

DRAFT: Please do not quote without permission of one of the authors; comments welcome at drkelly@purdue.edu, machery+@pitt.edu, or rmallon@philosophy.utah.edu

Word Count (text only, excluding footnotes and references): 11,132

Racial Cognition and Normative Racial Theory

By Daniel Kelly, Edouard Machery, and Ron Mallon¹

Racism is a problem with many facets, and a strategy of divide and conquer is useful in making the problem more tractable. One facet, which is also a core question of contemporary social morality, concerns how we ought to handle racial categorization, by which we mean, for instance, thinking of a person as *black*, *Korean*, *Latino*, *white*, etc. While it is widely agreed that racial categorization played a crucial role in past racial oppression, there remains wide disagreement among philosophers and social theorists about a role for racial categorization in future endeavors. At one extreme of this disagreement are short-term eliminativists who want to eliminate 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 opposite extreme, 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 1990, 1995, 1996). In between these two poles, there are many who believe that racial categorization is valuable (and perhaps necessary)

¹ We would like to thank the members of the Moral Psychology Research Group—particularly, John Doris and Tim Schroeder—as well as Luc Faucher for their useful comments on previous versions of this article. We are especially thankful to Erica Roedder for her many insightful comments on previous versions. Remaining mistakes are ours.

given the continued existence of racial inequality and the lingering effects of past racism. 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.²

Our purpose in this paper is not to explore the nuances of this debate, nor is it to weigh in on one side of this disagreement. Rather, we want to explore the intersection of these normative proposals with recent work on the psychology of racial cognition. Race theorists often trade in normative arguments for conservationist or eliminativist ends, and these normative arguments typically involve evaluations of the costs and benefits attached to those ends (e.g., Boxill 1984; Appiah 1995; Muir 1993; D'Souza 1996; Outlaw 1990, 1995, 1996). Some theorists make these commitments clear. For instance, Outlaw's discussions of conservationism (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, Chp. 4) discussion of eliminativism all weigh the value of employing racial categories in different ways against the costs. Such evaluations invariably involve background assumptions regarding the ease with which racial categorization and racism can be eliminated or reformed.

Given these background assumptions, one might expect similar discussions about 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 behavior often bear directly on the revisionist goals of

² As we will note below, theorists also differ regarding the *domains* within which they endorse or reject the use of racial categorization, e.g. public policy domains, social scientific domains, etc.

conservationists and eliminativists. Our aim is, then, to show the need for a closer engagement between normative racial philosophy and contemporary psychology of racial categorization and racial prejudice.

We begin section 1 by examining several positions within philosophy of race. We show that race theorists often advocate reforming our practices of racial categorization, but they pay little heed to facts about the psychology of race—even when these facts could importantly impact the feasibility of such reforms. 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 into domain-specific mechanisms of 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 idiosyncratic operation of evolved cognitive mechanisms. Moreover, we show that this research has significant 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. Surprisingly, this research program suggests that implicit racist biases can persist even in persons sincerely professing tolerant or even anti-racist views, and thus that implicit racial evaluations can be insulated in important ways from more explicitly held beliefs. We again argue that these findings bear on the feasibility of eliminativist and conservationist proposals. We conclude in Section 4 that taken as a whole, this recent empirical work on racial cognition is significant for both normative projects of race theory. We end by drawing out a few more specific implications from this work, to illustrate that the best hope for realizing

either normative program is a greater engagement with the specific details of racial psychology.

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. In particular, people endorsed a conception of race on which racial membership expressed itself in physical, cultural, moral, and emotional differences between members of various races.³ Such doctrines held that “natural” races exist, and that our racial categories are good predictors for a wide range of other socially and ethically significant features. We will call this *thick racialism*. While many racial theorists believe that members of a racial group share interesting biological properties (e.g., shared skin color, facial features, etc.), thick racialists holds that members of the same racial group share more than this, e.g. certain psychological tendencies or moral features. With the advent of modern genetics, it seemed obvious that the appropriate interpretation of such thick racialist claims was in terms of the emerging science of human heredity. Claims about psychological and moral differences between races were true as the result of specific genetic differences between races, namely the genetic differences that give rise to those different psychological tendencies and moral features. However, subsequent research has brought a host of thinkers across biology,

³ 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.

anthropology, social theory, as well as cognitive, social, and evolutionary psychology to a consensus that thick racialism is false.

The reasons for this *ontological consensus* that thick racialism is false are many, and increasing understanding of human genetic variation has played an important role.⁴ There still 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 disease screening or forensic science.⁵ Nevertheless, the important point for present purposes is that this ontological consensus against thick racialism is a shared, cross-disciplinary source 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 idealize 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:

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 expresses both of the ideas central to many contemporary eliminativist positions: the first being that thick racialism is false (a point Appiah expresses in saying

⁴ 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.

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

that there are no races), the second that continued use of the racial concept would be 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 value both the rejection of thick racialism and the elimination of any other forms of *racism*, but hold we should refrain from disposing of racial categories altogether.⁶ Outlaw (1995), 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" (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*" (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 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

⁶ 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.

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, suggesting 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)⁷

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 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 feasible.

With conservatism, the connections with psychology are more complicated, but it seems clear that conservationists, like Outlaw above, are typically committed to retaining

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

racial categorization while eliminating thick racialism and other forms of racism.⁸

Indeed, conservatism looks more plausible 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. But is this division of racial categorization from racial evaluation really possible? Here too, there is strong reason to think information about human psychology might be relevant to assessing conservationists' proposals.

In sum, both eliminativists and conservationists have, among their goals, patently psychological goals. 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 a reduction of pernicious racial evaluation.

The ease with which these psychological aims can be achieved looks to depend on the particular facts of racial cognition.

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

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. Naomi Zack (1998), for example, considers whether completing the project of racial eliminativism is politically or morally feasible given the protection the First Amendment provides to even mistaken thoughts and speech (16). 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

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

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 racial theorists? The question is not one that can be answered simply by reference to disciplinary boundaries, for philosophical racial theorists typically engage research from a wide range 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.⁹

Whatever the reason for this pervasive neglect of psychological considerations in the philosophic literature, we wish to highlight it, and to that end, we'll call it *the disregard of psychology*. Rather than speculate on what motivates the disregard of psychology, though, we simply argue that it is unjustified, and show that recent findings about racial cognition are directly relevant to assessing the feasibility of both eliminativism and conservationism. Realistically evaluating eliminativist and conservationist goals can only be accomplished if one takes into account some of the more robust and surprising findings produced by 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 simply untenable. Along the way, we also

⁹ 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 his exclusion of psychological factors is less so.

draw out some more detailed conclusions about how specific psychological results impact 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 eliminating thick racialism and mitigating the evaluations accompanying them. 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, Domain-Specificity, and Canalization

Racial categorization presents a puzzle for evolutionary-minded psychologists and anthropologists (Hirschfeld 1996; Gil-White 1999, 2001a; 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 they thereby group together people who share important biological properties (and perhaps also important psychological and moral properties as well). However, it is difficult to account for this phenomenon with the explanatory resources favored by evolutionary theorists, namely by appeal to an evolved, domain-specific cognitive system dealing with racial cognition. 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, 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 noticed in the first section, there is an ontological consensus amongst researchers of many stripes that whatever else they might be, races are not natural categories that support robust physical, social, psychological, and behavioral generalizations. 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 a *byproduct* of an evolved domain-specific cognitive system, but one that was initially selected for another function. 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 this component.

First though, some background will be useful in understanding these debates and the evolutionary-cognitive approach itself, for that approach stands in contrast with previous explanations of racial categorization in psychology and the social sciences, including *socialization* explanations, *perceptual saliency* explanations, and *group prejudice* explanations. Psychologists favoring explanations in terms of socialization have assumed that we are taught to draw the substantive distinctions used in racial categorization, thus that we pick them up from the social environment (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 disposition to

classify people racially is underwritten by an evolved cognitive system, whose development in children is to a large extent independent of teaching.

Another view at odds with the evolutionary approach holds that racial categorization results from domain-general categorization based on *perceptually salient* properties. On this view of categorization, which is not specifically about race, we classify a wide variety of objects into categories based on their perceptually salient features. The view straightforwardly applies to racial classification: since color is a salient visual property, skin colors trigger this domain-general categorization system, and as a result, we form 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 the specific domain of the social world, rather than with categories in general; we will see why below.

A final view held by many social psychologists maintains that racial categorization and racism are to be accounted for by a tendency to form *group* prejudices about social groups in general, be they women, races or social classes (e.g., Crandall and Eshleman 2003). Evolutionary psychologists reject this idea on the grounds that it conflates different types of social groups, and the different ways they are cognized. Evolutionary psychologists argue that not all social classifications and prejudices are produced by the same cognitive system, and conclude that racial cognition should be distinguished from other forms of cognition, such as gender classification and prejudices.

Evolutionary psychologists offer a variety of considerations in support of their distinctive approach to racial categorization, but we will focus our attention on five lines

of argument that favor the evolutionary-cognitive approach over the above explanations. First, and most controversially, evolutionary psychologists hold that *cross-culturally*, people rely on skin color and other bodily features to classify their fellows and further believe that such classifications pick out underlying biological categories. 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. *Pace* social constructionists, however, there is evidence that across cultures and times—e.g., in ancient Greece and in the Roman Empire (Isaac 2004)—people rely on classifications that are similar to modern racial categorization in two central respects: these classifications are based on phenotypic properties and are assumed to pick out biological categories. This is not to deny that racial categorization varies across cultures and times in many respects. But these core elements of racial categorization are not a merely parochial cultural phenomenon.¹⁰

Such cross-cultural commonality is just what an evolutionary psychologist would expect, since evolutionary psychologists view racial cognition as a byproduct of a stable component of the species-typical human cognitive endowment. In contrast, because it fails to give any explanation for why these core elements should recur across times and cultures, socialization accounts of racial categorization are, at best, incomplete.

In addition, despite having such beliefs about racial properties at an early age (see below), children are *not explicitly taught* to classify racially. Particularly, children's beliefs about races are not correlated with their parents' views (Branch and Newcombe

¹⁰ 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 2003), others with the development of scientific biological classifications by Linnaeus and Blumenbach in the 18th century (Banton 1978), while still others hold European social ideology from the end of the 19th century ultimately responsible (Guillaumin 1980).

1986; Aboud and Doyle 1996). This constitutes a second type of evidence against socialization explanations of the disposition to categorize racially.

Third, explanations of racial cognition that rely on perceptual saliency take for granted exactly what is to be explained. Assuming that skin color is a perceptually salient property begs the question: it simply assumes that people tend to pay attention to skin color and other skin-deep racial properties. Color is not always an intrinsically salient feature, though. For example, we usually neither pay attention to nor remember the color of artifacts. Thus, the saliency of *skin* color in particular needs to be explained.

Fourth, social psychologists' emphasis on group prejudice is unable to account for the differences between different types of social classification and prejudices. Stereotypes vary substantially across domains. To take only one example, disabilities-related stereotypes do not include the idea that disabilities-based categories are biological kinds. Races, on the other hand, *are* thought 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—a tendency to form stereotypes about social groups—we should not find such differences.

Fifth and finally, 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, intrinsic, innate, and inherited. This sort of essentialism is also characteristic of children's and adults' folk biological reasoning (Gelman and Wellman 1991). Because not all prejudices involve this form of essentialism, this is evidence against the group prejudice explanation of racial categorization.

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 problem for the view that the saliency of racial membership can be accounted for in terms of the perceptual saliency of skin color. 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 our folk biology. Thus, racial categorization seems to be neither the product of socialization alone nor of the perceptual saliency of skin color. 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 some species-typical, canalized cognitive system.¹¹ Because it is species-typical, environmentally canalized, and complex, this supposed cognitive system is plausibly the product of evolution by natural selection. Given the specific properties of racial categorization, this cognitive system is also plausibly domain-specific, treating race differently than other categories (including some other social categories). All this is grist for the mill of evolutionary psychologists.¹²

It is important to realize that without further argument, 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 arguably an adaptation. It too develops early, reliably, and cross-culturally. However, during development, several factors determine whether or not and how much people will be attracted by sweet foods (Rozin 1982). Thus, although canalized, our taste for sweetness isn't inevitable. Analogously, racial categorization may thus result from an evolved cognitive system without being inevitable. How difficult the elimination or modification of racial categorization is depends on the specific empirical details of its development and operation.

2.2 Controversies within Evolutionary Psychology

¹¹ Roughly speaking, a trait is environmentally canalized to the extent that its development is the same in different environments (for a more nuanced discussion, see Griffiths and Machery forthcoming).

¹² It is also worth noting that the evolutionary psychological approach need 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 between 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).

Against this backdrop of broad theoretic agreement, disputes have emerged about the specific character of our capacity to make racial classifications. Hirschfeld, Kurzban et al., and Gil-White have proposed three different accounts of the domain-specific cognitive system that is assumed to underlie racial categorization. The dust has not settled yet, but the resolution of their disagreements can impact on the debate between eliminativism and conservatism. In what follows, we briefly review the main proposals.

First, according to Hirschfeld (1996, 2001), racial categorization results from the interaction of an innate, evolved *folk sociology* and the specific social structure in which it is operating. The evolved function of this assumed folk sociology is to identify the social groups in the social environment of each of us. Given the importance of social life during the evolution of human beings (e.g., Dunbar 2003), the capacity to map the social world was most likely selected for. According to Hirschfeld, one aspect of this hypothesized folk sociology is that it essentializes whatever groups are salient in a given social environment. When societies are divided along racial lines, our folk sociology not only attends to those racial groups, but also essentializes them. In societies with a different social structure, 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 Indians' folk sociology will essentialize castes (for consistent evidence, see Mahalingam 2003).

A second proposal is made by Kurzban, Tooby and Cosmides, which is to a large extent similar to Hirschfeld's (2001, Cosmides et al. 2003).¹³ Instead of positing a folk sociology that picks out the salient social groups in a given social environment, they argue that racial categorization results from a cognitive system whose function is to track *coalitions* (that is, groups of individuals that help each other) in a given social environment. Kurzban et al. take for granted that races are coalitions in many modern societies, including the American society. Since the coalitional cognitive system tracks coalitions in the social environment, it picks out races in 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, subjects are presented with a fictional verbal exchange between the members of two basketball teams and with pictures of these members. Subjects are then given each sentence in the exchange, and asked who uttered it. Because subjects tend to confuse individuals who have been categorized as belonging to the same group, the resulting patterns of mistaken ascriptions are taken to indicate how members of the two teams are being categorized by those subjects. For instance, if subjects racially categorize the individuals involved in the verbal dispute, then when they make mistakes, they should be more likely to ascribe a statement made by a white character to another white character than to a black character. The results of this experiment were as follows. When coalitional membership (viz. membership in each basketball team) is not emphasized, subjects implicitly racially categorize the individuals involved in the verbal exchange. However, surprisingly, when

¹³ Among other subtleties, the two proposals differ largely in emphasis. While Hirschfeld focuses on the essentializing character of racial cognition, Kurzban and company stress their view that the proper function of the mechanism underlying racial cognition is linked to biological classification and folk biology.

coalitional membership is emphasized—by giving a colored jersey to each mixed race team—subjects appear to rely much less on racial categorization. This suggests 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 proposal is made by Gil-White, whose view is interestingly different from the first two (1999, 2001). Gil-White proposes that natural selection has selected for *an ethnic cognitive system*, that is, for a cognitive system whose evolved function is to identify ethnies. 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 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 our folk biology, the mechanism responsible for classifying and reasoning about species and biological kinds, was exapted for this purpose (for further detail, see Gil-White 2001). 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 biology 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 results from this cognitive system, because skin color and other racial properties (such as body type) are erroneously taken to be ethnic markers. In other words, races are mistaken for ethnies.

To summarize, controversies remain even among those who agree on the basic evolutionary-cognitive approach. Particularly, disagreements center around the evolved function of the domain-specific cognitive system believed to underlie racial categorization—tracking salient social groups, tracking coalitions, or tracking ethnies—as well as the reason 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 because of their disregard of psychology. To begin, each of these three pictures 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 proposal: during development, the cognitive system that underlies racialism is triggered by the use of race names ("black", "white"...) by parents, peers, etc., when parents, peers, etc., refer to social groups or characterize individuals. Children rely on such names to identify the important social groups in their social environment and they essentialize such groups. Race names are mapped onto specific visual cues (skin color, body shape...) later in development (Hirschfeld 1996, 136). Obviously, this proposal leaves many aspects of the development of racialism unspecified. However, it suggests that the feasibility of eliminativism is a function of the importance of races in people's

social environment. 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 are coalitions in our societies. Thus, if races are and remain coalitions, achieving the ideal of race blindness will be hindered by the structure of our evolved coalitional system.¹⁴ Kurzban et al.'s hypothesis also places interesting constraints on the type of programs we ought to use to promote eliminativism. Programs where African-Americans help other African-Americans (e.g., programs which assign junior minority professionals a senior minority mentor), for example, will tend to reinforce racial categories, if Kurzban et al. are right. On the other hand, programs in which *other* races help African-Americans (e.g., a classic affirmative action program in a predominantly white company) will 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 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 which ethnic cognition is supposed to have evolved. Nonetheless, like paleoanthropological ethnies, some modern groups may have substantial cultural

¹⁴ Kurzban et al. (2001) conclude 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 believe that this conclusion is hardly defensible. If they are right, the existence of racial categorization is bound to the existence of racially-based coalitions. Since these coalitions are an upshot of the economic and social structure, eliminating racial categorization requires extensive social reforms.

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 in some society, racial distinctions map onto such groups, the assumed 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 African-Americans in contemporary America.¹⁵ If Gil-White's theory is true, eliminativism could very well modifying the cultural structure of society. That is, it could require somehow weakening the cultural differences between racial groups (such as African-Americans and White-Americans in the USA). Given that such cultural differences are constitutive of people's identity, this is an important and potentially controversial cost for eliminativism. We also note that such reforms are of a different sort than the changes that would be required if Kurzban and colleagues were right.

Note that this may not be an inescapable difficulty for eliminativism. Social reform is not only possible, but it is already an integral component of our democratic societies (affirmative action, school integration, voting rights act, limits on redistricting, etc.). Moreover, as we noted in Section 1, eliminativism can come in different strengths, or be targeted on different social domains. It remains that the costs of social reform should weigh in the cost-benefit analysis of eliminativist proposals.

Conservationism, on the other hand, does not seem directly affected by these consequences. For, conservationists want to preserve racial distinctions. Thus, they do not have to change the cues that trigger the cognitive system that underlies racial categorization—that is, on the social or cultural structure of our societies. For instance,

¹⁵ Although this was not the case when African slaves arrived in the US. They came from different cultures in Africa.

they are not committed to weakening African-American cultural identity. As we saw, the evolutionary proposals considered in this section suggest that willy-nilly, eliminativists *are* committed to these projects.

The three hypotheses reviewed in Section 3.2 have another relevant consequence for the debate between eliminativists and conservationists. For, the plausibility of conservationism may be a direct function of what the correct psychological view is. Hirschfeld's and Gil-White's proposals tentatively suggest that racial categorization and essentialism—i.e., the belief that racial groups are biological groups characterized by essences that explain causally people's physical, behavioral and moral properties—are the product of the same cognitive system.¹⁶ Although details are scarce and evidence is lacking, their views 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 views, conserving racial categorization, while reforming its significance, 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 1986, 1994; Mills 1998; or Taylor 2004) may be defeated or at least complicated by the cognitive proclivity to interpret the categories in question in an essentialist matter. Of course, none of this implies a nonessentialist conservationism 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 it is

¹⁶ It should be noted that Hirschfeld does not give any evidence that moral properties are essentialized.

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 proposal turned out to be correct. For they propose that essentialism and the attention to racial physical properties stem from two different cognitive systems (Cosmides et al. 2003). Again, 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, species-typical, domain-specific cognitive system. If true, the evolutionary hypotheses would reveal that there are some specific problems for eliminativists and for conservationists. The three proposals considered here suggest that eliminativism is committed to some form of social reform. Moreover, as we saw, each proposal suggests that a distinct sort of social reform is needed for eliminativism and raises difficult specific normative questions about the way in which the cultural or coalitional unity of a minority group would have to be compromised in order to eliminate racial categorization. Hirschfeld's and Gil-White's proposals 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' proposal is congenial with this proposal. Neglecting psychology amounts to neglecting specific obstacles that

eliminativist and conservationists alike ought to tackle. The current disregard of psychology in the philosophic literature is untenable.

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 again have not been carefully discussed or examined because of the disregard of 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 prejudice and bias when subjected to indirect experimental methods. These methods were designed to bypass one's explicitly held views, i.e. those available via introspection and self-report, and instead probe the more subtle 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 inferring their mathematical abilities or wine sophistication from their performance on the respective tests. The first type of strategy depends for its reliability not only on the sincerity of the person's self-assessment and report, but also on his or her ability to introspectively access the information under consideration. The second type, though less direct in some ways, has the advantage of bypassing both of these. For similar reasons, strategies such as these have become trusted instruments for investigating many cognitive capacities. The research on implicit social cognition is no exception, and part of its progress has come by way of developing a number of measurement techniques based on this type of strategy, which we will call *indirect measures*.¹⁷

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)¹⁸

¹⁷ Phelps et al. 2000 and Phelps et al. 2003 use this term to distinguish them from those 'direct' measures that use techniques like subject interviews or questionnaires that rely on verbal and written self-report.

¹⁸ Thus characterized, indirect testing is not a particularly recent development to psychology (see, for instance, Stroop 1935).

This description isn't definitive, but it gets across the flavor of indirect measures. A variety of these indirect measures have been developed to investigate such mental states and processes. Below we will describe some of the most prominent indirect measures. In doing so, we will follow the terminology of Banaji et al. (2001) where it matters, using 'implicit' to describe those processes or mechanisms operating outside the subject's conscious awareness, and 'automatic' to denote those processes that operate without the subject's conscious control.¹⁹

*The Implicit Association Test (IAT) [EMAIL SENT 3/19 TO CHECK ON COPYRIGHTS
– Permission Granted from Banaji's lab]*

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

QuickTime™ and a
TIFF (LZW) decompressor
are needed to see this picture.

The IAT has been by far the most widely used indirect measure, and has been consequently subject to the most scrutiny.²⁰ 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

¹⁹ The term “implicit” is a source of potential confusion, as it is often applied to both the cognitive processes and the experimental measures used to probe them (Greenwald and Banaji 1995, Greenwald et al. 1998, Cunningham et al. 2001, Eberhardt 2005, Nosek et al. 2007), and is treated as a loosely synonymous with ‘automatic’ and ‘unconscious’, and various other terms.

²⁰ See Greenwald et. al. 1998 for the first presentation and initial results, Greenwald and Nosek 2001 and Nosek et al. 2007 for reviews and assessments of the methodological issues surrounding the use of the test and interpretation of results.

instance, an IAT might involve the category pair *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 the faces of blacks and whites, 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 of the task, exemplars are displayed one at a time, in random order, in the middle of a computer screen, and subjects must sort them as fast as they can.

Crucial to the logic of the test is the fact that subjects are required to sort the exemplars from the *four* categories using only *two* response options. For instance, they would be 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* (often comprised of 5 stages) tests, and the response options (the ‘e’ and ‘i’ keys) are assigned to different categories in different stages. So one stage might require the subject 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”). Subjects are to correct sorting errors as quickly as they can.²¹ Precise reaction times are

²¹ There are several variants of this basic paradigm (e.g., Cunningham et al. 2001).

measured by the computer on which the test is being taken, as is correction time and number of errors.²²

Coarse-grained interpretation of performance is fairly straightforward, given what the test was designed to probe. 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 responding (faster indicating stronger associations) and the frequency of errors (fewer errors indicating stronger association)” (Nosek et al. 2007, page 270). This is of course a relative, rather than absolute, measure, but the idea can be illustrated with our example case. If a subject 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 a relative, implicit preference for or bias towards whites over blacks; 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.

Modern Racism Scale (MRS)

²² See the previous citations 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/selectatest.html>.

By way of contrast with indirect measures like the IAT, the MRS is a *direct* measure of racial attitudes that is commonly used in conjunction with the indirect measures. This is standard self-report questionnaire that was designed to probe for more subtle forms of racial biases and prejudices (McConahay 1986). It asks questions explicitly about subjects' racial attitudes, and allows them to respond to statements with answers ranging from Strongly Disagree, through Somewhat Disagree, Neither Disagree nor Agree, Somewhat Agree, to Strongly Agree. Other direct measures sometimes used to assess subjects' explicitly endorsed racial attitudes, such as feeling thermometers and semantic differentials, similarly rely on self-report of feelings about whites and blacks.

The use of direct measures *together* with indirect measures is important because it is the conjunction of the two that really supports the inference to not just automatic but *implicit* processes and biases in the sense we discussed earlier, namely that they operate outside the introspective access and awareness of subjects. That subjects 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 self-report and introspection.

3.2 Evidence of Implicit Racial Biases

3.2.1 Implicit Racial Bias

While these types of measures have been used to probe and reveal a wide variety of implicit biases (age biases, gender biases, sexuality biases, disability biases, weight biases, religious biases), some of the first and most consistently confirmed findings yielded by these tests center on racial biases.²³ Subjects who profess tolerant or anti-

²³ Indeed, 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). The

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). The pattern of results is beyond dispute at this point. In other words, it is psychologically possible to be, and, it appears, many Americans actually are, *explicit racially unbiased while being implicit racially biased*.²⁴ This phenomenon is not only robust but counterintuitive. Not only is the presence of two sets of opposing racial attitudes within a single agent a significant departure from common sense, but as we will see, the ways in which implicit biases can be altered also deviate from common sense expectations as well.

3.2.2 Mitigating the Effects of Implicit Racial Biases

Perhaps the most important offshoot of this literature concerns the extent to which implicit biases can be moderated, regulated, eliminated, or otherwise controlled. Initial findings suggest that implicit biases are to some extent malleable.²⁵

For instance, some of these studies suggest that while implicit biases operate beyond the direct conscious control of the subjects themselves, they can be rather

results exhibited a now familiar, but still disturbing, pattern: while most (19 of 26) of the subjects 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 racist IATs have been reported so often and found so reliably that they have become commonplace (Kim and Greenwald 1998, Banaji 2001, Ottaway et al. 2001).

²⁴ 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 show complicated correlations (Greenwald et al. 2003, Nosek et al. 2007).

²⁵ 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.

dramatically influenced by aspects of the social environment. Focusing specifically on racial attitudes, Dasgupta and Greenwald (2001) showed subjects pictures of admired and disliked blacks and whites (Denzel Washington, Tom Hanks, Mike Tyson, Jeffrey Dahmer) and found that exposure to admired blacks and disliked whites weakened the pro-white IAT effect. Two corollaries to this finding were perhaps more striking: the weakening of the implicit bias found immediately after exposure to the pictures was still present 24 hours later, and the subjects' explicit attitudes remained unaffected. Lowery et al. (2001) found that implicit biases, again measured in white Americans by the IAT, could be lessened merely by having the subjects interact with a black experimenter rather than a white experimenter. Richeson and Ambady (2003) showed situational differences can affect implicit biases. White female subjects 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 superior role to a black in a subordinate role showed a greater IAT effect, i.e. greater traces of implicit racial bias, than those anticipating play a subordinate role to a black in a superior role.

Other studies investigate the extent to which a subject can obliquely influence their *own* implicit biases, either by suppressing their expression or indirectly affecting the implicit processes themselves. For instance, Blair et al. (2001) found that subjects 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.

Exploring malleability from another angle, Gregg et al. (2006) present data suggesting that implicit biases are at once easier to acquire than their explicit counterparts, but harder to alter once acquired, and are extremely difficult to eliminate. Strikingly, even when subjects were told that the information used to induce biases about one of two imaginary groups was incorrect (the mistaken result of a computer error), the implicit biases, as measured by an IAT, remained while the explicit biases disappeared.

These intriguing findings make up only the tip of an iceberg of highly active research on the nature and boundaries of malleability. The bulk of this work on malleability suggests at the very least that those underlying processes are not impossible to alter. Indeed, while they are inaccessible via direct introspection and appear not to require, indeed often *defy*, deliberation or conscious intention, the malleability data shows that they *are* amenable to more subtle forms of manipulation.

3.3 Consequences for the Debate between Eliminativism and Conservationism

3.3.1 Implicit Biases and Real World Behavior

A natural question to ask at this point is whether or not the biases revealed by indirect measurement techniques ever lead to prejudicial behavior 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. For, racial theorists 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 other largely inert mental representations that people neither endorse nor act upon in real world scenarios (see, for instance, Gehring et al. 2003).

We think this would be a mistake. First, we are impressed by mounting evidence that suggests implicit biases do indeed affect real world situations. In one study by Marianne Bertrand and Sendhil 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 African American sounding name or a very White sounding name. 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 African American resumes received a much smaller increase. Not only was the amount of discrimination fairly consistent across occupations and industries, but employers who explicitly list “Equal Opportunity Employer” in their ad discriminate just as much as other employers. Another recent statistical analysis of NBA games found an “opposite race bias” in the refereeing (Price and Wolfers, manuscript). The study, which took into account data from the 12 seasons from 1991-2003, found evidence of this in the number of fouls white referees called on black players, as well as the number of fouls black referees called on white 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, blocks, steals, and turnovers. This despite the fact that referees are subject to constant and intense scrutiny by the NBA itself, so much so that they have 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. While no indirect measures were used in either of these, both sets of findings certainly invite the interpretation that implicit biases are affecting practices in the real world. Indeed, especially in the NBA study, split second judgments must be made in high-pressure situations: exactly the type of situations where people’s behaviors are likely to be influenced by automatic processes.

Moreover, researchers have begun explicitly linking actual behavior with performance on indirect tests. Such studies further confirm that when subjects have to make instantaneous decisions, racial biases affect people’s behaviors. Payne (2006) reviews a large body of evidence concerning subjects who are asked to make snap discriminations between guns and a variety of harmless objects. Subjects, both White Americans and African-Americans, are more apt to misidentify a harmless object as a gun if they are first shown a picture of an African-American, rather than a picture of a White-American. This effect has become known as the “weapon bias.” Similar results are found with subjects who explicitly try to avoid racial biases. Moreover, presence of a weapon bias correlates with the indirect measures of racial biases (Payne 2005). This suggests that implicit racial biases may indeed lie behind the weapon bias.

The real world relevance of such findings is difficult to deny. It has been grimly illustrated by the tragic death of Amadou Diallo, who, in 1999, was shot 41 times by police officers who thought he as drawing a gun when in actuality he was just reaching for his wallet. In turn, the implications of this body of psychological research should not be ignored by those racial theorists seeking to offer feasible proposals about how best to deal with racial categorization and evaluation.

3.3.2 The High Costs of Conservatism

While it is fascinating in its own right, this body of work in social psychology is clearly relevant to the philosophical debate over race. To be forthright, the psychological story is still far from complete, and in a number of ways:

- (a) the extent and manner in which many of the results reported can be generalized to *other* cultures is still uncertain;
- (b) whether they can even be generalized to racial groups other than black and white within our *own* culture, such as Hispanics, Indian, Asian, etc. is unclear;
- (c) there is little systematic developmental data concerning the ontogenesis of implicit racial biases;
- (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 that 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 too unconstrained to be of much serious use (but see Section 2 for suggestions about more specific, psychologically important varieties of groups such as coalitions and ethnies).

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 how they impact 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 advocates a continuation of a standing trend in the United States on which racial categorization is retained, while racism is reduced. At first blush, proposals along these lines seem both sensible and realistically achievable. Indeed, as has been noted in a number of places (Bierniet and Crandall 1999, Shuman, Steeh and Bobo 1997, Phelps et al. 2000), the last couple of decades have show a significant decrease in the expression of explicit racist attitudes, as measured by self-report. While this is surely a sign of progress for both conservationists and eliminativists alike, the results reported in the previous section suggest that the actual state of affairs is much more complicated, and that the goals of conservationism are much less easy to bring about than it might initially seem. That it is psychologically possible to be, and that many white Americans indeed *are*, explicitly unbiased, but implicitly biased, suggests that a) maintaining racial categorization while b) purging racial categories of all of their derogatory evaluative baggage may be much more difficult than previously thought. At

the very least, racial categorizations and evaluations may be responsive to an entirely different set of techniques than we might have thought beforehand.

While some conservationists have recognized that racism may be difficult to eradicate,²⁶ by neglecting psychology, they have failed to grasp *why*, and the *extent* to which reforming racial classification of racism may be so difficult. By ignoring the existence and character of implicit biases in explicitly unbiased people, conservationists ignore some obstacles standing in the way of their own proposals.

To begin, implicit racial biases obviously differ from their explicit counterparts, and these differences need to be taken into account. In particular, we may point to three properties of implicit biases that should factor more prominently into conservationists' proposals and cost-benefit analyses. First, relative to explicit racial biases, implicit biases are *easier to acquire*, as shown by Gregg et al. (2006). This property indicates that any steps or methods of social reform designed to sheer racial categorization of negative implicit evaluation by blocking the acquisition of implicit biases by younger generations face a difficult challenge. Not only will those steps have to be quite rigorous to be effectual, but that so little is known about the mechanisms that support the acquisition of implicit biases only adds to the difficulties of proposals along these lines.

Second, for those that have acquired them, implicit racial biases appear suppressible, but require effort and attention to suppress. As noted above, some findings (Richeson et al. (2003), Richeson and Shelton (2003) and Govorun and Payne (2006)) indicate that suppressing implicit biases is effortful for individuals who harbor them. Indeed, Bartholow et al. (2006) have shown that alcohol consumption interferes with the capacity to intentionally control the expression of these biases. These findings raise

²⁶ E.g. Outlaw 1995, Taylor 2004.

problems for another variety of proposal conservationists might make that center not on alteration or eradication of implicit biases, but on their suppression. Even if eradicating or lessening the frequency of implicit racial biases is too costly to be feasible, the line of thought might go, the widespread *suppression* of implicit biases, so that they never or rarely manifest in overt behavior, will suffice. Again, the research suggests that implementing such proposals will be quite difficult. Implementing the widespread and consistent suppression of implicit biases seems to also require ensuring the vigilance and effort (and perhaps sobriety!) of those individuals who hold them.

Third, again relative to their explicit counterparts, implicit racial biases appear *more difficult to eradicate or reverse* once acquired. As mentioned above, this is given a striking experimental demonstration in Gregg et al. (2006). Thus, proposals urging that racial categorizations be shorn of racist evaluation, once they take the existence and character of implicit racial biases into account, will require more substantial amounts of reform than previously thought.

There is reason for hope, however. Perhaps the implicit biases are not so difficult to alter if the correct methods are used to alter them. Indeed, the research on malleability, while still very much a work in progress, suggests that this could very well be the case. Specifically, as we saw above, it shows that implicit biases are *not* completely immune to certain forms of influence. These results are highly suggestive of how such well-meaning efforts to affect implicit biases could most effectively be turned into concrete proposals, and put into practice. But again, this has an important upshot for assessing the feasibility of conservationist proposals as well. For not only are proposals calling for the lessening or eradication of negative racial evaluations committed to costly social reform, but if they

are to effectively impact on implicit biases as well as explicit ones, they are committed to enacting a significantly *different kind* of social reform than has been used previously.

Paying attention to psychology can provide hope by providing important insight into what form more effective social reform might take. Some results have suggested that affect and emotion are involved in the operation of implicit biases (e.g. Phelps et al. 2000). If this turns out to be the case, employing a class of emotion-based techniques that have successful in other areas might be more effective means by which conservationists can achieve their various proposals for sheering racism from racial categories. For instance, Rozin (1997) discusses 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 equally successful in the mitigation and elimination of even implicit racial biases.²⁷ Additionally, the preliminary data on malleability suggest that techniques such as increasing the visibility and prominence of counter stereotypic role models, could be used to separate racial categorizations from negative evaluative connotations, while retaining those classifications themselves.²⁸

To summarize, there is a substantial body of evidence that implicit racial bias can cohabit with a sincere rejection of racism. Behavioral and neuropsychological evidence suggests that implicit biases involve our emotions and can be, to a limited extent, put under deliberate control (Richeson et al. 2003). Though they function automatically and

²⁷ Other methods of social influencing that appeal to emotions, for instance those centering on shaming or emotion-based advertising (see any political campaign for instances of the later), might also be successfully applied to implicit racial biases.

²⁸ Though such proposals are certainly attractive, there are reasons to be cautious. Fessler et al. 2003 have argued that Rozin notwithstanding, moral vegetarianism may have little to do with disgust-based moralization.

are inaccessible to introspection, they are to some extent malleable by more exotic forms of influence. While the specifics of the malleability of implicit racial biases, as well as other aspects of implicit social cognition, await discovery and clarification in future research, it is evident that those forthcoming results, and many of the results already established, are directly relevant to the normative debate over race. At the very least, properties of the implicit biases have direct implications for our ability to realistically implement conservatism.

4. The Argument Against the Disregard of Psychology

Much of the philosophy of race 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 often *unacknowledged* consensus, that gives rise to the widespread disregard of psychology in that literature. Contributors write as if psychological research were also largely irrelevant to the normative racial debate, thus giving the appearance of assuming that there are no psychological facts that could affect, in either a positive or negative way, the feasibility of the various proposals under consideration.

We have argued that this disregard of psychology is unjustified. Rather, we have shown that findings in the research on racial cognition directly impacts on the feasibility of the goals held by normative racial theorists. We here provide a quick list of those findings and their consequences:

1. Racial categorization develops early, reliably, and cross-culturally. It is not a mere product of socialization, perceptual learning, or group-based cognition. Traditional educational reform will not necessarily guarantee the success of eliminativist or conservationist proposals.
2. The three evolutionary proposals reviewed in section 2 entail that eliminativists are committed to a substantial and particular type of social reform—weakening the saliency of races, if Hirschfeld is right, modifying the cooperative structure of our societies, if Kurzban, Tooby and Cosmides are correct, and transforming the cultural structure of our societies, if Gil-White is on the right track.
3. If racial categorization is a by-product of an ethnic cognitive system à la Gil-White or of a folk sociology à la Hirschfeld, then dissociating racial categorization from the belief that races are biological entities, as proposed by conservationists, may be difficult, for on these proposals both are produced by the same cognitive system. Kurzban et al.'s proposal, according to which racial categorization and the belief that races are biological kinds are produced by two different cognitive systems does not raise this problem for conservationists.
4. Implicit racial bias can coexist with the sincere rejection of racial prejudices. Thus, the existence and nature of implicit racial biases is a specific obstacle to conservationist proposals. Explicit prejudices have declined steadily, while implicit biases seem to be more robust. This suggests that the belief that racial evaluations can be revised and altered merely via processes of general social learning is wrong. This obstacle has been ignored by conservationists because of their ignorance of the psychology of racial prejudice.

5. To a large extent, implicit racial biases behave much differently than their explicit counterparts—they are not necessarily consistent with our explicit beliefs and seem to be little affected by rational argumentation and explicit teaching. Implicit biases may be to some extent malleable, but not by the usual common sense methods.
6. The suppression of existing implicit biases is itself a costly and effortful action, whose effectiveness degrades when those with implicit biases are under high stress or other less than ideal conditions.

Once again, our aim was not to weigh in on one side of this controversy, but to point out an assumption apparently made by *both* sides of the debate, and show it to be untenable. The disregard of psychology notwithstanding, the psychology of racial categorization and racial prejudice can no longer be neglected by serious race theorists involved in the debate between eliminativism and conservationism. Just as they have taken account of legal or social obstacles, race theorists would do well to begin taking the psychological obstacles into consideration as well.

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.
- Amadio, D., Harmon-Jones, E., and Devine, P. 2003. "Individual Differences in the Activation and Control of Affective Race Bias as Assessed by Startle Eyeblink Response and Self-Report." *Journal of Personality and Social Psychology*, 84(4): 738-753.
- Andreasen, R. O. 1998. "A New Perspective on the Race Debate." *British Journal of the Philosophy of Science*, 49: 199-225
- 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.
- Ariew, A. 1996. "Innateness and Canalization." *Philosophy of Science*, 63: S19-S27.
- Ariew, A. 1999. "Innateness is Canalization: In Defense of a Developmental Account of Innateness." In Valerie Hardcastle (ed.), *Where Biology Meets Psychology: Philosophical Essays*. Cambridge, MA: MIT Press. pp. 117-138.
- 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.
- Banton, M. 1978. *The Idea of Race*. Boulder, CO: Westview.
- 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
- Biernat, M. and Crandall, C. 1999. "Racial Attitudes." In Robinson, P., Shaver, D. 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.
- Bradley, M., Cuthbert, B. and Lang, P. 1999. "Affect and the Startle Reflex" In M. Dawson, A. Schell and A. Bohmelt (eds.), *Startle Modification: Implications for Neuroscience, Cognitive Science, and Clinical Science*. New York: Cambridge University Press.
- 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, Ryan A., and George J. Armelagos. 2001. "Apportionment of Racial Diversity: A Review." *Evolutionary Anthropology*, 10: 34-40.
- Buchanan, Allen. 2003. *Justice, Legitimacy, and Self-Determination: Moral Foundations for International Law*. New York: Oxford University Press.
- Cosmides, L. and J. Tooby. 1992. "Cognitive Adaptations for Social Exchanges." In J. H. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press. pp.163-228.
- Cosmides, L., J. Tooby and R. Kurzban. 2003. "Perceptions of Race." *Trends in Cognitive Science*, 7, (4): 173-179.
- Crandall, C. S. and A. Eshleman. 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.
- D'Souza, Dinesh. 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: 163-81.
- Eberhardt, J.L. 2005. "Imaging race." *American Psychologist*, 60: 181-190.
- 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. 2003. *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 H. M. Wellman. 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. 2001a. "Are Ethnic Groups Biological 'Species' to the Human Brain?" *Current Anthropology*, 42(4): 515-554.
- Gil-White, F. 2001b. "Sorting is Not Categorization: A Critique of the Claim that Brazilians Have Fuzzy Racial Categories." *Cognition and Culture*, 1(3): 219-249.
- 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.
- 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. (Forthcoming). Innateness, Canalization, and Biologising the Mind. *Philosophical Psychology*.
- Harris, Leonard. 1999. *Racism*. New York: Humanity Books
- 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.
- Isaac, B. H. Manuscript. "The Study of Ethnicity Needs Better Categories."
- Kitcher, P. 1999. "Race, Ethnicity, Biology, Culture." In L. Harris (ed.), 1999, pp 87-120.
- 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.
- 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.
- 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. "Concepts of Races are Biological: A Cross-Cultural Study."
- 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.
- 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."

- 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.
- Richerson, P., and R. Boyd. 2004. *Not by Genes Alone*. Chicago: The University of Chicago Press.
- Root, M. 2005. "The Number of Black Widows in the National Academy of Sciences." *Philosophy of Science*, 72: 1197-1207.
- Rozin, P. 1997. "Moralization." In Brandt, A. and P. Rozin, (eds.), *Morality + Health*. New York: Routledge.
- Rozin, P., Lowery, L., and Ebert, R. 1994. "Varieties of Disgust Faces and the Structure of Disgust." *Journal of Personality and Social Psychology*, 66(5): 870-881.
- Schuman H, Steeh C, BoBo L. 1997. "Racial Attitudes in America: Trends and Interpretations." Cambridge, UK: Harvard University Press.
- Sober, E. 1998. "Innate Knowledge." In Craig, E. (ed.) *The Routledge Encyclopedia of Philosophy*, Vol 4. London: Routledge. pp. 794-797.
- Stich, S. 1975. "Introduction: The Idea of Innateness." *Innate Ideas*. Berkeley, CA: University of California Press. pp. 1-22.
- 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.
- Tooby, J., and Cosmides, L. 1992. "The Psychological Foundations of Culture." In J. Barkow, L. Cosmides, and J. Tooby (eds.), *The Adapted Mind*. New York: Oxford University Press. pp. 19-136.
- Vanman, E. 2003. "Developing a Comprehensive Social Psychology with Shared Explanations of Primate Social Behavior." *International Journal of Comparative Psychology*, 16: 33-43.
- Vanman, E. and Miller, N. 1993. "Applications of Emotion Theory and Research to Stereotyping and Intergroup Relations." In D. M. Mackie and D. L. Hamilton (eds.), *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception*. Orlando, FL: Academic Press. pp. 213-238.
- Vidulich, R. and Krevanick, F. 1966. "Racial Attitudes and Emotional Response to Visual Representations of the Negro." *Journal of Social Psychology*, 68: 85-93.

- 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.
- 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.