ACADEMIC PRESS

Journal of Experimental Social Psychology 38 (2002) 384–396

Journal of Experimental Social Psychology

www.academicpress.com

Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons

B. Keith Payne,* Alan J. Lambert, and Larry L. Jacoby

Department of Psychology, Washington University, One Brookings Drive, Campus Box 1125, St. Louis, MO 63130, USA

Received 2 May 2001; revised 5 November 2001

Abstract

This study applied process dissociation (PD) to investigate the role of conscious goals in controlling automatic influences of stereotypes. A priming procedure was used to show that the presence of Black (vs. White) faces caused stereotypical misidentifications of objects. Specifically, Black primes caused lures to be misidentified as weapons, whereas White primes caused weapons to be misidentified as non-threatening objects. By manipulating the goals with which participants completed the experiment, we demonstrated that the stereotype bias was invariant across conscious goals to avoid vs. use the influence of race. PD analysis provided three major insights. First, the impact of race was mediated solely through an unintentional accessibility bias. Second, requiring participants to respond rapidly increased the impact of stereotypes, an effect that was mediated by a reduction in controlled discrimination among stimuli. Finally, calling attention to race increased the stereotype accessibility bias, regardless of whether race was made salient with the intention to use, or the intention to avoid the influence of race.

Keywords: Prejudice; Stereotypes; Automatic processes; Cognitive control; Goals

Research showing the ubiquity of unconscious prejudice is both intriguing and disturbing. The fact that people who explicitly espouse egalitarian values may, nonetheless, be prejudiced carries the specter that anyone might be an implicit bigot without the power to know or control his or her own biases (e.g., Banaji & Greenwald, 1995; Bargh, 1999; Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995). The extent to which unconscious and automatic cognitions are controllable is currently a matter of lively discussion (e.g., Bargh & Gollwitzer, 1994; Blair & Banaji, 1996; Devine, 1989). Although it is a good question how much of everyday life is actually under mindful control (Bargh, 1999), there appears to be some consensus that people can sometimes restrain their automatic tendencies, given ample motivation and opportunity (Fazio, 1990; Fazio et al., 1995; Devine & Monteith, 1999; Greenwald & Banaji, 1995; Lepore & Brown, 1997).

In this paper, we take a new approach to study the role of goals in controlling the automatic influence of stereotypes. We use the process dissociation procedure which can decompose judgments into controlled and automatic components (Jacoby, 1991). This approach is especially valuable for our purposes because it allows us to track the influence of racial stereotypes and higher-order goals on each component separately. Moreover, this procedure allows us to address several issues with a degree of precision that would not have been possible using more traditional strategies for measuring automatic and controlled mechanisms.

A recent article by Payne (2001) used the process dissociation procedure to investigate the processes underlying stereotype-based inferences. The goal of that research was to understand how the mere presence of a Black face could cause people to misidentify harmless objects as weapons. Specifically, when a lure object (e.g., a tool) was immediately preceded by a Black face, it was more likely to be mistaken for a handgun, compared to conditions in which this same object was preceded with a White face. This stereotypic difference emerged mainly when participants were required to respond rapidly, a condition that mimics the time pressure involved in actual police confrontations.

* Corresponding author. Fax: +1-314-935-7588.
E-mail address: bkpayne@artsci.wustl.edu (B.K. Payne).

0022-1031/02/$ - see front matter
PII: S0022-1031(02)00006-9

On one hand, the fact that mere presence of a Black vs. White face produced systematic errors (e.g., misidentifying a hand tool as a gun) is of theoretical and practical interest in its own right. However, the most notable aspect of Payne (2001) was use of process dissociation to show that the biasing effect of race was isolated in the automatic, rather than the controlled, component. (We shall describe below in more detail the manner in which process dissociation is able to yield this kind of information.) These findings are important in that they generated more direct evidence than has previously been available that the impact of stereotypes on perceptions can operate in an automatic fashion, independent of perceivers' intentions (see Hense, Penner, & Nelson, 1995, for an application of process dissociation to memory for age stereotypes).

However, Payne (2001) left unresolved several important theoretical issues. First, the study did not manipulate participants' higher-order goals, such as the goal to behave without prejudice. As such, it remains important to determine whether, and under what conditions participants could eliminate the impact of race on their judgments if they were explicitly motivated to do so. This question is motivated by both theoretical interest and applied concerns. If the effect of race in biasing perceptual judgments depends on people's goals, then the effect reported by Payne (2001) would carry quite different theoretical implications than if the effect was found to occur regardless of participant's goals. On the practical level, if race commonly and automatically affects the perceptual judgments that people (e.g., police officers) make, then this effect may have serious consequences for minority groups. It is therefore important to understand whether and how perceivers' goals might moderate the impact of race on those judgments. By measuring the accuracy of participants' judgments and applying process dissociation analysis, the present approach sheds light on both the practical consequences of racial stereotypes, and the mechanisms that give rise to those effects.

In the following section, we will first discuss recent research concerning the intentional control of automatic influences, and the questions that work raises about the nature of stereotype biases. Next, we will briefly discuss the logic of the process dissociation approach. We argue that it may help clarify the unresolved issues arising from the conceptual and methodological limitations in studying automatic processes and cognitive control. We then apply the logic of that approach to test the impact of conscious goals on race-based errors.

Goals and the control of automatic influences

One essential feature of the automatic processes is that they can operate unintentionally (Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Uleman (1999) has noted a useful distinction between two meanings of "unintentional" that may help to avoid confusions in terminology. First, unintentional can mean *not intending* to do something, reflecting a simple absence of intention. This sense of the word corresponds to what Uleman calls "spontaneous" processes. For example, imagine a professor walking across the campus deep in thought. Suddenly, he realizes that he is talking aloud to himself. Blushing, he goes on and continues his monologue internally. This is a case of speaking while *not intending* to. A second use of "unintentional" means *intending not* to do something. If a behavior is unintentional in this sense, it is truly *against* a person's will. Now imagine our professor has just noticed that he is talking to himself. But this time, when he attempts to stop, he finds he cannot stop no matter how hard he tries. This hypothetical case of verbal incontinence would be a case of speaking while *intending not* to. The first situation may end in slight embarrassment; the second may end in a psychiatrist's office. Behaviors executed against the explicit intentions of the actor represent a degree of automaticity that is different than a simple lack of intention.

Many of the phenomena labeled automatic in social cognition research are unintentional in the first sense, i.e., simply not intending. Priming paradigms in which word completions or response latencies for lexical decisions are measured, for example, are subtle procedures that capitalize on the fact that the associations being measured may not depend on the participants' intentions. Less is known about the processes that take place when people actively try to avoid unintentional influences (intending *not* to).

The intentional control of automatic biases has received some attention in stereotyping research because, although stereotypes are prevalent, they are often seen as socially inappropriate and inconsistent with egalitarian values (Gaertner & Dovidio, 1986). However, research on this topic has generally measured the effects of goals to avoid stereotypes on outcome judgments, without directly addressing the underlying mechanisms (Devine & Monteith, 1999). When people are able to successfully overcome their automatic tendencies toward stereotyping and make an even-handed judgment, is it because the automatic process has been halted? Or is it because the output behavior has been severed from the automatic influence, and based on a more controlled process?

Process dissociation analysis of discriminability and bias

The present research used a modification of the weapons-identification paradigm described above, coupled with the process dissociation procedure, to clarify these and other related matters. Process dissociation decomposes performance on a task into two estimates: discriminability and accessibility bias. The procedure is similar to other models such as signal detection theory (Green & Swets, 1966) in that it yields two independent parameters that jointly contribute to performance. Signal detection relies on a single dimension (strength of the perceptual signal or memory trace) whereas process dissociation identifies the two estimates with separate processes, which depend on distinct bases for responding.

In some cases, the bias parameter can be characterized as an estimate of automatic processing, and the discriminability parameter as an estimate of cognitive control. Payne (2001) characterized the bias estimate in the weapon identification paradigm as automatic because it (a) influenced responses regardless of whether it facilitated or interfered with task performance, and (b) occurred rapidly (within a stimulus-onset asynchrony of 200 ms) and was invariant across different levels of processing time participants were allowed. Conceptually, the bias estimate in this paradigm can be defined as a tendency to respond in a stereotypical way when a person is unable to exert control over responses. In terms of the paradigm used by Payne (2001), this reflected the tendency for participants to be more likely to respond "gun" if they had just been primed with a Black rather than a White face.

The discriminability estimate, in contrast, was used to operationalize cognitive control because it (a) was defined by the ability to respond in accordance with task instructions, and (b) was sensitive to the manipulation of processing time. Discriminability may be conceptualized here as the ability to respond based on the objective features of the stimulus. That is, discriminability is the ability to respond "gun" to guns and "tool" to tools, corrected for systematic biases. The race of the faces selectively influenced the accessibility bias parameter, but not discriminability, suggesting that the stereotypical pattern of errors was driven by automatic racial bias.

As in other applications of the process dissociation procedure, analysis of responses in this paradigm requires both *congruent* conditions and *incongruent* conditions. For congruent conditions, automatic, and controlled processes act in concert to produce the same response. For incongruent conditions, however, the two processes oppose one another, driving *opposite* responses. In the paradigm used by Payne (2001), the computer first presented a Black or White face followed by a target object that was either a handgun or a hand tool. Participants were instructed to identify the target object as either a gun or a tool by pressing one of two keys. Here, automatic stereotypes and controlled responding based on the actual target are the two processes that are either matched or mismatched in their expected effects.

Congruent conditions are the White–tool pairs and the Black–gun pairs. In these cases, responding either on the basis of the racial category *or* on the basis of the actual target will lead to the correct answer. For example, when a black face is followed by a gun, either responding based on racial stereotypes or the objectively correct information would lead to the response "gun." The incongruent conditions include the White-gun and Black-tool pairs. In these conditions, the racial stereotypes and the target object should lead to contradictory responses. For example, when a black face is followed by a tool, race stereotypes would lead to an erroneous "gun" response, while the objective information would lead to the "tool" response.

When both sources of information favor the same response, then correct responses can result from either controlled discrimination (D) of the target, or a stereotypical accessibility bias (A) when target discrimination does not occur $(1-D)$. This relationship may be expressed mathematically in the equation: $P(\text{correct} \mid \text{congruent}) = D + (1-D)A$. When a black face is paired with a gun, this equation shows that the correct response, "gun" can be consistently achieved in two ways. The first way is by successfully controlling the response. The second is by unintentionally responding "gun" because of the race prime, even when unable to respond based on the actual target. Because on congruent trials discriminability and accessibility bias are perfectly confounded, it is impossible to tell which process drives responses from these trials alone.

On incongruent trials, the two processes are not only unconfounded, but set in opposition to one another. When controlled discrimination and stereotypic bias are opposed to one another, false alarms occur when an accessibility bias (A) operates in the absence of target discrimination $(1-D)$. Mathematically, this can be written $P(\text{stereotypic error} \mid \text{incongruent}) = (1-D)A$. As an example, consider the trials in which a black face is paired with a tool. If discriminability fails and a participant is influenced by accessible race information, then he or she should respond "gun".

These two sets of equations may be used to derive estimates of each process. Discriminability is solved algebraically as the difference between correct responses in the congruent condition, and errors in the incongruent condition: $D = P(\text{correct} \mid \text{congruent}) - P(\text{stereotypic } error \mid \text{incongruent})$. To see that this follows from the equations above, recall that $P(\text{correct} \mid \text{congruent}) = D + (1-D)A$ and $P(\text{stereotypic error} \mid \text{incongruent}) = (1-D)A$. Thus, subtracting $P(\text{stereotypic error} \mid \text{incon}\,gruent)$ from $P(\text{correct} \mid \text{congruent})$ provides an estimate of D, the discriminability parameter. This formulation is analogous to correcting for guessing in recognition memory by subtracting false alarm rates from hit rates. To keep with our example, it amounts to subtracting the false "gun" responses after a black prime and a tool from the true "gun" responses after a black prime and a gun. The incongruent trials are used to "unconfound" processes in the congruent trials. Conceptually, discriminability is the extent to which participants are able to respond as they intend to, based on the actual targets and the task instructions.

Given the estimate of discriminability, one can solve for the accessibility bias estimate. Recall that the probability of mistaking a tool for a gun after a black prime is $P(\text{stereotypic error}\,|\,\text{incongruent}) = (1 - D)A$. To solve for A, we divide the false alarm rate in the incongruent condition, $(1 - D)A$ by the probability of a failure of discrimination, $(1 - D)$. Thus, $A = P$ (stereotypic error | incongruent)/(1 − D). In this way, the pattern of hits and false alarms can be decomposed into estimates of the ability to control responses based on target objects, and the bias to respond in a stereotypical direction when control fails. (For further discussion of the process dissociation procedure, and its applicability to stereotyping research, see Payne, Lambert, & Jacoby, 2002.)

Goals of the present experiment

The experiment reported addressed two main questions. First, it is important to know how "unintentional" the stereotypic false perception effect is. Does it simply occur in the absence of an intention, or is it more pervasive, so that it proceeds regardless of one's goals? To investigate this question, some participants were made aware of the potentially biasing effect of racial information, and assigned different explicit goals. Participants were assigned to one of the three goal conditions. The control condition was given no explicit goal, and no mention was made of race. In the "avoid race" condition, participants were informed that previous research had shown that the race of the faces influences people's judgments, and they were asked to make sure their responses were not biased by the race of the primes. In the "use race" condition, participants were also informed about the potentially biasing effect of race, and asked to play the role of a "racial profiler" who used race to help identify possible weapons.

If the effect of race on the identification of targets is unintentional only in the sense that it occurs in the absence of explicit intentions, then the racial primes should influence responses in the control and "use race" condition but not in the "avoid race" condition. However, if the effect is unintentional in the sense that the racial primes influence performance regardless of goals to the contrary, then the effect should be evident in all the three conditions. Note that both assigning the goal to avoid racial stereotypes, and assigning the goal to use racial stereotypes necessarily entails calling participants' attention to race. Thus, both the "avoid race" group and the "use race" group were similar in that their attention was directed toward the influence of the racial primes. The control group, however, was not induced to give special attention to race. The comparison between the control group on the one hand, and the two goal groups on the other, allows an additional test of the effect of *attention to race*, independent of the *reason* for which participants were attending to it.

The second major question was how these different goals affect the distinct processes of discrimination and accessibility bias. Process dissociation analysis offers more direct evidence than has previously been available about the processes underlying the control of automatic stereotype influences. In particular, if participants are able to alter the effect of race on their perceptual judgments in some conditions, then it is important to know *how* those changes are brought about. The processes of discriminability and accessibility bias are assumed to be independent, not mutually exclusive (see Hintzman & Curran, 1997; Jacoby, Begg, & Toth, 1997 for a discussion of the assumptions underlying the procedure). As a result, decreased use of accessible information does not necessarily mean increased control.

One potential mechanism for reducing the impact of race is inhibiting the automatic processing of racial cues. If the racial bias obtained in this paradigm is characterized by goal-dependent automaticity, then there should be a lower accessibility bias estimate when participants aim to avoid race, compared to when they aim to use race.

A second potential mechanism lies in maximizing cognitive control by increasing discrimination between target items. A goal implemented in this way would be reflected in the discriminability estimate, with greater discrimination among items when participants intend to avoid racial considerations. It is worth considering this mechanism in more conceptual detail to clarify what control means here. By control, we mean responding in the way that one intends to, according to the task at hand. In this paradigm, the task requires discrimination between weapons and lures. Similarly, in a real confrontation, the important aspect of control is to respond based upon a clear and present threat rather than "losing control" and responding based on a snap reaction.

One might expect the goals that participants have should affect the controlled component. At the same time, the goal to respond without prejudice is a higher-order goal than to identify objects based on their features. Cognitive control in our approach is defined by this lower-order goal of task requirements. One way to achieve that higher-order goal would be to invest greater effort in successfully carrying out the basic task instructions by discriminating between the items. However, the goal to avoid or use stereotypes does not necessarily entail a change in cognitive control as it is defined here. As such, the processes that result from higher-order goals present an interesting question that is a primary motivator of this research.

In addition to creating congruent and incongruent conditions using racial primes and targets, the present experiment manipulated processing opportunity using a response deadline. The deadline manipulation provided a means for validating the assumption that the accessibility bias estimate reflects automatic processing, whereas the discriminability estimate reflects controlled processing. If this assumption is valid, then the bias estimate should remain *invariant* as processing time decreases, but the discriminability estimate should decline.

Methods

Participants and design

Ninety-seven non-Black undergraduates (66 females, 31 males) participated in return for course credit. Because analyses found no main effects or interactions involving participants' sex, all analyses to follow are collapsed over this factor. The main independent variables included the goal of the participant (control, ignore race, and use race), race of the prime (Black vs. White), the type of target (gun vs. tool), and response deadline. Because the dynamics of automatic and controlled processes within this time range are not well known, it was deemed useful to measure the same participants' performance at multiple levels of response deadline. To track the changes in processing over different processing times, three levels of deadline were chosen: 700, 450, and 200 ms. The task began at the longest deadline, and progressed in blocks to shorter deadlines. In sum, the design of the experiment was a 2 (Race) × 2 (Target) × 3 (Deadline) × 3 (Goal) factorial with goal manipulated between participants, and all other variables manipulated within participants.

Procedure

During preliminary instructions, participants were informed that we were interested in perceptions of both people and of inanimate objects. All participants were then informed, "in this experiment you will see pairs of pictures presented briefly. The first picture will be a face. The second picture will be either a gun or a hand tool. Your job is to respond to the *second picture* by deciding whether it is a gun or a tool." They were also instructed that they should be prepared to respond within a relatively short, pre-determined interval as indicated by the presence of a red response cue and that the pace would become faster as the task went on. (See next section for the exact nature of this priming task.) These were the only instructions given to the control group. Of particular importance is the fact that, at this point, participants were given no advance indication that the study was specifically concerned with race. In contrast, the "avoid race" group and the "use race" group were both given the following additional instructions to alert them to the presence of racial cues:

> The faces will be from either White (European American) or Black (African American) people. Research has shown that the race of the face sometimes impacts the ways that people classify the second object. People are sometimes faster and more accurate in responding to guns after a Black face than after a White face.

Participants in the "avoid race" group were then instructed:

> You have been randomly assigned to take the perspective of a completely unbiased person. Regardless of your personal views, we would like you to base your responses only on whether the second object looks more like a gun or tool. Try not to let the race of the face influence your decisions.

Participants in the "use race" group were told:

> You have been randomly assigned to the "racial profiling" condition. Regardless of your personal views, we would like you to play the role of someone engaged in racial profiling. That is, try to make correct classifications, but we would like you to use the race of the faces to help you identify the gun or tool in question.

As can be seen from these instructions, these latter two conditions differ from the first in that they make explicit mention of race. In other words, although participants in these latter two conditions are being given "opposite" instructions (to avoid using vs. use race), they are similar in that race is, in itself, being raised as a salient issue in a way that is not true in the control condition. This fact shall become relevant in the context of the results to be reported.

Priming task

Prime and target stimuli were digitized photographs 5.3 × 4 cm in size. Prime photos included four Black and four White faces, including two male and two female faces of each race. Target photos included four handguns with varying features, and four hand tools (e.g., wrench, and pliers). On each trial the prime face was presented for 200 ms in the center of the screen. It was replaced immediately by the target picture (a gun or a tool), which remained on the screen for 100 ms. Pilot testing showed that the prime and target pictures were visible to all participants at this speed. The target was followed immediately by the response probe. The probe was a 3.5 × 3.5 cm cross that remained red for 700, 450, or 200 ms. After that interval, the probe became black and remained on the screen for 500 ms. Responses were still recorded during the dark phase of the response probe. Thus, the effective deadline ranged from 1200 to 700 ms. However, participants were not told that their

Table 1
The mean proportions of errors by Prime race, Target item, Deadline, and Goal conditions (*SD*'s in italics)

Goal condition										
Prime	Item	Control ($n = 35$)			Avoid race ($n = 30$)			Use race ($n = 32$)		
		Deadline (ms)								
		700	450	200	700	450	200	700	450	200
Black	Gun	0.058	0.110	0.210	0.035	0.120	0.180	0.043	0.130	0.170
		(0.051)	*(0.087)*	*(0.100)*	*(0.046)*	*(0.100)*	*(0.080)*	*(0.049)*	*(0.090)*	*(0.077)*
	Tool	0.056	0.130	0.210	0.060	0.150	0.220	0.070	0.180	0.240
		(0.047)	*(0.110)*	*(0.120)*	*(0.061)*	*(0.130)*	*(0.110)*	*(0.130)*	*(0.160)*	*(0.120)*
White	Gun	0.049	0.140	0.240	0.041	0.170	0.210	0.074	0.190	0.220
		(0.046)	*(0.089)*	*(0.100)*	*(0.040)*	*(0.140)*	*(0.090)*	*(0.087)*	*(0.140)*	*(0.120)*
	Tool	0.050	0.110	0.200	0.047	0.130	0.190	0.046	0.130	0.170
		(0.048)	*(0.085)*	*(0.110)*	*(0.049)*	*(0.100)*	*(0.087)*	*(0.043)*	*(0.096)*	*(0.091)*

responses would be accepted after the probe had darkened. Immediately after the blackened response probe disappeared, the next trial began.

All participants received 64 practice trials with a response deadline of 800 ms to help them respond within the deadline for critical trials. Each practice trial presented a prime, a target, and the response probe as in the critical trials, but the responses for these trials were not used. The critical trials were arranged in three blocks in which the response interval decreased between each block.[1] Between blocks, participants were allowed a brief rest. The computer provided instructions to participants that the next block would require faster responses than the previous block, and reminding them to respond while the probe was red. Within each block, each of the eight primes was paired with each of the eight targets twice, yielding 128 trials per block. Over three critical blocks, this yielded a total of 384 observations per participant. The pairings were presented in a new randomized order for each participant.

Results

Analyses of identification errors

The most interesting aspect of these data for the present purpose is the errors: the misidentification of tools as guns, or of guns as tools. For this reason, results are reported as error rates rather than accuracy. (Accuracy is simply the complement of the error rate.) First error results will be described, followed by analyses of discriminability and bias estimates. As a preliminary analysis, a Prime race × Target × Deadline × Goal ANOVA was performed on error rates. This analysis revealed a significant main effect for Deadline, a two-way Prime race × Target interaction, a Prime race × Target × Deadline interaction, and a Prime race × Target × Goal interaction, all p's $< .05$. Table 1 provides the proportions of errors and their standard deviations for each experimental condition. Because each of these effects raises a distinct set of issues in its own right, the implications of each will be considered in turn below.

Effects of deadline

The main effect of deadline indicated that errors increased linearly as processing time decreased, $F(2, 188) = 172.38$, $p < .001$. This effect was true for the control group (M's = .05, .12, .22 for the long, medium, and short deadlines, respectively), as well as the "avoid race" group (M's = .05, .14, .20) and the "use race" group (M's = .06, .16, .20). In other words, collapsed over all other factors, participants were less able to base their responses on the actual targets as the response deadline decreased. This result validates the assumption that participants would have less control over their responses as processing time was depleted.

Effects of race

The predicted pattern of race-stereotypic misidentifications was indicated by a highly significant Prime race × Target interaction, $F(1, 93) = 37.08$, $p < .001$. This pattern took the form of a symmetrical cross-over interaction, in which tools were misidentified as guns more often after a Black prime ($M = .15$) than after a White prime ($M = .12$), and guns were misidentified as tools more often after a White prime ($M = .15$) then after a Black prime ($M = .12$). For ease of display, this

[1] Employing a fixed order for the deadline from long to short produces the problem that deadline is confounded with time. However, counterbalancing the order of different response deadlines is problematic, because participants who experience a fast deadline before a slow deadline may continue responding in the hurry to which they have become accustomed. We do not believe this potential confound compromises our results. The reason is that fatigue over time, and progressively shortened deadlines are both likely to impair control. Further, Payne (2001) used a between-subjects manipulation of deadline and found results that parallel the present findings. It was judged that the benefit of tracking the processing dynamics within participants over time outweighed the limitation of order.

effect is shown in Fig. 1 as stereotype-congruent errors (e.g., mistakenly calling a tool a gun when primed with a Black face, or mistakenly calling a gun a tool when primed with a White face) and stereotype-incongruent errors (Table 1 shows all the component means). Overall, stereotype-congruent errors were more likely than stereotype incongruent errors, and this was true in all the three goal conditions.

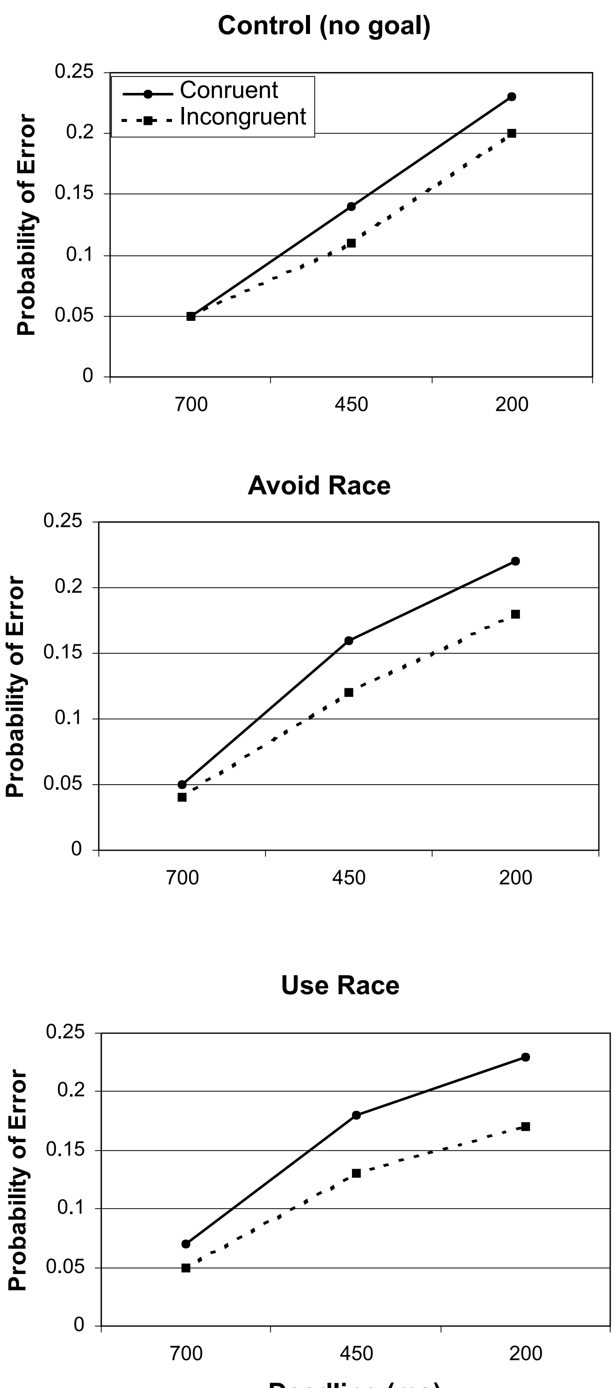

Fig. 1. Stereotype-congruent and stereotype-incongruent errors by deadline and goal conditions.

The difference between stereotype-congruent and incongruent errors became greater in each of the three goal groups as processing time decreased, as shown in the Prime race $\times$ Target $\times$ Deadline interaction, $F(2, 186) = 3.93$, $p < .02$. This finding is important because it shows convergence with other research demonstrating that stereotypes have greater impact on judgments when processing resources are limited (e.g., Macrae, Milne, & Bodenhausen, 1994).

Effect of goals on race-based errors

One of the critical questions in this paper concerned whether participants' goals would moderate the impact of the racial primes on their perceptual judgments of the targets. In particular, would the pattern of stereotype-congruent versus incongruent errors vary as a function of the goal set to which participants were assigned? The difference between stereotype-congruent and incongruent errors did, in fact, change across the three goal conditions, as reflected by a significant Prime race $\times$ Target $\times$ Goal effect, $F(2, 93) = 3.88$, $p < .02$.

Simple effect tests were performed to clarify the nature of this interaction. On the one hand, and as can be readily seen in Fig. 1, participants made more stereotype-congruent than incongruent errors in all three groups collapsed over response deadline and this was true in all three goal sets, all *p*s $< .01$. Nevertheless, the *relative* size of these effects varied across goal condition. Specifically, post-hoc Tukey tests showed that the difference between stereotype-congruent and stereotype-incongruent errors in the "use race" condition (.162 vs. .117; difference = .045) was greater than that in the control condition (.138 vs. .123; difference = .015), $p < .02$. The stereotypic error difference for the "avoid race" group (.143 vs. .116; difference = .027) was intermediate between the two, and not significantly different from either, $p > .26$.

One of the most important aspects of these findings is that, compared to the control condition, participants aware of and motivated to avoid using stereotypes were unable to reduce the impact of racial primes on judgments. However, this conclusion does not rest on an argument from a null finding. Indeed, instructions to avoid using race *increased* the extent to which participants made more stereotype congruent versus incongruent errors (see Fig. 1). One interpretation of these findings is that both the "avoid race" and the "use race" groups were alerted to the presence of racial cues that could influence their judgments, while the control group was not.

To test this interpretation, we compared the control group with both goal groups. This comparison tests the effect of making race salient, independent of the goal with which participants attended to race information. For this comparison, a focused contrast was carried out to compare the size of the difference between stereotype-congruent and stereotype-incongruent errors in the control group to the other two groups. Results indicated that the average stereotypic difference in errors across the two race-salient goal groups (.153 vs. .116; difference = .037) was greater than in the control group (.138 vs. .123, difference = .015), $F(1,95) = 5.16$, $p < .03$. Thus, making race salient increased the tendency to stereotypically misidentify objects, regardless of whether race was focused on with the intent to avoid its influence, or the intent to employ it.[2]

Summary of error analysis

Results showed that aware and motivated participants were able to avoid the influence of the racial primes on their identifications of the targets only when adequate processing time was available. As processing opportunity was depleted, however, all groups performed in a stereotypical manner, misidentifying hand tools as guns after a Black face, and misidentifying guns as tools after a White face. Although participants were unable to avoid the influence of racial stereotypes, participants' goals did affect their responses. Ironically, the effect of making race salient–even when race was highlighted as a means to prevent being biased by it–was to create an even more stereotypical pattern of mistaken identifications. In order to generate further evidence in support of this interpretation, and to examine the underlying mechanisms by which these differences operated, process dissociation analyses were conducted.

Process dissociation analyses

Estimates of discriminability and bias were calculated for each condition.[3] In the following sections, the effects of each variable will be reported for the discriminability estimate, and next for the bias estimate. For each of these analyses, a Prime race × Deadline × Goal condition ANOVA was used. Because Target type was used to compute discriminability and bias estimates, it is not included as an independent variable in these models.

Discriminability

Higher values for the discriminability estimate reflect greater control over responses based on the actual target items. Fig. 2 displays discriminability estimates broken down by Deadline, Goal group and Prime race. The deadline interval produced a strong effect on the discriminability estimate, $F(2,188) = 172.38$, $p < .001$, indicating that as processing opportunity decreased, control was reduced.

Importantly, discriminability was *not* affected by: (a) prime race, $F(1,94) < 1$, *ns*., (b) goals, $F(2,94) < 1$, *ns*, nor by (c) the interaction between prime race and goals, $F(2,94) < 1$, *ns*. These null findings are important for several reasons. First, they show that the race of primes did not impair people's ability to distinguish between actual guns from actual tools. Thus, the influence of the racial stereotype in this paradigm was *not* driven by reductions in people's ability or motivation to respond in a controlled manner. Similarly, participants' assigned goals did not affect their ability to discriminate between guns and tools. Thus, no support was found for the possibility that people would implement their higher-order goals by changing the extent to which they carefully discriminated between objective target items. Finally, to test whether calling attention to the race of primes influenced the ability to distinguish guns from tools, a focused contrast was performed comparing the discriminability estimate in the control condition to the other two goal groups. However, there was no difference between these groups, and no interaction with Prime race, $F(1,95) < 1$, *ns.* (see Fig. 2).

Accessibility bias estimates

Accessibility bias estimates were constructed so that higher values reflect a greater bias toward responding "gun." Therefore, a stereotypical race bias is shown if bias estimates are higher for the Black prime condition than for the White primes. Fig. 3 displays the bias estimates for each prime, deadline, and goal condition. As Fig. 3 shows, bias estimates were consistently higher after a Black prime than after a White prime, $F(1,94) = 27.66$, $p < .001$. This critical result shows that the effect of the racial cues on misperceptions of the target items was mediated solely through the process of accessibility bias. This effect was not qualified by deadline, Deadline × Prime interaction $F(2,188) < 1$. The fact that processing time did not systematically affect this difference between bias estimates suggests, replicating earlier results, (Payne, 2001) that this bias process

[2] Because both male and female faces were used as primes, it was important to check whether the sex of the face influenced the results described. A Prime sex × Prime race × Target × Deadline × Goal ANOVA was performed to test the effects of prime sex. An unexpected four-way Prime race × Target × Deadline × Prime sex interaction emerged, $F(2,186) = 3.69$, $p < .03$. When this interaction was decomposed, the predicted Prime race × Target interaction was in place for both male primes, $F(1,93) = 31.50$, $p < .001$, and female primes, $F(1,94) = 22.18$, $p < .001$. The higher-order interaction appeared to be driven by the fact that this two-way interaction became steadily stronger as deadline time decreased for male primes, Prime race × Target × Deadline interaction $F(2,186) = 5.25$, $p < .006$. The increase with faster deadlines was in the same direction but less pronounced for female primes, $F(2,188) = 1.85$, $p < .16$. Gender of the prime did not have any other effects, all *p*s > .20. It is not clear why the sole effect of prime gender would have emerged in this higher-order interaction and, pending further replication of it, a theoretical interpretation of it would appear premature. At any rate, this finding does not at all qualify the conclusions to be drawn from our data.

[3] At the longest deadline, a few participants showed perfect discriminability, indicating no errors in distinguishing between guns and tools. As the equations described earlier show, this creates the problem of dividing by zero when the bias estimate is computed. To solve this problem, before calculating estimates, an adjustment was applied to all data as described by Snodgrass and Corwin (1988). See Snodgrass and Corwin (1988) for a description of this and other related methods.

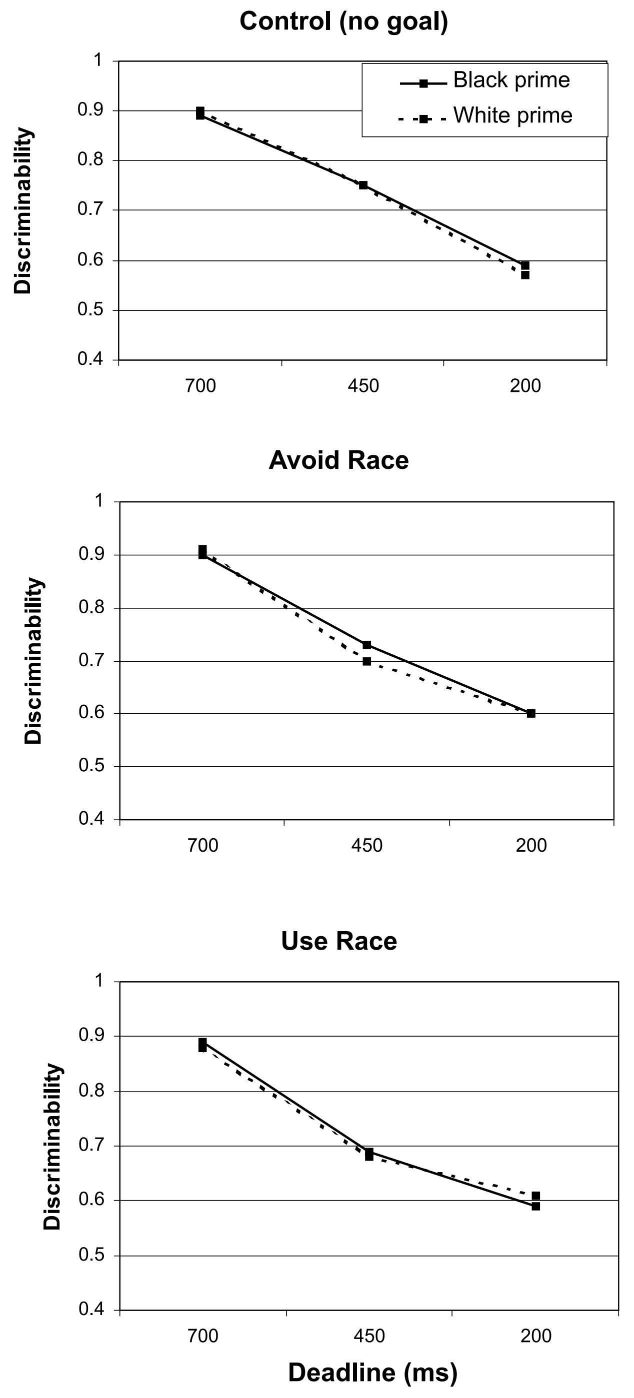

Fig. 2. Discriminability estimates by deadline, prime race, and goal conditions.

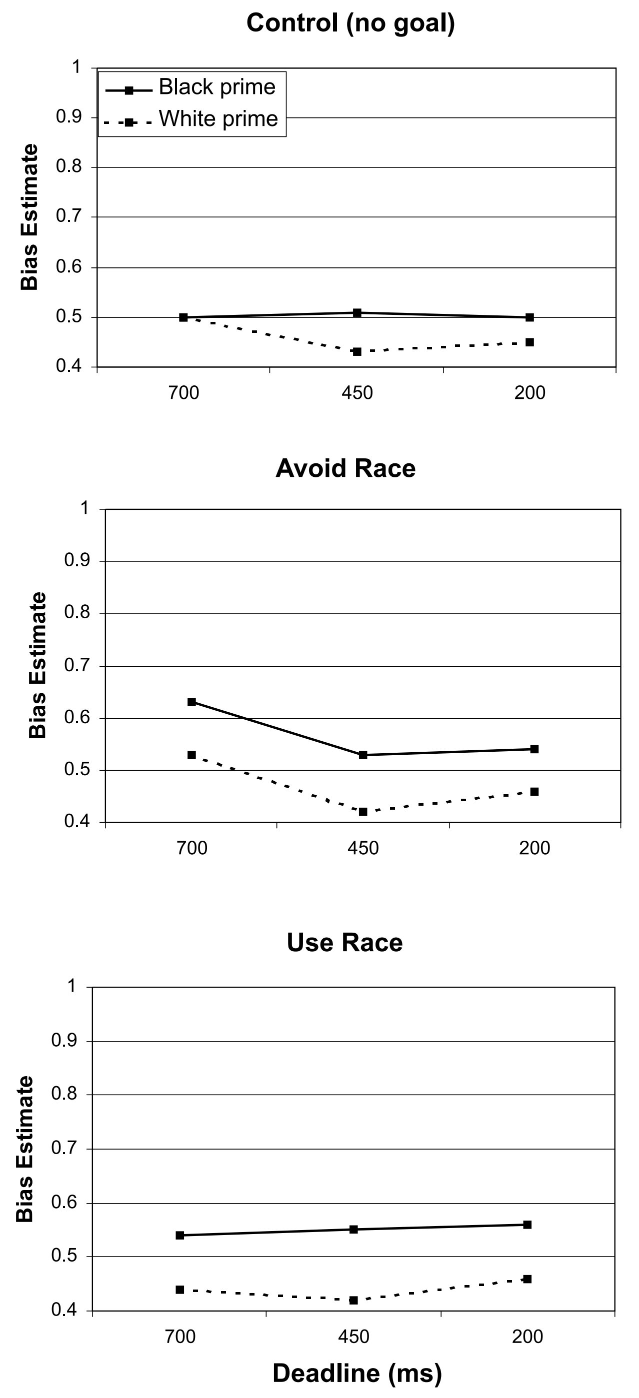

Fig. 3. Accessibility bias estimates by deadline, prime race, and goal conditions.

operates automatically insofar as it takes place to the same extent *regardless* of whether processing time is plentiful or scarce.

The fact that the race-based difference in accessibility bias was not affected by response deadline allows a conclusion from our data that might not have been apparent otherwise. If we had not used the process dissociation procedure, the fact that the overall rate of errors strongly increased as response deadline decreased (see Fig. 1) might be interpreted to suggest that race information was being processed to a greater extent when time was limited, possibly as an effort-saving strategy (e.g., Bodenhausen & Lichtenstein, 1987). However, the constant race-difference in accessibility bias suggests that race information was processed to the *same extent* regardless of processing time. Thus, the observed increase in stereotype-congruent errors under reduced response deadline was wholly due to the fact that discriminability declined and, when this happened, participants' responses were determined by accessibility bias. Our model claims that accessibility bias drives responses when control (i.e., the ability to coordinate behavior to respond based on the objective features of the target) fails. As processing opportunity decreased, control failed more often. Consequently, an accessibility bias of the same magnitude caused more stereotypical behaviors as control declined.

Effects of goals

An important issue in this research is whether accessibility bias was amenable to intentional control. Inspection of Fig. 3 shows that the difference in bias between Black and White primes was similar across all the three goal conditions, a fact that was reflected statistically by the absence of a reliable Goal × Prime race interaction, $F(2, 94) = 1.75$, $p < .18$. Thus, the bias exhibited by participants appears to be unintentional not only in the sense that it operated without an explicit intention to use race, but also in a stronger sense. Namely, the bias operated even when participants tried strategically to avoid the influence of race. The effect of racial cues on the misidentification of objects was thus mediated by a process that is "automatic" with respect to at least three properties. First, the bias estimate is defined by responses that occur regardless of whether they facilitate or interfere with task performance (see Jacoby, Kelley, & McElree, 1999; Payne, 2001). Second, the bias was invariant with respect to processing time. Third, this experiment showed that the bias was invariant with respect to individuals' conscious intentions. However, the bias estimate is not simply an insensitive index that remains constant regardless of manipulations. It was affected by racial primes, and as discussed next, by the salience of racial information.

A focused contrast was performed to compare the race difference in bias levels for the control condition to the difference in the other two goal groups (in which race was made salient). The contrast revealed that the bias difference between Black and White prime conditions was significantly higher in the race-salient goal conditions (the bottom two panels of Fig. 3) than in the control condition, $F(1, 95) = 3.81$, $p < .05$. It appears that when the race of primes was made salient, regardless of the intent, the effect was a strengthening of racial accessibility bias.

Supplemental analyses

In isolation, the similarity of the findings to emerge from the "avoid race" and "use race" conditions leaves open a trivializing alternative interpretation of our data. Specifically, one could argue that, despite the very different instructions given to participants, participants either ignored these instructions or were unmotivated to follow them. If so, then it could be that our results reported thus far are only due to the fact that race was made more salient in the avoid and use race conditions compared to the control condition. We were able to rule out this alternate explanation through administering a set of supplemental questions completed by a subset of participants ($n = 55$) regarding their subjective reports of their goals and their perceived performance. This subset was roughly the second half of the participants who participated; questions were completed after the main dependent variables were collected, and were administered evenly across experimental conditions. There were no significant differences in task performance between this subset and participants who did not complete the supplemental measures.

To verify that participants in the different goal conditions did actually adopt different conscious intentions, two questions tapped the extent to which participants actively tried to use or avoid race in completing the identification task. The first question asked "How much did you *use* the faces in helping you make your decisions?" The second question asked "How much did you *avoid* using the faces in making your decisions?" The scales for these questions were anchored by 0 (*did not use at all*) to 8 (*used strongly*) and 0 (*did not avoid at all*) to 8 (*avoided strongly*), respectively. These two items were averaged together (with the "avoided using race" item reverse coded) to form an index of the extent to which participants subjectively tried to use versus avoid the race of the faces in making their perceptual judgments; higher values indicate a goal to use race.

Consistent with the interpretation offered thus far, the experimental condition to which participants were assigned significantly affected their self-reported goals, $F(2, 54) = 9.45$, $p < .001$. Post-hoc Tukey tests indicated that participants in the "use race" group reported greater intentions to use race ($M = 3.26$) compared to either the control group ($M = 1.23$) or the "avoid race" group ($M = 1.94$), $p < .03$. However, the latter two groups did not differ from each other ($p > .25$). This suggests that, even in the absence of explicit instructions, control participants had a goal not to use the race of the faces as a basis for their judgments. Importantly, the goal manipulation successfully influenced participants' conscious intentions as measured by self-report, but not their performance.

Because the participants' subjective reports were sensitive to experimental manipulations of goals, but their performance was not, it is of interest to know how well participants' subjective experience was calibrated to their actual performance. A third question was thus posed to measure participants' subjective judgments of their own accuracy in identifying tools vs. guns. This question asked "How accurate do you believe you were in identifying the pictures?" and was anchored by 0 (*not* *at all accurate*) and 8 (*very accurate*). To examine the relationship between subjective experience and objective performance, we tested the correlation between this item and actual accuracy scores. This correlation was significant, $r = .41$, $p < .002$, indicating that participants who were objectively more accurate tended to know that they were relatively accurate.

Of greater theoretical interest, we can ask which component(s) of their performance participants were attuned to. Would subjective accuracy be related to discriminability or bias? According to our framework, discrimination is a consciously controlled process, but bias is not. If so, then participants' subjective experience would be expected to track the former to a greater extent than the latter. This is, in fact, what the data show. Averaging across prime race and deadline conditions, the correlation between participants' subjective sense of accuracy and their discriminability estimates fully accounted for the correlation between confidence and accuracy, $r = .41$, $p < .002$. Because discriminability changed across deadline conditions, it is important to test the correlation at each level of processing time.

Table 2 shows the correlation between subjective accuracy ratings and discrimination and bias estimates at each level of deadline and prime race. As expected, subjective accuracy was significantly correlated with target discrimination in each of the six conditions (r's ranged from .27 to .34, all p's $< .05$). In contrast, subjective accuracy was not correlated with bias scores in any of the conditions (r's ranged from $-.14$ to .09, all p's $> .50$). Because accessibility bias scores differed across prime race conditions but not deadline conditions, a similar analysis was performed testing the correlation between subjective accuracy and bias in each prime condition. The relationship was not significant for either condition ($r = .03$ for White primes, $r = -.13$ for Black primes, both p's $> .35$).

It is worth noting that collapsing across deadline intervals caused the overall confidence-discriminability correlation to be inflated because of between-condition variance. Thus, the smaller but reliable within-cell correlations are probably the best index of the relationship. These analyses suggest that participants' subjective experience was reliably attuned to their ability to execute controlled discriminations, but not at all attuned to the biases that drove their responses when control failed. This pattern is important because it suggests one possible reason that participants' biases were not amenable to moderation by conscious goals: when responses were driven by accessibility bias, participants may not have felt they were being particularly inaccurate. Because people often use subjective experience as a guide in controlling behavior, this dissociation of the processes that people are able to successfully monitor has important implications for the control of automatic influences.

Table 2
Correlations between subjective accuracy and discriminability and bias estimates

Deadline (ms)	Prime race	$r_{\text{subjective,discrim.}}$	$r_{\text{subjective,bias}}$
700	Black	.33[a]	−.06
	White	.28[a]	−.05
450	Black	.34[a]	−.14
	White	.27[a]	.09
200	Black	.33[a]	−.04
	White	.30[a]	.05

[a] *Note.* $p < .05$.

General discussion

The present research asked two main questions. First, can people control at will their tendency to stereotypically misidentify objects as guns when paired with Black racial cues, and to misidentify objects as "not guns" when paired with White racial cues? Second, in conditions where this stereotypical tendency is reduced or exacerbated, how do the processes of stimulus discrimination and accessibility bias mediate those changes?

Regarding the first question, participants' explicit goals did not effectively regulate their stereotypical pattern of misidentifications. This bias appears to be very difficult to control. It is not only "unintentional" in the sense that it occurs in the absence of a specific intention; it appears to be "unintentional" in the sense that the bias operates even against the express will of participants. The failure of conscious goals to regulate participants' stereotypical misidentifications of objects as weapons is unlikely to be the result of insensitive measures. Both accessibility bias and discriminability estimates were demonstrated to be sensitive to manipulations expected to affect them. Discriminability was sensitive to limits on processing time, and accessibility was sensitive to racial cues.

Regardless of the participants' goals, they made few stereotypical misidentifications when ample processing time was available. Also, regardless of goals, participants made more stereotypical (vs. counter-stereotypical) errors when race was made salient than when it was not. Together, these latter two findings imply an alarming puzzle for the practical considerations of reducing harmful effects of prejudice. The two recommendations this research suggests for reducing prejudicial false identifications are: (1) "slow down" and (2) "don't notice race." However, these are precisely the sorts of suggestions that may be impossible to implement in natural settings. Consider urging a police officer to react slowly during a confrontation with a potentially armed suspect! Speed is obviously important in this situation, and the time pressure immense.

The second suggestion, to ignore race, may be equally impractical. A growing body of research shows that suppressing particular thoughts can be quite difficult, and even result in "rebound effects" in which the unwanted thought becomes more influential than it would have been without suppression attempts (e.g., Macrae, Bodenhausen, Milne, & Jetten, 1994; Wegner, 1994). If so, suggesting either to experimental participants or to police officers that they should *not pay attention* to race may be as ineffective as asking readers not to think of a "white bear," (Wegner & Erber, 1992). Although our results yielded no direct evidence of a rebound effect, the fact that "avoid bias" instructions make salient the very information that one is imploring participants to ignore similarly highlights the difficulties inherent in successful inhibition of stereotypical processing.

When estimates of discriminability and accessibility bias are examined, the reasons for these results become clear. The salience of race heightened the accessibility of racial expectations. Unfortunately, it did not matter whether participants were focusing on race with the intention to avoid its influence, or with the intention to employ their racial prejudices. Participants' conscious goals did not reliably affect either discriminability or accessibility bias, except through the (partially ironic) effects of race-salience.

A limitation of the present research that deserves note is the type of goals that were induced. The goals investigated in this experiment were temporary, situationally induced goals. Some researchers have recently suggested that for a goal to regulate highly over-learned, automatic processes, the goal itself must become "automatized" through repeated use (Bargh, 1999; Moskowitz et al., 1999). It is possible that extensive practice in implementing egalitarian goals could control the operation of stereotypes by preventing the accessibility bias observed in this study (see Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000).

We have conceptualized cognitive control here as the ability to flexibly regulate responses based on conscious goals. Consistent with this rationale, goals were implemented as conscious, temporary attempts. An exciting direction for continuing research concerns whether well-practiced goals can regulate the impact of accessibility biases. If so, it is worth asking whether that regulation retains the *flexibility* that typically characterizes conscious control over actions, so that participants can choose to use or avoid stereotypes. An intriguing alternative possibility is that the implementation of such proceduralized goals may simply become a different source of accessibility bias, which influences participants' behavior when control fails, whether they intend for it to or not.

The use of process dissociation analysis provided insights into the operation of stereotype processes that would not have been evident using other methods such as accuracy alone, or response time. This study demonstrated important boundary conditions on the ability to control automatic influences in a task with potentially weighty implications, and specified the processes underpinning those boundaries. These processes were shown to operate independently in multiple ways. Not only were they affected differently by manipulations intended to influence accessibility and control selectively, but they also showed distinct relationships to participants' subjective experience. It is the convergence of evidence from these multiple sources that not only validates the assumptions of the present approach, but also makes more complete the emerging picture of how goals, subjective experience, and cognitive control interact to regulate, and sometimes exacerbate, psychological biases.

References

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68*, 181–198.

Bargh, J. A. (1999). The cognitive monster: the case against the controllability of automatic stereotype effects. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford.

Bargh, J. A., & Gollwitzer, P. M. (1994). Environmental control of goal-directed action: automatic and strategic contingencies between situations and behavior. In W. D. Spaulding (Ed.), *Integrative views of motivation cognition and emotion* (pp. 71–124). Lincoln, NE: University of Nebraska Press.

Blair, I., & Banaji, M. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142–1163.

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information-processing strategies: The importance of task complexity. *Journal of Personality and Social Psychology, 52*, 871–880.

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

Devine, P. G., & Monteith, M. J. (1999). Automaticity and control in stereotyping. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology*. New York: Guilford.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: the MODE model as an integrative framework. *Advances in Experimental Social Psychology, 23*, 75–109.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (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.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. Dovidio, & S. Gaertner (Eds.), *Prejudice discrimination and racism*. Orlando, FL: Academic Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review, 192*, 4–27.

Hense, R. L., Penner, L. A., & Nelson, D. L. (1995). Implicit memory for age stereotypes. *Social Cognition, 13*, 399–415.

Hintzman, D. L., & Curran, T. (1997). More than one way to violate independence: reply to Jacoby and Shrout (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 511–513.

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jacoby, L. L., Begg, I. M., & Toth, J. P. (1997). In defense of functional independence: violations of assumptions underlying the process-dissociation procedure? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 484–495.

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: early selection vs late correction. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 383–400). New York: Guilford.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*, 871–888.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: is prejudice inevitable? *Journal of Personality and Social Psychology, 72*, 275–287.

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

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: a peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*, 37–47.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Shaal, B. (1999). Preconscious control over stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167–184.

Payne, B. K. (2001). Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181–192.

Payne, B. K., Lambert, A. J., & Jacoby, L. L. (2002) Attitudes as accessibility bias: dissociating automatic and controlled processes. In: J. S. Uleman, J. A. Bargh, & R. Hassin, (Eds.), The unconscious mind (in press).

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solo (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127–190.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.

Uleman, J. S. (1999). Spontaneous and intentional inferences in impression formation. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology*. New York: Guilford.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*, 34–52.

Wegner, D. M., & Erber, R. (1992). The hyperaccessibility of suppressed thoughts. *Journal of Personality and Social Psychology, 63*, 903–912.