[NB: Readers who feel that this paper is itself an example of implicit racism, please see the appendix, "A note on language."]

Implicit Racism: Cognitive Origins and Potential Interventions

The study of racism encompasses the entire field of psychology: it touches on our attributions and inferences about other people, our construction of social groups and of humanity in general, and of our ability to rapidly learn and adapt to cultural information. It also involves our ability to systematically ignore or distort evidence that disagrees with our beliefs, and to segregate behavior so that we can be generous and humane in some circumstances and brutal in others. It demonstrates the potential the study of internal mental processes has to speak to vital political and human rights issues. Implicit racism, racist thought and behavior which many people exhibit even when they consciously hold strongly anti-racist beliefs, adds another level of complexity: individuals come into conflict not only with their society's ideals or general beliefs about moral behavior, but with their own explicit goals and intentions. This paper is in no way meant as a comprehensive review of the literature on racism, or even implicit racism. My goal is to find useful general properties of the phenomenon of implicit racism, and interpret them through the lens of social cognition and automatic cognitive processes. Greater attention to the low-level cognitive processes that give rise to racist thought and belief will, I hope, give rise to a greater understanding of the origins and ethical implications of racism, and help unify our knowledge about stereotyping, discrimination, and prejudice with our knowledge about other aspects of psychology.

Implicit Racism

The concept of implicit racism began to develop in the 1970's, as it became clear that large decreases in survey measures of anti-black beliefs and attitudes did not translate into colorblind egalitarianism (McConahay & Hough, 1979). School desegregation was seen as a particularly lamentable failure. Initially offered up as a way to decrease racist attitudes, improve minority-group self-esteem, and reduce racial gaps in educational achievement, it instead produced schools with homogeneous and non-interacting racial subpopulations, in which white students continued to outperform other groups (Aronson & Bridgeman, 1979). Psychologists began to search for ways in which self-reports of egalitarian beliefs among white Americans might be misleading.

McConahay & Hough (1976) developed the Modern Racism Scale, based on the concept of "symbolic racism." The MRS asks subjects about their beliefs about policies (such as busing) and broad social statements (e.g., that racism is no longer a problem and black people no longer need to push for social change) which bear on possibilities for improving the social status of black people, without mentioning overt hostility or dislike for them. Other researchers developed behavioral tests, measuring the extent to which white subjects were differentially helpful, aggressive, or interpersonally warm towards blacks as opposed to whites (reviewed in Crosby, Bromley, & Saxe, 1980, Gaertner & Dovidio, 1986). In addition to the usual benefits of avoiding self-report measures, these situations -- some of which were highly naturalistic, and all of which avoided being overt tests of racism -- allowed measurement of behaviors which effectively discriminated[1] against or communicated dislike to black people, regardless of what the actors involved thought about them. The MRS, while still a survey measure, also characterizes beliefs by their results rather than the belief-holder's understanding of them.

This line of research permitted an investigation of de facto racism, which promised to explain persistent racial disparities in education and socioeconomic status in the face of massive changes in both social norms and reports of racist beliefs by whites. However, it made the situation of its white subjects much more puzzling: are they racists, or not? A typical finding by

[1] I follow the tripartite division of attitudes used by Fiske (1998): stereotyping refers to overly-broad beliefs and cognitions about group differences, prejudice to affective reactions based on the target's group membership, and discrimination to behavioral differences based on the group of the target.

Gaertner from 1973 (discussed in Gaertner & Dovidio, 1986) demonstrates the difficulty: white residents of Brooklyn, New York who were registered as either liberal or conservative party members received a phone call from a stranger. The stranger "realized" that he had called the wrong number, but went on to say that he'd been trying to call a tow truck because his car had broken down, that he was at a pay phone, and that he had no more change. Predictably, conservatives were more likely to refuse to assist a black caller than a white caller, while liberals helped both about equally. However, liberals were more likely to hang up on a black caller before he could ask for assistance than a white one, and this disparity was greater than the disparity for conservatives. Clearly this kind of behavior is not the same as old-fashioned, overt hostility towards black people (what researchers once referred to as "redneck racism"), especially considering that other liberal party members, when asked about the situation hypothetically, asserted that they would help anyone, without regard to race. Yet this behavior is clearly a case of racial discrimination, and has clear implications for the ability of black people to function effectively and feel comfortable among white people, especially in educational and professional contexts in which whites still predominate and hold most positions of power.

Crosby, Bromley, & Saxe (1980) argued that these results were a straightforward indication that white people in America continued to dislike black people and wanted to oppress and hurt them. Changing laws and social norms had made this an unacceptable attitude to express, so whites instead used discourtesy, interpersonal distance, and institutionalized forms of aggression (such as administering electric shocks in the context of a learning experiment). In their narrative review of the literature, they presented studies showing that white subjects were more aggressive towards blacks in anonymous rather than face-to-face conditions, and switched from direct aggression (e.g., delivering more or more intense shocks than necessary) to indirect aggression (e.g., holding down the shock button longer) when there was the possibility of formal censure or retaliation for direct aggression. "We inferred from the literature," they wrote, "that whites today are, in fact, more prejudiced than they are wont to admit" (p. 557).

Several authors, working from the same body of evidence, developed a different view. Devine (1991), writing 11 years later but reviewing a similarly troubling body of evidence, agreed that racism remained prevalent but concluded that "many of the subjects in the present research... appear to be embroiled in the arduous task of breaking the prejudice habit" (p. 829). Crosby, Bromley, & Saxe had located racism in individuals, and progressivism within the culture; these more optimistic authors claimed that racist socialization is still endemic, alongside socialization with more egalitarian beliefs, and the two compete in the minds of individuals. Gaertner & Dovidio (1986) took at face value the frequent and consistent claims of non-racist beliefs among white subjects, and noted that when some aspect of the situation reminded them of egalitarian or humanitarian norms, signs of racism in their behavior suddenly disappeared. This could be seen as a social desirability effect, but there is other evidence that their desire to be egalitarian is genuine. Devine et al. (1991) asked subjects to rate how they believed they should and, separately, how they would feel or act in several situations involving contact with or judgment of oppressed minorities (blacks or gay men). Subjects who reported low prejudice were willing to report that they would feel or act more negatively than they thought appropriate, and reported feelings of discomfort and guilt which were consistent with the discrepancy. Devine (1989) and Monteith (1992) also reported evidence that low-prejudiced[2] individuals would, when possible, make use of mental control tactics to reduce their levels of prejudicial or discriminatory action, but high-prejudiced individuals would not.

The theory developed to explain this type of unintended racism states that the awareness of stereotypical beliefs and racist norms is unavoidable. Gaertner & Dovidio (1986) point out that images and beliefs which subtly denigrate blacks are still common in mainstream white culture. These authors also follow Allport (1954) in noting that racist beliefs can be picked up quite innocently: because black people have historically been forced into low-status jobs and poor, crime-prone neighborhoods, a naive observer would be correct in noting that they are, in

[2] As in the literature, the terms "high-prejudice" and "low-prejudice" will be used to refer to responses to self-report measures of subtle prejudice, including the MRS and similar measures.

fact, more likely to be poor, undereducated, or criminal. This knowledge is consumed by constantly-operating, omnivorous unconscious processes that work to automate trait inferences and similar heuristics, and is not directly modifiable through the less-experiential knowledge that there are situational explanations for the group differences (this can be seen as an instance of the fundamental attribution error [Ross, 1977]). Yet in spite of these powerful cognitive pressures, individuals often do avoid expressing stereotypical beliefs or acting in discriminatory ways. If we accept that most people have some desire to be genuinely egalitarian, and sometimes succeed, we open a new class of questions: why are stereotypical beliefs so easy to acquire and so difficult to control? What control processes can override implicitly acquired stereotypes? Which situations are relatively more or less amenable to control, and what are the short- and long-term effects of exerting control? We will see that these questions illuminate the interaction between automatic and controlled cognition, the motivational and affective aspects of controlling automatic processing, and the potential benefits and dangers of this type of control.

Automatic Processes and the Cognitive Components of Implicit Racism

The automatic social cognition model of implicit racism specifies the mechanism for the process described above: contact with denigrated minority groups produces automatic cognitions and affective reactions, which can only later be altered or suppressed by deliberate mental control. Bargh (1994) identifies four properties which distinguish automatic processing: it is unintentional, in that certain environmental stimuli are sufficient to activate it without effort; unaware, in that either the initiating stimulus or the process itself does not enter consciousness, although its output might; efficient, in that it is well-adapted to processing a certain type of information and does so more quickly than deliberate processing; and difficult to control, in that it often functions in spite of attempts to suppress it or alter its course. Automatic processes are diverse and intuitively familiar: we swerve our heads in the direction of an unexpected noise; we see a computer prompt and type our passwords, without having to consciously remember them or see feedback on the screen; we see a person in a blue uniform and interact with them in a guarded and deferential fashion, without having to effortfully scrutinize their outfit and retrieve a comparison representation of a police uniform from memory.

Automatic behavior can be seen as having high adaptive value: automatized physical processes allow us to respond quickly and reliably to important stimuli (Wegner & Bargh, 1998), while cognitive heuristics help us make rapid and satisficing decisions in a world with overwhelming amounts of often-confusing information (e.g., Griffin & Kahneman, 2003). The speed and accuracy of automatic processes, and the value of the attention they free up for other tasks, so impressed early psychologists that it inspired the proclamation "there is no freedom except through automation" (Bryan & Harter, 1899; quoted in Shiffrin, 1988). However, this autonomy comes at a cost: a fifth characteristic of automatic processes is that they lack direct feedback (Wegner & Bargh, 1998), and are thus difficult to change when our goals or a change in the environment renders them unsuitable. This, too, is familiar: it happens when an intermittently noisy refrigerator makes us twitch every time it starts to growl, even though we know consciously that there's nothing important to orient towards; when we notice that we've typed our password in response to a prompt that was actually asking for our e-mail address; and when we discover that we just turned our ID over to a private security guard who is dressed to mimic a police officer. It also occurs when we react negatively towards a member of an underprivileged minority group, even though we had intended to treat them fairly.

Another process tying together this diverse range of phenomena is spreading activation. The spreading activation concept models memory as a network, in which facts, beliefs, sensory memories, etc., are nodes, and these nodes are connected by paths representing conceptual linkages or observed co-occurrence. When a thought, emotion, or external stimulus is associated with something in memory, all the nodes connected to that node also increase in activation, which is operationalized as an increased likelihood for that concept to be remembered, spontaneously expressed, or used as a set for understanding other stimuli (Higgins & King, 1983). This increased activation is referred to as priming. Although there are debates about the details of the network structure and the nature of activation, this basic statement of the associative model is broadly accurate (McNamara, 1992). Furthermore, while priming through spreading activation has most often been investigated with simple word-associations, studies have shown that primes can affect higher-level processes ranging from judgment of another person's behavior to one's own level of performance on an intelligence test (e.g., Dijksterhuis & van Knippenberg, 1998).

Higgins & King (1983) discuss different ways in which we can draw inferences based on social category information: we can impute unobserved characteristics which are necessary aspects of category membership, or we can impute characteristics which are associated with that category, without being necessary. An example of this distinction would be learning that a person is a nurse, and assuming that he must have medical training (necessary), as opposed to assuming that he must be homosexual (associated). Recognizing this difference may give us a good model for understanding the change racism has undergone. Old-fashioned racism treated negative characteristics such as aggression and low intelligence as an integral part of black category membership, while implicit racists may have perfectly egalitarian definitions of racial categories, but still find some of those categories linked to negative attributes.

This automatic activation of attitudes can be unexpectedly powerful due to its dissociation from conscious belief or information processing. One of social psychology's classic findings is that people tend to make inappropriate attributions for emotions and sensations when the real cause is not apparent (Schachter & Singer, 1962; Zanna & Cooper, 1976), and Bargh (1994) notes that this may be generally true of automatic processes which operate outside of conscious awareness. An individual with negative associations for black people might meet a black stranger, experience mild feelings of dislike or discomfort, and end up with a negative impression of them which was caused entirely by internal associations. These associations can also bias conscious cognition by introducing a mental set for the interpretation of future social information (Higgins & King, 1983). A common effect in the racism literature is that white subjects will judge ambiguous behavior as more hostile when the actor is black as opposed to white. They may also act this way themselves: exposure to images of a group believed to be hostile (blacks) or prejudiced (skinheads), either subliminally (Bargh, Chen, & Burrows, 1996) or consciously (Macare, Bodenhausen, Milne, & Jetten, 1994; Kawakami, Dovidio, & Dijksterhuis, 2003), has been found to increase the hostility and prejudice in viewers.[3] Hostility is especially disturbing because it evokes hostility from others, meaning that a mild predisposition to act aggressively towards a black interaction partner can quickly become a self-fulfilling prophecy (Chen & Bargh, 1997). Following the dual influences on perception and action, the white subject then has another experience of stereotypical behavior to add to a growing negative impression of black people or skinheads.

Associative effects of this type can be particularly insidious: they escape attempts at control by never entering the subject's consciousness in the first place. Greenwald & Banaji (1995) cite evidence that priming effects are reduced if subjects focus on the prime or are made aware of the potential influence. This would be helpful to someone whose definition of being black necessarily involved aggression, if they wanted to temporarily reduce their tendency to perceive black people as aggressive, since the biasing belief would be known to them. But people whose explicit beliefs are non-racist could attempt to be aware of potential bias, find that there is none in their beliefs about black people, and thus decide there is no problem -- even if their mental representation of black people has strong associative ties to many of the same negative stereotypes. It might be possible to trace these activations through extended introspection, but this is unlikely: Gaertner & Dovidio (1986) point out that whites have strong social and moral pressures to believe that they are non-racist, and this makes it easier for them to blind themselves to subtle signs of racism which are apparent only in their behavior, or to find non-racial explanations for them. Devine, Monteith, Zuwerink, & Elliot (1991) found experimental evidence of discomfort when people who expressed low-prejudice beliefs were

[3] the phenomenon appears to generalize to adopting positive traits of valued groups, such as professors (Dijksterhuis & van Knippenberg, 1998), but to my knowledge no one has tested whether subjects will exhibit positive stereotypical traits of groups which are denigrated overall (e.g., athleticism and extraversion for blacks).

forced to confront signs of prejudice in their behavior, which suggests that people are unlikely to frequently engage in these thought experiments without strong motivation.

All this raises the question of where negative racial beliefs come from, and why they are so compelling. Allport's (1954) answer is still regrettably plausible: because black people have been historically oppressed, we are likely to see disproportionate numbers of black people with undesirable characteristics. Furthermore, our contact with other individuals who have made this observation, or who learned negative beliefs overtly at a time when they were socially normative, causes us to pick up these responses ourselves. We also have a whole stable of cognitive biases which make racial judgments difficult to disconfirm: we are biased towards making dispositional attributions for others' behavior; we attribute homogeneous dispositions to out-groups (Pettigrew's "ultimate attribution error"); and we display a bias towards evidence which confirms our pre-existing beliefs, seeking it out, attending to it, and remembering it more than disconfirming evidence (Jones & Nisbett, 1972).

In addition to domain-general learning and communication skills, and heuristics specific to social judgment, we also appear to have a specific predisposition to think about kinds of people.[4] Hirschfeld (1998) argues that we have a natural tendency to parcel people into essentialized, indivisible “natural kinds” which we believe to predict their behavior, and that children have an innate tendency to want to learn about natural kinds from others. and other evidence (discussed in Wegner & Bargh, 1998) suggests that social category information appears to enjoy a "privileged" status and is more likely to capture attention and influence our judgments (information about undesirable behavior is also privileged, and this may be one reason why negative stereotypes are more common than positive ones). Bartmess (2003) found that autistic children, who are developmentally impaired in language, communication, and theory of mind, nonetheless know about the racial and gender categories used in their culture and make use of the

[4] in the United States these kinds are conceptualized as biological races and differentiated by physical characteristics, but other cultures which discriminate based on smaller ethnic groups or heredity occupations (caste) behave similarly (e.g., Mahalingam, 2002).

associated stereotypes just as much as non-impaired children of the same mental age. Thus, it appears that our judgments about different racial groups feed into a mental module which is highly sensitive to racial cues and at least partially distinct from higher-level conscious thought, and this may explain why the judgments it produces are persistent, compelling, and difficult to change.[5]

Our intuitive styles of intergroup relation may also interact with our development of racist beliefs. Tajfel & Turner's social identity theory (1986) posits that in-group enhancement and out-group derogation are basic human ways of feeling secure and competent. Their use of the minimal-group paradigm, in which strangers are broken into groups at random, has demonstrated that even with no reason to perceive similarity to group members and no prospects for future interaction, people will still create intergroup conflict. This holds even when it is clear that each groups profit would be maximized by cooperating. Although social identity theory was originally developed as an alternative to theories based on practical conflict over limited resources, it's easy to see how this kind of devaluation can lead one group to try to deprive or marginalize others, causing an intrapsychic self-enhancement process to transition to one of overt conflict over naturally- or artificially-limited resources.

In a spreading-activation model of cognition, frequency of activation contributes to ease of future activation. Each of these explanations for the intuitive nature of racial stereotyping also helps explain why stereotypes are likely to come to mind even when they are inconsistent with our beliefs, questionably relevant to the current situation, and contrary to our desires. In addition to making activation more likely, this type of "practice" also makes a variety of skills and attentive processes faster and more automatic (Shiffrin, 1988). Devine (1989) demonstrates that low-prejudiced white subjects have the same knowledge of negative cultural stereotypes and

[5] Note that the proposition of a hard-wired racial categorization module is not inconsistent with the social construction of racial categories. Humans appear to have many innate adaptations which mold themselves based on environmental input, such as the ability to learn whatever language(s) are spoken around one as a child (see Tooby & Cosmides, 1991, for a detailed explanation).

racial slurs as high-prejudiced subjects, presumably as a result of growing up in a society in which racial beliefs and racial divisions are ubiquitous. She argues that these ways of thinking thus have a "history of personal activation" which builds up long before the complex cognitive abilities necessary to evaluate or control them are present. This is similar to Crosby, Bromley, & Saxe's (1980) argument that the discrepancy between espoused attitudes and behavior is due to new nonracist norms not being fully "internalized," but it deals with a multi-level view of self which can internalize beliefs at some levels while leaving others intact.

Challenges in the Control of Racist Cognitions

Although these implicit processes help explain where prejudice originates, they do not offer a full explanation for its persistence. Humans are not automatons, after all, and it seems that after years of overt attitude change, people should have been able to control or deliberately alter these automatic cognitions. This section will discuss three general reasons why automatic social cognitions are difficult to alter: insufficient conscious awareness, degradation of source and contextual information, and limited control capacity.

The first difficulty is that we can't exert conscious control over a process that occurs too subtly or rapidly to be noticed. Priming effects often take place on a time-scale that prevents us from recruiting effort in time to redirect them. In one ingenious demonstration (Neely, 1977), subjects were shown a prime word, then asked to categorize a target string as word- or non-word. When the target was a word, it bore either a conceptual relation or an arbitrary but reliable relation to the prime word, which varied by prime word (for example, the prime word "bird" was always followed by either a nonsense string or the name of a bird, while the prime word "building" was always followed by either a nonsense string or the name of a body part). The activation of the prime-target association was measured by improved reaction time on the word-categorization response. When subjects were shown the target word 250ms after the prime, the conceptual association improved reaction time, but the arbitrary association produced worse reaction times than no prime. When the time between prime and target was extended to 2000ms, both the arbitrary and the conceptually-related prime improved reaction time. This effect of prime type represents the difference between the time necessary for automatic spreading activation to make an association and the time necessary for a consciously-recognized but non-intuitive piece of knowledge to make a similar association. If we map endemic racist responses onto conceptually-related word associations and newer egalitarian beliefs onto arbitrarily learned associations, we have a model for the ability of racist beliefs to pop into our minds, leaving our conscious beliefs rushing to catch up. The one-and-three-quarters seconds interval between them seems like a short period of time, but it's sufficient to adopt a posture, make a snap judgment, or choose four or five words that have the ability to direct the future course of an interaction (see Kalma, 1991 and Rosenthal, 2003, for evidence that people can make reliable judgments about others' social interaction style after very brief exposure to limited information). This is especially true if, as discussed above, they evoke similar behavior from the interaction partner and set up a self-reinforcing behavioral cycle.

A second reason racist beliefs are difficult to dispute or change is that we often automatically process information which we would consciously reject. One of the oldest examples of this is the "sleeper effect" (e.g., Pratkanis, Leippe, Greenwald, & Baumgardner, 1988), in which subjects initially give high weight to information from credible sources and low weight to information from dubious sources, but over time come to use them equivalently. It appears that attributions and inferences which we would discount as racist show the same kind of forgetting-of-source effect, and exert gradually more influence on us over time (Higgins & King, 1983). Even information which a reliable source tells us is false outright is sometimes used in anchoring judgments about ourselves and others (Wegner, Coulton, & Wenzlaff, 1985, Schul & Burnstein, 1985). Information from a fictional story which subjects know perfectly well to be counterfactual (e.g., Al Gore as the current president) produces longer responding time to questions about the real world (Gerrig & Prentice, 1991). Daniel Gilbert (Gilbert, 1993; Gilbert, Tararrodi, & Malone, 1993) goes so far as to argue that the very act of comprehending a proposition involves believing it, and only later can we process it and add a tag indicating falsehood. Even if we do manage to catch ourselves thinking of unemployed black people as lazy or helpless and correct ourselves, that serves as no guarantee that those thoughts will not be strengthened by the activation and not be used in subsequent evaluations of them or others. Over time, the tags fall off.

Finally, even when we are able to recognize and successfully dispute automatic thoughts, we can't keep it up forever. Control over automatic processing requires deliberate attention-switching or thought-suppression, both of which require effort, and effort appears to be a limited resource (Baumeister, Bratslavsky, Muraven, & Tice, 1998). When executive control is exhausted, we are especially prone to act impulsively, reason heuristically, and fail to control or evaluate emotional reactions. This process is especially noticeable when it occurs as part of what Wegner (Wegner & Bargh, 1998) calls the ironic processing model of postsuppressional rebound: when we attempt to suppress a thought, we also have to be vigilant for its occurrence, which serves to increase its level of activation and make it more likely to arise. When control is exhausted or relaxed, this priming effect causes the thought to occur with greater frequency than it would have otherwise. Research on suppression of racist beliefs (Macare, Bodenhausen, Milne, & Jetten, 1994; Foerster & Liberman, 2001) demonstrates that this does in fact occur; subjects who attempt to avoid stereotyping skinheads or foreign workers on one task will later stereotype them even more later. Furthermore, Foerster & Liberman (2001) interpret their results as indicating that subjects begin to think of themselves as racists based on the occurrence of these thoughts -- although, as we will see later, other coexisting beliefs may cause this self-attribution to have a positive net effect.

A Digression: Intention, Responsibility, and Morality

Readers who are concerned with social justice could easily interpret this review as a list of excuses for mistreatment and discrimination against minorities. If people who perpetrate racist acts are themselves victims of uncontrollable cognitive processes, are we still able to condemn them and insist that they change? To the extent that people who act in unintentionally racist ways genuinely are ignorant of their actions and do desire an egalitarian society, I would argue that a general condemnation of them is unreasonable and, additionally, counterproductive. Research on learning and effort has determined that people vary in the extent to which they see intelligence, personality, and ability as continuous quantities which can change gradually ("incremental theorists"), as opposed to innate and largely unchangeable ("entity theorists") (Dweck, 1999). Incremental theorists are typically more willing to invest effort in learning and self-change[6], suggesting that a narrow focus on "racist thoughts" and "racist behavior" may be more effective than a thoroughgoing condemnation of people labeled "racists." This is the outlook of the following section, on potential control mechanisms: concern with moral condemnation and culpability is unproductive, and unfair to people who do not identify with their automatic behaviors and beliefs; developing and disseminating accurate knowledge and helpful techniques will serve them better than offering reasons to feel guilty. This focus on education is mirrored in Fiske's (1989) discussion of intentionality: she expresses concern that research on implicit racism will normalize racist behavior, due to the strong role that malicious intent plays in United States law and cultural morality. She concludes, however, that an individual can be judged "intentionally racist" if they declined to learn about and use readily-available techniques for reducing implicit racism; thus our goal should be to ensure that such techniques are available.

Possibilities for Control

Despite the difficulties discussed above, we do have control over some of the contents of our consciousness, and most automatic processes are ultimately amenable to some degree of change. All the researchers cited here, from Allport on, hold out some hope of reducing implicit racism, and there are several programs of research bearing on the precise processes that can give us control over our automaticity. Roughly speaking, they divide the field into individual acts of

[6] This outlook also predicts fewer dispositional attributions towards others, and more flexibility in initial impressions (Dweck, 1999) -- making this outlook directly beneficial in reducing racist behavior, as well as useful for inspiring effort.

control, and institutional- or social-level changes which offer different stimulus or input to individual cognitive systems.

Research on discrepancies between self-reports concerning appropriate behavior towards minorities and self-reports of imagined reactions to contact situations (Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1989) indicates that while the classic aversive / implicit racism effects still exist, subjects recognize them and feel bad about them. Subjects reporting low-prejudice were particularly likely to express shame and guilt, although even high-prejudice subjects (who imagined that they would respond even more aversively than was justified by their prejudices) felt some negativity. Gaertner & Dovidio (1986) predicted that such general negativity would be transferred to the minorities in question and make white subjects dislike them even more, but this may not be true for the self-directed negativity of Monteith's low-prejudice subjects. Indeed, Monteith sees these "compunctions" as a promising start on the road to self-regulation. She notes that negative emotions serve to focus attention, which can assist us in observing and disputing automatic thoughts. She identifies negative reactions to our thoughts with Gray's (1990) behavioral inhibition system (BIS), which causes us to stop, reorient towards the source of the negative emotion, and label it as aversive. Over time, this labeling can build up and cause either reduced activation or enhanced attention for the undesired racist thoughts, which will become less common or easier to control directly. The affect associated with an automatic cognition is often brought to mind at the same time as the cognition itself (Morris, Squires, Taber, & Lodge, 2003), so every experience of compunction upon having a racist thought should make future experiences of compunction and inhibition more likely.

This proposed system for change collides head-on with the Wegnerian concept of postsuppressional rebound. Any attention paid to the idea, especially attention which leads to suppression, would be predicted to enhance the idea's chronic level of activation, making it more likely to occur in the future. A variety of studies demonstrates that suppression, whether the object is racism, physical clumsiness, sleeplessness, or the image of a white bear, often enhances the very thing one is trying to suppress (Wegner, 1990; Wenzlaff & Wegner, 2000). Suppressing material which is emotionally important, as racism may be for people with egalitarian ideals, can also lead to emotional distress, reduced performance on important life tasks, and increased physical illness and mortality (Pennebaker, Zech, & Rimè, 2001; Wenzlaff & Wegner, 2000). In order to navigate a path between the dangers of suppression and the harm caused by uncontrolled automatic thoughts, we will need to examine the concepts of inhibition, suppression, and cognitive change in more detail.

Suppression is carried out by "an intentional operating process that will promote the preferred state (i.e., anything other than the unwanted thought)" (Wenzlaff & Wegner, 2000). Although it may require a good deal of mental effort to tear the mind away from a compelling or affectively-laden thought, suppression is fundamentally a matter of occupying attention elsewhere so that the undesired thought cannot occupy it. This can be conceptualized as raising the threshold required for a cognition to enter consciousness, but raising the threshold does not by itself reduce the activation which these cognitions already had. The distraction process can be problematic in three ways: first, if the suppressor has experienced the desire to entertain the thought but inhibited it, this lack of closure can actually enhance the concept's activation. This is the basis of the classical Zeigarnik effect, in which uncompleted tasks are remembered better than completed ones; it is also perhaps operational in findings that lack of closure contributes to psychopathologies such as depression and PTSD (see Wenzlaff & Wegner, 2000). Second, the act of monitoring consciousness to make sure that the undesired thought does not occur requires representing and thus activating that very thought (hence the name "ironic monitoring process"), to some extent. If this activation persists long enough or spreads to enough related cognitions, the thought will eventually break through to consciousness, especially if the distraction process has been relaxed or exhausted. Finally, distraction from the undesired thought does not necessarily entail distraction from the stimuli that provoked it. If the stimulus is transient, then it is quite possible that suppression will hold out until the opportunity for the thought has passed, in which case suppression has succeeded and the associative link between the stimulus and the thought will be weakened, as in Monteith's (1989) model. However, if the stimulus remains (for example, if a white person is in a conversation with a black person), the undesired thought will continue increasing in activation and suppression will be continuously challenged.

Overall, suppression functions to increase a thought's "deep cognitive activation," the extent to which it is subconsciously activated and capable of subtly influencing other processes such as memory, attention, and the activation of related thoughts (Wegner & Smart, 1997). In the case of racist cognitions, this might emerge as the behaviors Gaertner & Dovidio (1986) characterized as "aversive racism": avoidance, unfriendly (but still civil) behavior, and reduced helping -- but only when social norms provided an excuse for not helping, so the behavior did not have to be confronted as explicitly racist. The dissociation between spontaneous emotional behavior and effortfully-crafted conscious behavior is perhaps most visible in codings of speech during personal interactions (reported in Crosby, Bromley, & Saxe, 1980): when white subjects spoke with a white confederate, the friendliness they expressed in their words was strongly correlated with the warmth of their tone. When speaking with a black confederate, they typically said friendly things, but in a substantially less friendly tone of voice. These are good examples of suppressed, or simply consciously inaccessible, thoughts continuing to influence behavior, and a convincing argument against suppression as the sole strategy for reducing discrimination.

If suppressing automatic cognitions isn't the answer, what should we do once a sense of compunction (or a need for dissonance reduction, or an intellectual desire for egalitarianism) has alerted us to a racist thought or emotion? An alternate response is to take the awareness of the thought as an opportunity to respond. Although it seems inadvisable to hold a repugnant thought in mind longer than necessary, and doing so should indeed increase its level of activation, it also takes us out of the brief window in which we are compelled to accept only its strongest associations (Neely, 1977) or to treat it as true without opportunity for critical processing (Gilbert, 1993; Gilbert, Tararrodi, & Malone, 1993). There is, in fact, substantial evidence that attending to a potential source of bias can reduce its effects (reviewed in Greenwald & Banaji, 1995).

This model is also at the heart of cognitive therapy, a school of psychotherapy that is based on learning to recognize, make conscious, and dispute automatic thoughts[7] (Beck, 2000). Clients are assisted in identifying typical misperceptions and cognitive errors and in working out covert responses which remind them of their fallaciousness. Over time, this disputation is believed to become automatized itself, and to progressively weaken the associations that make the targeted misperceptions so easily activated. Although I am not aware of any direct longitudinal tests of Monteith's (1993) theories about weakening the chronic activation of racist beliefs, a number of studies of cognitive therapy for depression (Teasdale, et al. 2001; Segal, Gemar, & Williams, 1999; Whisman, 1993) demonstrate a parallel phenomenon: use of disputation leads to improvement in psychological health, which is associated with a reduction in the strength of the targeted automatic thoughts. Thus, if we focus on the automatic thought once it has been isolated and elaborate on its contradiction with our consciously-held principles and on the lack of valid evidence for it, we can, over time, reduce its chances of being activated. This is a somewhat cumbersome mental act and not always possible during real-world interactions, which is why cognitive therapy typically recommends working out arguments and refutations in advance, and practicing them until they become somewhat automatic in themselves. The disputation can then become a matter of "delegation" (Wegner & Bargh, 1998), in which we effortfully initiate a process which then runs automatically without further demands on executive resources. The evidence discussed above on sleeper effects and the difficulty of disregarding fictional information indicates that tagging an idea as "false" will not work very well, but with repeated applications the tag can become increasingly strongly associated with the idea, and eventually it should stick.

In addition to associating racist thoughts with disputations, we can imagine an intervention which is less direct but probably more emotionally engaging: associating racist thoughts with non-racist thoughts. If perceiving a member of a stereotyped group brings up

[7] in practice cognitive therapy is typically seen as a component of cognitive-behavioral therapy (CBT).

negative associations, but also memories of positive interactions or of situations in which stereotypes failed, the net result will at least be less negative. This also offers a more relevant distractor to turn our minds towards, perhaps eliminating some of the working memory load and ironic activation effects of suppression through irrelevant distraction. Forming positive associations may not be easy: the early activation history of negative stereotypes (Devine, 1989), the tendency to argue that individuals who disconfirm our stereotypes are exceptions or subtypes (Allport, 1954; Fiske, 1998), and the hypothesized drive to denigrate out-groups in order to enhance our in-group (Tajfel & Turner, 1986) all seem to dispose us towards negative perceptions of people perceived as being part of a different group. Creating opportunities for frequent and emotionally significant positive interactions may be the strength of institutional- and societal-level interventions, to which we now turn.

Allport (1954) long ago predicted that intergroup contact would reduce racism if the groups in contact were perceived as equal-status, if the contact was socially sanctioned and encouraged, and if the groups came to perceive common interests and personal similarities. These criteria, while their necessity and sufficiency have not been tested directly, can help us understand why some interventions have succeeded, and others have failed. School integration has often been seen as insufficient, because it did not address the economic disparities that continued to keep students separate even when physically intermingled, and it did not unite the students in any common goals. The value of cooperation in reducing group conflict inspired the creation of the jigsaw classroom, a teaching style developed throughout the 1970's by Elliot Aronson and colleagues. Students in a jigsaw classroom are placed in small groups of mixed sex, race, and academic ability. Each student is given a subset of the information necessary to complete the assignment, and they must pool their resources to achieve a good group grade (for a more complete description, see Aronson & Patnoe, 1997). Students in jigsaw classrooms demonstrate less racism and more interpersonal attraction (Aronson & Patnoe, 1997; Aronson & Bridgeman, 1979). Generally speaking, students in cooperative learning environments become less likely to commit the dispositional attribution errors that help feed into group stereotyping

(Gilbert & Malone, 1995; Stephan, Kenney, & Aronson, 1977). Even single, time-limited experiences with cooperative activities enhance students' ability to take the perspective of others (Tjosvold, Johnson, & Johnson, 1984), which may also be a key component in reducing willingness to demean or unnecessarily compete with members of other groups.[8]

The manipulations used in the educational context are likely to generalize to business and government as well, but in less-formalized situations the role of society may be more prominent in the declaration of acceptable values. Monteith (1993) writes:

> Societal-level changes in the laws and norms concerning people's responses to setereotyped groups have undoubtedly encouraged some people to adopt and internalized low prejudiced attitudes. Such attitudes will become more accessible and more likely to provide a basis for responding if societal institutions repeatedly communicate nonprejudiced messages. Also, negative stereotypes will become less accessible and less likely to provide a basis for responding to the extent that societal institutions avoid negative stereotyped depictions of racial, ethnic, and minority group members (p. 483).

Based on the way that social norms can help activate personal norms, Wegner & Erber (1993) discuss social control as a supplement to the limited reservoir of personal control. Greenwald & Banaji (1995) endorse affirmative action programs on this basis: not only do they offer more opportunities for positive interaction between racial groups, but they also promote a view of regular, cooperative intergroup contact as normal and desirable -- or, at least, unavoidable. This constant reinforcement of egalitarian norms may not force individuals in society to internalize those norms fully, but by keeping them prominent they may help push our dissonance-reduction efforts in the direction of reconciling our automatic thoughts and beliefs with our conscious

[8] There is also evidence that students in cooperative environments perform as well as or better than students in traditional independent or competitive environments (Springer, Stanne, & Donovan, 1999, Johnson, Marumaya, Johnson, & Nelson, 1981). While school performance per se is orthogonal to reducing racism, it may be critical to students' perceiving the interaction as positive and worthwhile, which will determine the affect they will attach to their memories of intergroup contact.

values, rather than changing our conscious values and deciding that prejudice isn't so bad after all.

An Anti-Racism Toolbox: Techniques for Different Situations

The discussion of potential for change affords a primary position to disputation, and I do believe that disputing racist thoughts and beliefs is a necessary component of change. However, it would be unreasonable to use this technique in all situations: paying attention to internal mental processes when trying to interact with another person can be just as disruptive as any unintentional, automatic display of disapproval. Additionally, entering an interaction with a goal based on performance (e.g., "don't do anything offensive") instead of attention to and interest in the other person is directly detrimental to performance (Grant-Pillow & Dweck, 2002). Cognitive disputation is best used in deliberative situations, such as admissions decisions and criminal trials (or, in an everyday context, in determining whether we like someone), or in private rehearsal to help it become automatized. When snap judgments or spontaneous social interaction are important, automatically-activated positive impressions may play a greater role. Real-life positive intergroup experiences are thus just as important as intellectual comprehension of potential biases. When we catch an instance of prejudicial thought or behavior, we can respond by reminding ourselves of successful, enjoyable interactions in the past and move on.

Similarly, the de-stigmatization of racism which I recommended in the section on moral judgment and intentionality may be helpful in quickly discounting racist thoughts, allowing them to be identified as unimportant cognitive errors rather than serious personal failings *or* indicators of genuine belief. In a study on suppression of racist impressions, Foerster & Liberman (2001) led some subjects to believe that suppression should be easy, but warned others that it would be difficult and that occasionally producing racist thoughts is normal. The latter group showed less postsuppressional rebound. The authors viewed this as evidence that the easy-expectancy subjects were attributing racist beliefs to an internal disposition and thus acting consistently with the assumed disposition later; I believe another possible interpretation is that expecting failures of suppression makes intrusive thoughts less interesting, hence they attract less attention and are not further activated by our responding to them. While paying attention to automatic racist thoughts is critical to reducing their occurrence, the proper time is when they can be dealt with in depth, not when our resources are limited and all they can do is distract us.

Future Directions

I have attempted in this paper to bring research on implicit racism together with research on the low-level cognitive processes that underlie implicit thought and cognition of all kinds. This research offers a great deal of explanatory power, but it also raises new questions and leaves many old questions answered only in theory. Following is a list of some research programs that should be undertaken and new theoretical links that remain to be forged:

· *Direct testing of thought disputation*: The techniques of cognitive therapy could be adapted to deal with typical stereotypical beliefs and sources of intergroup conflict. Individuals could be taught about the cognitive biases that lead to stereotyping and prejudice, and assisted in developing and rehearsing responses to racist thoughts when they arise. This would be more difficult than the usual clinical trial: in addition to the usual difficulties of longitudinal studies, we would need to find a population that was in frequent contact with different racial groups and interested in reducing implicit racism. But since most Americans do not believe that they are racist (Devine, Monteith, Zuwerink, & Elliot, 1991), and may be strongly motivated to avoid confronting their implicit racism (Gaertner & Dovidio, 1986), it would be difficult to recruit sufficiently committed participants. A demonstration of the discrepancies between participants' beliefs and their actions (such as the experimental procedures used in Monteith, 1993, and Devine, Monteith, Zuwerink, & Elliot, 1991, or the implicit associations test of Banaji & Greenwald [1997]) might be useful to help educate participants and drive home the value of the intervention. In any case, envrionmentally-valid tests of the techniques presented here will be required to determine their actual usefulness, and ultimately to bring any potential benefits to the public.

· *Generalization*: Most of the research reviewed here has focused on white Americans' attitude towards black Americans. A few studies have extended this their scope to other groups, such as gay men (Devine, Monteith, Zuwerink, & Elliot, 1991) and skinheads (Macare, Bodenhausen, Milne, & Jetten, 1998), and some research has found that measures similar to the modern racism scale predict anti-immigration attitudes in a wide variety of European countries (Meertens & Pettigrew, 1997). However, our theories about the specific cognitive processes underlying implicit racism would be greatly strengthened by testing across cultures and across target minorities within a society. It may be especially important to study Asian cultures which lack some of the dispositional attribution biases common in Western psychology but which still show stereotyping, prejudice, and discrimination against social out-groups.

· *Evolutionary / Modular viewpoints*: Evolutionary psychology predicts that tasks which are basic to survival will show some or all of the properties of automatic processes (lack of intention, lack of conscious awareness, efficiency, and difficulty in inhibition [Bargh, 1994]) and operate as "modules" which are separate from other cognitive processes in terms of their input, processing, or output (Tooby & Cosmides, 1992). If racism is one expression of a cognitive module that learns about the group differences a society constructs and responds to them in standardized ways, we should be able to find certain types of sensory input which preferentially activate it. If there are specific behavioral cues which we unconsciously use to tag another person as an in-group or out-group member, manipulating or selectively ignoring or attending to those cues could have a disproportionate influence on the extent to which our automatic cognitions about race are activated and the ease with which we can dispute or suppress them.

· *Malleability of sensitivity to group information*: Even if the modular model of race-sensitivity is correct, we may still have some amount of influence over the extent to which we attend to information about groups and use that information to make inferences or guide emotional responses to other people. Reducing our sensitivity to group membership, or increasing the inclusivity of our in-group concept, would strike racism at its root, and potentially be a great deal more efficient than laboriously defeating the emotionally powerful inferences we draw once group membership perception has been activated. Some research on positive emotions has been especially promising: Isen has found that people in good moods conceptualize object categories more broadly and are more willing to include valid but unusual exemplars, are less prone to confirmation bias and perseverance with unsuccessful beliefs, and are more effective at problem solving overall (Isen, 2000), These effects translate to group interactions, and have been found to help with the establishment of a common ingroup identity (Dovidio, et al., 1998). Fredrickson's research on cognitive broadening (Fredrickson, 2001) has uncovered a number of similar effects. Very recent results from her lab suggest that cognitive broadening, mediated by positive emotions, can even reduce extremely rapid, implicit effects of race on face perception (Johnson, in preparation). If it turns out that positive emotion, which is relatively easy to induce, can eliminate implicit biases which are resistant to conscious intervention, we will have found a very powerful technique for indirectly controlling implicit racism. This should inspire research concerning the effects of positive emotions on other forms of automatic social cognition.

Conclusion

Reducing institutionalized racism in western society was a prolonged, challenging, and collaborative project, and we should not expect a reduction of implicit racism within individuals to be easier. In the short term postsuppressional rebound, misattribution, and frustration are likely to result. Only with prolonged application can deliberate control processes begin to alter the automatic thoughts and responses that determine much of our daily behavior. However, the view of implicit racism I have offered has a hopeful message as well: modern observations of stereotyping, prejudice, and discrimination need not be taken to indicate that nothing has changed in the past century, or that people who act in racist ways are just as hateful and fearful as their racist predecessors, and deceitful as well. The cognitive processes behind racism, while compelling and universal, are still understandable and open to deliberate manipulation. The proper way to produce reform is not moral condemnation or open racial conflict, but gradual research and education, combining findings from clinical, social, and cognitive psychology, as well as broader models from the other social sciences. In this respect I hope that our understanding of implicit racism can serve as a model for future efforts to regulate ingrained behaviors, experiment with social organization, and enhance human happiness and functioning.

References

Allport, Gordon (1954). *The Nature of Prejudice.* Cambridge, MA: Addison-Wesley Publishing Company, 1954, pp. 3-28.

Aronson, E. & Patnoe, S. (1997). *The Jigsaw Classroom : building cooperation in the classroom*. New York : Longman

Aronson, E., & Bridgeman, D. (1979). Jigsaw groups and the desegregated classroom: In pursuit of common goals. *Personality and Social Psychology Bulletin, 5(4),* 438-446.

Bargh, J. (1994). The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In R. Wyer, Jr. & T. Srull (Eds.), *Handbook of Social Cognition (2 ed).* (pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum.

Bargh, J., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71(2)*

Bartmess, E. (2003, June 10th). Presented at University College, London, England.

Baumeister, R., Bratslavsky, E., Muraven, M, & Tice, D. (1998). "Ego Depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74(5),* 1252-65.

Chen, M. & Bargh, J. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology 33(5),* 541-60.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin, 87(3),* 546-63.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology. 56(1),* 5-18.

Devine, P., Monteith, M., Zuwerink, J., & Elliot, A. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60(6),* 817-30

Dijksterhuis, A. & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology, 74(4)*, 865-877.

Dovidio, J.F., Gaertner, S.L., Isen, A.M., Rust, M. & Guerra, P. (1998). Positive affect, cognition, and the reduction of intergroup bias. In C. Sedikides, J. Schopler, & C.A. Insko (Eds.), Intergroup cognition and intergroup behavior (pp. 337-366). Mahway, NJ: Erlbaum.

Dweck, C. (1999). *Self-Theories: Their Role in Personality, Motivation, and Development.* Philadelphia, PA: Psychology Press.

Fiske, S. (1998). Stereotyping, Prejudice, and Discrimination. In D. GIlbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology, Volume II, Fourth edition,* (pp. 357-411). New York: McGraw-Hill.

Foerster, J. & Liberman, N. (2001). The role of attribution motivation in producing postsuppression rebound. *Journal of Personality and Social Psychology, 81(3)*, 377-90.

Fredrickson, B. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions.Ê*American Psychologist, 56(3),* 218-226.

Gaertner, Samuel L. and John F. Dovidio "The Aversive Form of Racism," in *Prejudice, Discrimination, and Racism*, edited by John F. Dovidio and Samuel L. Gaertner Orlando: Academic Press, 1986, pp. 61-89.

Gerrig, R. & Prentice, D. (1991). The representation of fictional information. *Psychological Science, 2(5),* 336-40.

Gilbert, D. (1993). The assent of man: Mental representation and the control of belief. In J. Pennebaker & D. Wegner (Eds.), *Handbook of Mental Control,* (pp. 57-87). Englewood Cliffs, NJ: Prentice-Hall.

Gilbert, D., & Malone, P. (1995). The correspondence bias. *Psychological Bulletin, Vol 117(1),* 21-38.

Gilbert, D., Tafarrodi, R., & Malone, P. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology, 65(2),* 221-33.

Goleman, D. Useful modes of thought contribute to prejudice. (1987, May 12). *The New York Times,* pp. 17-20.

Grant-Pillow, H. & Dweck, C. (2002). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85(3).*

Gray, J. (1990). "Brain systems that mediate both emotion and cognition". *Cognition and Emotion, 4(3),* 269-88.

Greenwald, A. G. and M. R. Banaji "Implicit Social Cognition: Attitudes, Self-Esteem, and Sterotypes." *Psychological Review*, Vol. 102, 1995, pp. 4-27.

Griffin, D. & Kahneman, D. (2003). Judgmental heuristics: human strength or human weakness? In L. Aspinwall & U. Staudinger (Eds.), *A Psychology of Human Strengths: Fundamental Questions and Future Directions for a Positive Psychology*. (pp. 165-178). Washington, DC: American Psychological Association

Higgins, T. & King, G. (1983). Accessiblity of social constructs: processing consequences of individual and contextual variability. In N. Cantor & J. Kihlstrom (Eds.), *Personality, Cognition, and Social Interaction.* (pp. 69-122). Hillsdale, NJ: Lawrence Erlbaum

Hirschfield, L. (1998). *Race in the Making*. Cambridge, MA: MIT Press.
Isen, A.M. (2000). Positive affect and decision making. In M. Lewis & J. Haviland-Jones (Eds.). *Handbook of Emotions*, 2nd Edition (pp. 417-435). NY: Guilford.

Johnson, D, Maruyama, G., Johnson, R., & Nelson, D. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. *Psychological Bulletin, 89(1)*, 47-62.

Johnson, K. (in preparation). Positive emotions eliminate the own race bias in face perception.

Kalma, A. (1991). Hierarchisation and dominance assessment at first glance. *European Journal of Social Psychology, 21,* 165-181.

Kawakami, K., Dovidio, J., Dijksterhuis, A., (2003). Effect of social category priming on personal attitudes. *Psychological Science, 14(4),* 315-319.

Macare, C., Bodenhausen, G., Milne, A., & Jetten, J. (1994). Out of mind but back in sight: stereotypes on the rebound. *Journal of Personality and Social Psychology, 67(5)*, 808-817.

Mahalingam, R. (2002, September). *Culture, essentialism, and marginality: A developmental perspective*. Paper presented at a conference on ÒExploring an epidemiology of success in children and youth who experience social inequalities,Ó University of Michigan.

McNamara, T. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99(4)*, 650-662.

Meertens, R. & Pettigrew, T. (1997). Is subtle prejudice really prejudice? *Public Opinion Quarterly, 61(1),* 54-71.

Monteith, M. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology, 65(3),* 469-85.

Morris, J., Squires, N., Taber, C., & Lodge, M. (2003). Activation of political attitudes: A psychophysiological examination of the hot cognition hypothesis. *Political Psychology, 24(4),* 727-737.

Neely, J. (1977). Semantic priming and retrieval from lexical memory: Roles of inihibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106(3),* 226-254.

Pennebaker, J.W., Zech, E., & RimŽ, B. (2001). Disclosing and sharing emotion: Psychological, social and health consequences. In M.S. Stroebe, R.O. Hansson, W. Stroebe, & H. Schut (Eds.), *Handbook of bereavement research: Consequences, coping, and care* (pp 517-544). Washington, DC: American Psychological Association.

Pratkanis, A., Leippe, M., Greenwald, A., & Baumgardner, M. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead. Long live the sleeper effect. *Journal of Personality and Social Psychology, 54(2),* 203-218.

Rosenthal, R. (2003). Covert communication in laboratories, classrooms, and the truly real world. *Current Directions in Psychological Science, 12(5),* 151-155.

Ross, Lee. (1977). The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process. *Advances in Experimental Social Psychology, (10),* 173-220.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69(5),* 379-399.

Schul, Y. & Burnstein, E. (1985). When discounting fails: conditions under which individuals use discredited information in making a judgment. *Journal of Personality and Social Psychology, 49(4),* 894-903

Segal, Z., Gemar, M., Williams, S. (1999). "Differential Cognitive Response to a Mood Challenge Following Successful Cognitive Therapy or Pharmacotherapy for Unipolar Depression". *Journal of Abnormal Psychology 108(1),* 3-10.

Shiffrin, R. Attention. (1988). In R. Atkinson, R. Herrnstein, G. Lindzey, & R. Luce, (Eds.), *Stevens' Handbook of Experimental Psychology* (pp. 739-812). New York: Wiley

Springer, L., Stanne, M., & Donovan, S. (1999). Effects of small-group learning on undergraduates in science, math, engineering, and technology: A meta-analysis. *Review of Educational Research, 69(1),* 21-51

Stephan, C., Kennedy, JC, & Aronson, E. (1977). The effects of friendship and outcome on task attribution. *Sociometry, 40,* 107-111

Tajfel, H & Turner, J. (1986). The Social Identity Theory of Intergroup Behavior. In S. Worchel (Ed.) *The Psychology of Intergroup Relations.* (pp. 7-24). Chicago: Nelson-Hall

Teasdale, J., Scott, J., Moore, R., Hayhurst, H., Pope, M., Paykel, E. (2001). "How Does Cognitive Therapy Prevent Relapse in Residual Depression? Evidence From a Controlled Trial". *Journal of Consulting and Clinical Psychology, 69(3),* 347-357

Tjosvold, D., Johnson, D., & Johnson, R. (1984). Influence strategy, perspective-taking, and relationships between high- and low-power individuals in cooperative and competitive contexts. *Journal of Psychology, 116,* 187-202

Tooby, J., & Cosmides, L. (1992). *The Psychological foundations of culture.* In J. Barkow, L. Cosmides, & J. Tooby, (Eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* New York: Oxford University Press.

Walker, I., & Crogan, M. (1998). "Academic performance, prejudice, and the jigsaw classroom: New pieces to the puzzle". *Journal of Community and Applied Social Psychology, 8,* 381-93.

Wegner, D. & Bargh, J. (1998). Control and automaticity in social life. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology, (4th ed.)* New York: McGraw-Hill.

Wegner, D. & Erber, R. (1993). Social foundations of mental control. In D. Wegner & J. Pennebaker (Eds.), *Handbook of Mental Control*. (pp. 36-56). Englewood Cliffs, NJ: Prentice-Hall.

Wegner, D. & Smart, L. (1997). Deep cognitive activation: A new approach to the unconscious. *Journal of Consulting and Clinical Psychology, 86(6),* 984-995

Wegner, D. (1990). *White Bears and Other Unwanted Thoughts: Suppression, Obsession, and the Psychology of Mental Control.* New York: Penguin Books

Wegner, D., Coulton, G., & Wenzlaff, R. (1985). The transparency of Denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology, 49*, 382-91

Wenzlaff, R., & Wegner, D. (2000). Thought Suppression. *Annual Review of Psychology, 51,* 59-91.

Whisman, M. (1993). "Mediators and moderators of change in cognitive therapy of depression" *Psychological Bulletin, 114(2),* 248-65.

Zanna, M & Cooper, J. (1976). Dissonance and the Attribution Process. In J. Harvey, W. Ickes, & R. Kidd (Eds.). *New Directions in Attribution Research* (pp. 199-221). Hillsdale, NJ: Lawrence Erlbaum Associates.

Appendix: A note on language

This paper contains some fairly obvious biases in its use of language. I frequently refer to the audience with the term “we,” indicating that I am targeting my writing towards people who are typically the agents of racism and only rarely its victims. However, I feel that attempting to use consistently sterilized third-person language referring to “whites” or “majority-group members” would cost a great deal in terms of style and implied applicability to readers. I would welcome advice on how to properly navigate this situation.