA: Academic Press.
- Blair, I., Ma, J. and Lenton, A. 2001. "Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery." *Journal of Personality and Social Psychology*, 81(5): 828-841.
- Blanton, H., & Jaccard, J. (2008). "Unconscious racism: A concept in pursuit of a measure." *Annual Review of Sociology*, 34: 277-2097.
- Boxill, B. 1984. *Blacks and Social Justice*. Totowa, NJ: Rowman and Allenheld.
- Branch, C. W. and Newcombe, N. 1986. "Racial Attitude Development Among Young Black Children as a Function of Parental Attitudes: A Longitudinal and Cross - Sectional Study." *Child Development*, 57: 712-721.
- Brown, R. A. and Armelagos, G. J.. 2001. "Apportionment of Racial Diversity: A Review." *Evolutionary Anthropology*, 10: 34-40.
- Cosmides, L., Tooby, J. and Kurzban, R. 2003. "Perceptions of Race." *Trends in Cognitive Sciences*, 7(4): 173-179.

- Crandall, C. S. and Eshleman, A. 2003. "A Justification-Suppression Model of the Expression and Experience of Prejudice." *Psychological Bulletin*, 129(3): 414-446.
- Cunningham, W., Preacher, K and Banaji, M. 2001. "Implicit Attitude Measures: Consistency, Stability, and Convergent Validity." *Psychological Science*, 12(2): 163-170.
- Dasgupta, N. and Greenwald, A. 2001. "On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals." *Journal of Personality and Social Psychology*, 81(5): 800-814.
- Dasgupta, N., McGhee, D., Greenwald, A. and Banaji, M. 2000. "Automatic Preference for White Americans: Eliminating the Familiarity Explanation." *Journal of Experimental Social Psychology*, 36: 316-328.
- Devine, P. 1989. "Stereotypes and Prejudice: Their Automatic and Controlled Components." *Journal of Personality and Social Psychology*, 56: 5-18.
- Devine, P., Plant, E., Amodio, D., Harmon-Jones, E. and Vance, S. 2002. "The Regulation of Explicit and Implicit Race Bias: The Role of Motivations to Respond Without Prejudice." *Journal of Personality and Social Psychology*, 82(5): 835-848.
- Devos, T., Nosek, B. A. and Banaji, M. R. 2007. "Aliens in their Own Land? Implicit and Explicit Ascriptions of National Identity to Native Americans and White Americans." Unpublished Manuscript. Accessed @ <http://projectimplicit.net/articles.php> (06/08/2007).
- Diamond, J. 1994. Race without color. *Discover Magazine*, 15(11): 83-89.
- D'Souza, D. 1996. "The One-Drop-of-Blood Rule." *Forbes*, 158(13): 48.
- Dunbar, R. I. M. 2003. "The Social Brain: Mind, Language, and Society in Evolutionary Perspective." *Annual Review of Anthropology*, 32: 163-181.
- Dunham, Y., Baron, A. and Banaji, M. 2008. The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, 12(7): 248-253.
- Eberhardt, J. L. 2005. "Imaging race." *American Psychologist*, 60: 181-190.
- Faucher, L. and Machery, E. Forthcoming. "Racism: Against Jorge Garcia's Moral and Psychological Monism." *Philosophy of the Social Sciences*.
- Fazio, R., Jackson, J., Dunton, B. and Williams, C. 1995. "Variability in Automatic Activation as an Unobtrusive Measure of Racial Attitudes: A Bona Fide Pipeline?" *Journal of Personality and Social Psychology*, 69: 1013-1027.
- Fessler, D. M. T., Arguello A. P., Mekdara J. M. and Macias R. 2003. "Disgust Sensitivity and Meat Consumption: A Test of an Emotivist Account of Moral Vegetarianism." *Appetite*, 41(1): 31-41.
- Fredrickson, G. M. 2002. *Racism: A Short History*. Princeton University Press.
- Gannett, L. 2005. "Group Categories in Pharmacogenetics Research." *Philosophy of Science*, 72: 1232-1247.
- Gelman, S. A. and Wellman, H. M. 1991. "Insides and Essences: Early Understandings of the Non-Obvious." *Cognition*, 38: 213-244.
- Gil-White, F. 1999. "How Thick is Blood? The Plot Thickens ... : If Ethnic Actors are Primordialists, What Remains of the Circumstantialists/Primordialists Controversy?" *Ethnic and Racial Studies*, 22(5): 789-820.
- Gil-White, F. 2001. "Are Ethnic Groups Biological 'Species' to the Human Brain?" *Current Anthropology*, 42(4): 515-554.

- Govorun, O. and Payne, B. K. 2006. "Ego Depletion and Prejudice: Strong Effects of Simple Plans." *Social Cognition*, 24: 111-136.
- Greenwald, A. and Banaji, M. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review*, 102(1): 4-27.
- Greenwald, A., McGhee, D. and Schwartz, J. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology*, 74(6): 1464-1480.
- Greenwald, A., Nosek, B. and Banaji, R. 2003. "Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm." *Journal of Personality and Social Psychology*, 85: 197-216.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. In press. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*.
- Gregg, A. P., Seibt, B., and Banaji, M. R. 2006. Easier Done than Undone: Asymmetry in the Malleability of Implicit Preferences. *Journal of Personality and Social Psychology*, 90: 1-20.
- Griffiths, P. E. and Machery, E. 2008. Innateness, Canalization, and Biologising the Mind. *Philosophical Psychology*, 21: 397-414.
- Haslanger, S. (2000). "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" *Noûs* 34(1): 31-55.
- Hirschfeld, L. W. 1996. *Race in Making: Cognition, Culture, and the Child's Construction of Human Kinds*. Cambridge, MA: MIT Press.
- Hirschfeld, L. W. 2001. "On a Folk Theory of Society: Children, Evolution, and Mental Representations of Social Groups." *Personality and Social Psychology Review*, 5(2): 107-117.
- Isaac, B. H. 2004. *The Invention of Racism in Classical Antiquity*. Princeton, NJ: Princeton University Press.
- Kelly, D. and Roedder, E. 2008. "Racial Cognition and the Ethics of Implicit Bias." *Philosophy Compass*, 3(3): 522-540.
- Kurzban, R., J. Tooby and L. Cosmides. 2001. "Can Race Be Erased? Coalitional Computation and Social Categorization." *Proceeding of the National Academy of Science*, 98(26): 15387-15392.
- Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: Procedures and validity. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59-102). New York: Guilford Press.
- Lemm, K., & Banaji, M. R. (1999). Unconscious attitudes and beliefs about women and men. In U. Pasero & F. Braun (Eds.), *Wahrnehmung und Herstellung von Geschlecht (Perceiving and performing gender)* (pp. 215-233). Opladen: Westdeutscher Verlag.
- Lengbeyer, L. 2004. Racism and Impure Hearts. In Michael Levine & Tamas Pataki, eds., *Racism in Mind: Philosophical Explanations of Racism and Its Implications* (Cornell UP, 2004), pp.158-78.
- Levy, B., & Banaji, M. R. 2002. "Implicit ageism". In T. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 49-75). Cambridge, MA: MIT Press.

- Lewontin, R. C. 1972. "The Apportionment of Human Diversity." *Evolutionary Biology*, 6: 381-398.
- Lowery, B., Hardin, C. and Sinclair, S. 2001. "Social Influence Effects on Automatic Racial Prejudice." *Journal of Personality and Social Psychology*, 81(5): 842-855.
- Machery, E. and Faucher, L. 2005a. "Why do we Think Racially?" In H. Cohen and C. Lefebvre (eds.), *Handbook of Categorization in Cognitive Science*. Orlando, FL, Elsevier. pp. 1009-1033.
- Machery, E. and Faucher, L. 2005b. "Social Construction and the Concept of Race." *Philosophy of Science*, 72: 1208-1219.
- Machery, E. and Faucher, L. Ms. "The Folk Concept of Race: Experimental Studies."
- Mahalingam, R. 2003. "Essentialism, Culture and Power: Representations of Social Class." *Journal of Social Issues*, 59: 733-749.
- Mallon, R. 2004. "Passing, Traveling, and Reality: Social Construction and the Metaphysics of Race." *Noûs*, 38(4): 644-673
- Mallon, R. 2006. "'Race': Normative, Not Metaphysical or Semantic." *Ethics*, 116(3): 525-551.
- McConahay, J. 1986. "Modern Racism, Ambivalence, and the Modern Racism Scale." In J. F. Dovidio and S. L. Gaertner (eds.), *Prejudice, Discrimination, and Racism*, Orlando, FL: Academic Press.
- Mills, C. 1998. *Blackness Visible: Essays on Philosophy and Race*. Ithaca: Cornell University Press.
- Muir, D. 1993. "Race: The Mythic Root of Racism." *Sociological Inquiry*, 63: 339-350.
- Nei, M. and Roychoudhury, A.K. 1972. "Gene Differences Between Caucasian, Negro, and Japanese Populations." *Science*, 177: 434-436.
- Nei, M. and Roychoudhury, A.K. 1974. "Genetic Variation Within and Between the Three Major Races of Man, Caucasoids, Negroids, and Mongoloids." *American Journal of Human Genetics*, 26: 421-443.
- Nosek, B. A., Greenwald, A. G., and Banaji, M. R. 2007. "The Implicit Association Test at Age 7: A Methodological and Conceptual Review." In J. A. Bargh (ed.), *Automatic Processes in Social Thinking and Behavior*. Philadelphia, PA: Psychology Press. pp. 265-292.
- Ottaway, S. A., Hayden, D. and Oakes, M. 2001. "Implicit Attitudes and Racism: The Role of Word Familiarity and Frequency in the Implicit Association Test." *Social Cognition*, 18(2): 97-144.
- Outlaw, L. 1990. "Toward a Critical Theory of 'Race'." In D. T. Goldberg (ed.), *The Anatomy of Race*. Minneapolis, MN: The University of Minnesota Press. pp. 58-82.
- Outlaw, L. 1995. "On W.E.B. Du Bois's 'The Conservation of Races'." In L. A. Bell and D. Blumenfeld (eds.), *Overcoming Racism and Sexism*. Lanhan MD: University Press of America. pp. 79-102.
- Outlaw, L. 1996. *On Race and Philosophy*. New York: Routledge.
- Payne, B. K. 2005. "Conceptualizing Control in Social Cognition: The Role of Automatic and Controlled Processes in Misperceiving a Weapon." *Journal of Personality Social Psychology*, 81: 181-192.
- Payne, B. K. 2006. "Weapon Bias: Split-second Decisions and Unintended Stereotyping." *Current Directions in Psychological Science*, 15: 287-291.

- Phelps, E., Cannistraci, C., and Cunningham, W. 2003. "Intact Performance on An Indirect Measure of Race Bias Following Amygdala Damage." *Neuropsychologia*, 41: 203-208.
- Phelps, E., O'Connor, K., Cunningham, W., Funyama, S., Gatenby, C., Core, J. and Banaji, M. 2000. "Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation." *Journal of Cognitive Neuroscience*, 12(5): 729-38.
- Phelps, E. and Thomas, L. 2003. "Race, Behavior, and the Brain: The Role of Neuroimaging in Understanding Complex Social Behaviors." *Political Psychology*, 24(4): 747-758.
- Price, J. and Wolfers, J. Manuscript. "Racial Discrimination Among NBA Referees." Accessed @ <http://bpp.wharton.upenn.edu/jwolfers/research.shtml> (06/08/2008).
- Rankin, R. and Campbell, D. 1955. "Galvanic Skin Response to Negro and White Experimenters." *Journal of Abnormal and Social Psychology*, 51: 30-33.
- Richeson, J. and Ambady, N. 2003. "Effects of Situational Power on Automatic Racial Prejudice." *Journal of Experimental Social Psychology*, 39: 177-183.
- Richeson, J., Baird, A., Gordon, H., Heatherton, T., Wyland, C., Trawalter, S. and Shelton, N. 2003. "An fMRI Investigation of the Impact of Interracial Contact of Executive Function." *Nature Neuroscience*, 6(12): 1323-1328.
- Richeson, J. A., and Shelton, J. N. 2003. "When Prejudice Does Not Pay: Effects of Interracial Contact on Executive Function." *Psychological Science*, 14: 287-290.
- Root, M. (2000). "How We Divide the World." *Philosophy of Science* 67(Proceedings): 628-639.
- Root, M. 2005. "The Number of Black Widows in the National Academy of Sciences." *Philosophy of Science*, 72: 1197-1207.
- Rozin, P. 1982. "Human food selection: The interaction of biology, culture and individual experience." In L. M. Barker (Ed.), *The psychobiology of human food selection* (pp.225-254).
- Rozin, P. 1997. "Moralization." In A. Brandt and P. Rozin (eds.), *Morality + Health*. New York: Routledge.
- Schwartz, A. & Rashbaum, W. 2007, July 21. "N.B.A. Referee is the Focus of a Federal Inquiry." *The New York Times*.  
<http://www.nytimes.com/2007/07/21/sports/basketball/21referee.html>
- Schwartz, M. B., Vartanian, L. R., Nosek, B. A., & Brownell, K. D. (2006). The influence of one's own body weight on implicit and explicit anti-fat bias. *Obesity*, 14(3): 440-447.
- Shelby, T. (2002). "Foundations of Black Solidarity: Collective Identity or Common Oppression." *Ethics* 112.
- Shelby, T. (2005). *We who are dark : the philosophical foundations of Black solidarity*. Cambridge, Mass., Belknap Press of Harvard University Press.
- Stroop, J. 1935. "Studies of Inteference in Serial Verbal Reactions." *Journal of Experimental Psychology*, 18: 643-662.
- Sundstrom, R. 2002. "Racial Nominalism." *Journal of Social Philosophy*, 33(2): 193-210.
- Taylor, P. 2000. "Appiah's Uncompleted Argument: Du Bois and the Reality of Race." *Social Theory and Practice*, 26(1): 103-28.
- Taylor, P. 2004. *Race: A Philosophical Introduction*. Cambridge, UK: Polity Press.

- Taylor, S., Fiske, S., Etcoff, N., and Ruderman, A. 1978. "The Categorical and Contextual Bases of Person Memory and Stereotyping." *Journal of Personality and Social Psychology*, 36: 778-793.
- Vanman, E.J., Paul, B.Y., Ito, T.A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology*, 73, 941-959.
- Wasserstrom, R. 2001 (1980). "Racism and Sexism." *Philosophy and Social Issues: Five Studies*. Notre Dame, IN: Univ of Notre Dame Press. Reprinted in (ed. B. Boxill) *Race and Racism*. New York. Oxford University Press. pp. 307-343.
- Webster, Y. 1992. *The Racialization of America*. New York: St. Martin's Press.
- Young, I. M. (1989). "Polity and Group Difference: A Critique of the Idea of Universal Citizenship." *Ethics* 99: 250-74.
- Zack, N. 1993. *Race and Mixed Race*. Philadelphia, PA: Temple University Press..
- Zack, N. 1998. *Thinking About Race*. Belmont, CA: Wadsworth Publishing.
- Zack, N. 2002. *Philosophy of Science and Race*. New York: Routledge.