How Dissociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach

Nier, Jason A.

How Dissociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach

Jason A. Nier
Connecticut College

The current study examined the implicit and explicit attitudes of White Americans toward African-Americans. A variation of the Bogus Pipeline procedure was employed to determine if the apparent dissociation between implicit and explicit measures of racial attitudes that is reported in previous research might be exaggerated. The results indicated that the relationship between implicit and explicit attitudes was only significant under Bogus Pipeline conditions, while implicit and explicit attitudes were largely dissociated when they were measured under normal circumstances. Thus, it appeared that as the motivation to accurately report explicit attitudes increased, the implicit–explicit relationship strengthened and the dissociation between implicit and explicit racial attitudes was substantially reduced. The results indicate that Whites' implicit and explicit attitudes toward African-Americans may not be as greatly dissociated as some theories of racial attitudes have presumed.

KEYWORDS   implicit attitudes, racial attitudes

FOR DECADES social psychologists have observed that White attitudes toward African-Americans have become increasingly positive (Jones, 1997). Despite the increased positivity of self-reported attitudes, subtle measures of prejudice suggest that negative attitudes persist (Dovidio & Gaertner, 1998), calling into question the validity of the increasingly positive self-reports. Recently, however, the introduction of implicit measures of racial attitudes, such as the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998), have promised to enhance the accuracy of racial attitude assessments. These measures are believed to be less reactive than explicit measures since they do not rely upon participants' self-reports (Greenwald et al., 1998). Furthermore, implicit measures of White attitudes toward African-Americans generally yield attitude assessments that are more negative than explicit measures, suggesting that they may relate to more subtle forms of negative inter-racial behavior (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997).

Author's note
Address correspondence to Jason A. Nier, Department of Psychology, Box 5305, Connecticut College, 270 Mohegan Avenue, New London, CT 06320, USA
[email: janie@conncoll.edu]

The relationship between implicit and explicit measures of racial attitudes

While implicit measures seem to yield attitude assessments that are more negative than self-reported attitudes, interpreting the precise meaning of these measures has proven problematic. Perhaps the most distinctive aspect of implicit measures is their apparently inconsistent relationship to traditional, explicit measures of racial attitudes (for reviews see Blair, 2001; Dovidio, Kawakami, & Beach, 2001; Fazio & Olson, 2003). Although there have been numerous exceptions (e.g. Cunningham, Preacher, & Banaji, 2001; Wittenbrink, Judd, & Park, 1997), many studies examining the relationship between the two measures have found that implicit attitudes are often *not* correlated with explicit attitudes (e.g. Boniecki & Jacks, 2002, *r* = .03; Dasgupta, McGhee, Greenwald, & Banaji, 2000, *r* = .02; Greenwald et al., 1998, *r* = .07). This has led researchers to disagree about what implicit measures are actually assessing and has even led some to question the utility of implicit measures (Cameron, Alvarez, & Bargh, 2000).

While it remains clear that there is a considerable degree of inconsistency between implicit and explicit measures of White attitudes toward African-Americans, there have been several different explanations offered to explain this poor correspondence. One explanation focuses on methodological concerns, particularly the psychometric properties of implicit measures. The reliability of implicit measures is generally lower than the reliability of explicit measures (Kawakami & Dovidio, 2001), and as a result the error variance that is present in the measurement of implicit attitudes may suppress the implicit–explicit relationship. Consistent with this interpretation, procedures that have been introduced to improve the reliability of implicit measures strengthen the magnitude of the implicit–explicit relationship (Cunningham et al., 2001; Greenwald, Nosek, & Banaji, 2003).

Another explanation for the weak implicit–explicit link, which Blair (2001) called the *representational* perspective, holds that implicit and explicit attitudes are independent cognitive representations that are the result of different psychological processes. While there are differing theoretical variations on this general theme (Devine, 1989; Wilson, Lindsey, & Schooler, 2000), the basic idea is that the poor implicit–explicit relationship is due primarily to the fact that implicit and explicit measures assess different evaluations of the same attitude object. In those circumstances where implicit and explicit evaluations differ, as is often presumed to be the case for White attitudes toward African-Americans, attitudes are said to be 'dissociated' (Greenwald et al., 1998) or 'independent' (Wilson et al., 2000).

Furthermore, some theories of racial attitudes are based, in part, on the notion of 'dissociation' that is associated with the representational perspective. For example, Devine's (1989) influential model of prejudice holds that all Whites hold automatically activated negative beliefs, yet individuals vary in the negativity of explicit beliefs toward African-Americans. As evidence for this model, Devine pointed out that responses on the Modern Racism Scale, an explicit measure of racial attitudes, were unrelated to the degree to which racial stereotypes were automatically activated using a subliminal priming procedure.

Further evidence for the dissociation between implicit and explicit attitudes comes from the differential predictive validity of these measures. Implicit measures tend to predict spontaneous behaviors that are difficult for individuals to consciously monitor and control, while explicit measures tend to predict deliberative responses that are easily monitored and controlled (Dovidio et al., 1997; Dovidio, Kawakami, & Gaertner, 2002; Wilson et al., 2000). Thus the notion that implicit and explicit attitudes are largely dissociated seems to be consistent with prevailing theories of attitudes and the available empirical evidence.

However, the empirical evidence upon which the representational perspective is based does not account for the impact of social desirability concerns, a motivational variable that can influence the expression of explicit racial attitudes. Specifically, since most individuals are

motivated to present a positive self-image (Schlenker & Weingold, 1992), they may manipulate their explicit responses so as to appear unprejudiced. Supporting this view, Sigall and Page's (1971) classic 'bogus pipeline' study indicated that Whites expressed more negative attitudes toward African-Americans when they believed that a device was able to accurately assess their true attitudes. Thus when participants were led to believe that a machine was a 'pipeline' to their true inner states and attitudes, social desirability forces presumably exerted less influence on the self-reported racial attitudes and as a result they reported attitudes that more closely reflected their true beliefs. More recent research indicates that social desirability forces continue to influence Whites' explicit attitude toward African-Americans (e.g. Fazio, Jackson, Dunton, & Williams, 1995).

Whereas the social desirability perspective has focused exclusively on the malleability of explicit measures under different conditions (Roese & Jamieson, 1993) and there seems to be little doubt that social desirability forces influence the expression of explicit attitudes, less attention has been directed toward the influence of social desirability on implicit measures. This is not entirely surprising, since implicit measures are perhaps less likely to be influenced by social desirability forces than explicit measures. Nevertheless, the strength of the *relationship between* implicit and explicit measures may be moderated by the influence of social desirability on explicit measures (Dunton & Fazio, 1997), since explicit attitudes may shift under conditions that minimize the impact of social desirability motivations. Specifically, if Whites report more negative explicit attitudes toward African-Americans due to a bogus pipeline procedure and these explicit responses are less contaminated by social desirability concerns, it is possible these attitudes may be more highly correlated with implicit measures, since implicit measures also tend to yield fairly negative attitudes assessments and are also less likely to be influenced by social desirability forces. To date, there is no published research that has systematically examined the degree to which the implicit–explicit relationship may

strengthen under conditions designed to minimize socially desirable responding. This question has important theoretical implications since the representational explanation holds that the dissociation between implicit and explicit attitudes is due primarily to the different psychological processes that produce these two types of attitudes, rather than social desirability forces.

Current study

In the current study, the implicit and explicit attitudes of Whites toward African-Americans were assessed under different circumstances in order determine the degree to which implicit and explicit attitudes are dissociated when social desirability forces are minimized. A variation of the bogus pipeline procedure was employed to examine whether the implicit–explicit relationship would be stronger under bogus pipeline conditions, relative to the circumstances under which implicit and explicit measures have traditionally been administered. Participants in all conditions were asked to complete the Implicit Association Test (IAT), an implicit measure (Greenwald et at., 1998), and the Modern Racism Scale (MRS), an explicit measure of racial attitudes (McConahay, 1986). The order in which participants completed the IAT and MRS was counterbalanced so that the presence of order effects could be determined.

In order to manipulate participants' motivation to provide accurate self-reports, the information that participants received about the nature of the IAT was systematically varied. In one condition (the *Accurate* condition) participants were told that the IAT was an accurate measure of racial attitudes, not unlike a lie detector. Thus the motivation to accurately report one's explicit attitude was strong since participants believed that the ostensibly accurate implicit measure would be used to corroborate their self-reported attitude. In a second condition, participants received information that stressed that the implicit measure was *not* an accurate measure of their racial attitudes. Thus in this condition (the *Inaccurate*

condition), participants were aware that their responses on the explicit measure would be compared to their performance on an implicit measure, but they did not have to be concerned with the ability of the implicit measure to accurately assess their true attitude. In the final condition (the *No Information* condition) participants did not receive any information about the nature of the IAT. Thus this condition reflects the context in which the implicit–explicit relationship is generally assessed; the motivation to accurately report an explicit attitude was weak and social desirability motivations were presumably high.

It was predicted that participants in the Accurate condition would express more negative explicit racial attitudes compared to participants in the No Information and Inaccurate conditions. Consequently, the relationship between implicit and explicit racial attitudes was predicted to be stronger in the Accurate condition, relative to the Inaccurate condition and the No Information condition. Such a pattern of results would indicate that the increased motivation to accurately report explicit racial attitudes, which is strong in the Accurate condition and weaker in the Inaccurate and No Information conditions, reduces the dissociation between implicit and explicit attitudes, indicating that the true degree of dissociation between implicit and explicit racial attitude may be smaller than some theories of racial attitudes have presumed.

Additionally, the order in which the MRS and the IAT were administered was counterbalanced in order to examine the plausibility of a potential alternative interpretation of the results. As Whites complete the IAT they often realize that the results of the test will indicate that they are biased against African-Americans (Monteith, Voils, & Ashburn-Nardo, 2001). Consequently, when participants then complete an explicit measure they may adjust their explicit responses to 'fit' with their IAT performance. Thus a participant's perception of their performance on the IAT, rather than changes in social desirability forces, might cause a shift in explicit attitudes and a strengthening of the implicit–explicit relationship. The

counterbalancing of the order in which the MRS and IAT were administered allowed this alternative explanation to be empirically tested. Specifically, if a stronger implicit–explicit relationship is observed only when the IAT is completed first and the MRS second, thereby allowing participants to adjust their responses on the MRS to fit with their perceived performance on the IAT, then a strengthened implicit–explicit relationship could indeed be due to the fact that participants adjusted their MRS responses to fit with their IAT performance. However, if the magnitude of the implicit–explicit correlation is similar regardless of the order in which the measures are completed, then the results cannot be due to the adjustment of MRS responses to fit with perceived IAT performance, since participants who completed the MRS before the IAT could not be aware of their IAT performance at the time they completed the MRS. Such a pattern of results would strongly suggest that the minimization of social desirability forces due to the bogus pipeline manipulation would be the more likely explanation for a strong implicit–explicit relationship.

Method

Participants
Participants were 112 White undergraduates enrolled at a small college in New England. Participants received course credit for their participation.

Design
A 3 (Information about IAT: Accurate, Inaccurate, No Information) × 2 (Order of Measures: IAT First, IAT Last) between-subjects factorial design.

Measures
The IAT was used to assess implicit racial attitudes toward Blacks and Whites (Greenwald et al., 1998). The IAT requires individuals to categorize target concepts (in this case, Black and White faces) and attributes (good and bad words). In the version of the IAT used in the current study, 10 faces (5 male and 5 female; 5

Black and 5 White) were employed as stimuli to represent Black and White social groups; these faces were taken from the group of 12 faces used by Cunningham et al. (2001). The following evaluative words (drawn primarily from Greenwald et al., 1998) were utilized as 'good' words: lucky, rainbow, love, peace, heaven, pleasure, cheer, and happy. The following words were used as 'bad' words: evil, death, pain, disaster, ugly, vomit, stink, and rotten.

These stimuli were employed in a procedure very similar to previous IAT research where the IAT task was composed of five blocks; blocks 1, 2, and 4 were practice blocks and the data were collected during blocks 3 and 5, with data being collected on 40 trials from each block. For one of these blocks, participants classified White faces and good words on one key and Black faces and bad words on another key (*white+good/black+bad* block). For the other block, participants classified White faces and bad words on one key and Black faces and good words on another key (*white+bad/black+good* block). The order of these two data collection blocks was determined randomly for each participant. The accuracy and latency of participants' responses were recorded on each trial. As suggested by Greenwald et al. (2003) if a participant gave an incorrect response on a trial, the trial continued until the correct response was given and these latencies were included in subsequent IAT analyses.

A six-item version of the MRS was administered to assess explicit attitudes toward African-Americans (McConahay, 1986). This version of the MRS contained the following items: 'Blacks are getting too demanding in their push for equal rights', 'Discrimination against Blacks is no longer a significant problem in the United States', 'It is easy to understand the anger of Black people in America' (reverse scored), 'Over the past few years the government and news media have shown more respect for Blacks than they deserve', 'Blacks should not push themselves where they are not wanted', 'Over the past few years Blacks have gotten more economically than they deserve'. Participants expressed their degree of agreement with each statement on a

1 to 7 scale, where 1 indicated strong disagreement and 7 indicated strong agreement. The reliability of the scale was acceptable at pre-test (alpha = .74) and post-test (alpha = .86). The average of the six MRS items constituted the explicit measure of participants' racial attitudes at each time period.

Procedure

The study consisted of two separate sessions; one session was a pre-test session in which the MRS was administered and the second session was the laboratory portion of the study.[1] Upon arrival at the laboratory, participants were informed that they would be participating in a study of word associations and attitudes. Participants were then randomly assigned to one of three conditions.

In the *Accurate* condition, participants completed the IAT and the MRS. Also, the experimenter gave the participants information that ostensibly established the accuracy of the IAT as a measure of racial attitudes. Specifically, immediately prior to completing the MRS participants were told that the IAT was designed to assess White attitudes toward Black Americans. Participants then watched a five-minute videotape of a segment from Dateline, a television news program, which attested to the accuracy of the IAT (Banaji & Greenwald, 2000). After the clip was over, the experimenter read aloud the following information about the accuracy of the IAT:

> Previous research by social psychologists has demonstrated that these computer tasks are very accurate measures of people's true racial attitudes. For example, social psychologists at Yale University, University of Washington, University of Colorado, University of Delaware, and Colgate University have all successfully used these computer measures of racial attitudes. Although these computer tasks may not be a perfect measure of your racial attitudes, *it's the closest thing to a lie detector that social psychologists can use to determine your true beliefs about race.*

Similar to most research employing the Bogus Pipeline procedure (Roese & Jamieson, 1993), where participants are asked to predict the output of the pipeline device, when completing

the MRS participants were instructed that their response should reflect 'what the computer would predict' their response to be for each of the six statements, rather than indicating their personal attitude toward African-Americans.[2] In order to test for the presence of order effects, the order of completion of the IAT and the MRS was counterbalanced so that half the participants first completed the IAT, were then given the information about the IAT, and then completed the MRS. The remaining participants first were given information about the IAT, completed the MRS, and then completed the IAT.

The *Inaccurate* condition was identical to the Accurate condition, except that the experimenter stated that the IAT was a poor measure of racial attitudes. Specifically, the experimenter read aloud the following information just before participants completed the MRS.

> [The IAT was] originally proposed to assess individual's attitudes toward Black Americans. Psychologists hoped that an individual's performance on these computer tasks would predict whether people hold negative attitudes about members of different racial groups. Unfortunately, it was discovered that people's performance on these computer tasks is virtually unrelated to their racial attitudes. For example, Dr. Russell Fazio, a social psychologist at Indiana University, has found that people's performance on these computer tasks is unrelated to their actual racial attitudes and other beliefs about race. In other words, he found that *the computer tasks don't tell us anything about your racial attitude.*

Participants then completed the MRS immediately afterward. Similar to the Accurate condition, as participants completed the MRS they were instructed to indicate 'what the computer would predict' their response to be for each of the six statements and the order of the completion of the MRS and IAT was counterbalanced.

In the *No Information* condition participants completed the MRS and IAT without receiving any information about the accuracy of the IAT as a measure of racial attitudes, and the order of the completion of the IAT and MRS was again counterbalanced. Also, participants were asked to indicate their true beliefs about race (rather than predicting the results of the computer's analysis) when completing the MRS. Thus the No Information condition reflected the procedure through which implicit and explicit measures of racial attitudes are often administered (e.g. Dasgupta et al., 2000; Kawakami & Dovidio, 2001).

Results

Implicit attitudes

Extreme IAT reaction time outliers were recoded, such that responses quicker than 300 ms were recoded as 300 ms and responses greater than 1200 ms were recoded as 1200 ms (similar to Dasgupta et al., 2000). Mean response latencies and error rates were calculated for each data collection block of trials (the *white+good/black+bad* block and the *white+bad/black+good* block).[3] Replicating previous IAT research, response latencies were significantly longer for the *white+bad/black+good* trials (M = 708.0 ms), relative to the white+good/black+bad trials (M = 586.1 ms), $t(111) = 11.3$, $p < .001$, indicating that participants displayed an implicit evaluative preference for White over Black.

A similar analysis was also conducted on error rates. The proportion of trials in which a participant gave an incorrect response was calculated for both of the data collection blocks. A *t* test revealed that the proportion of incorrect trials was significantly higher for *white+bad/black+good* trials ($M = 6.88\%$), relative to *white+good/black+bad* trials ($M = 5.20\%$), $t(111) = 3.35$, $p < .001$, which again indicated that participants displayed an implicit preference for White over Black.

Next, a 3 (Information: Accurate, Inaccurate, No Information) × 2 (Order: IAT First, IAT Last) × 2 (Block: white+good/black+bad, white+bad/black+good) mixed multivariate analysis of variance (MANOVA) was conducted to examine whether implicit attitudes varied across experimental condition, where the IAT effects for reaction time and error rates were treated as the dependent variables, Block was treated as the within-subjects factor, while Information and Order were treated as

Table 1. Changes in Modern Racism Scores as a function of Information condition

Experimental condition	Pre-test		Post-test	
	M	SD	M	SD
Accurate	2.30[a]	0.92	3.55[b]	1.17
Inaccurate	2.13[a]	0.77	2.96[c]	1.08
No Information	2.02[a]	0.63	1.82[a]	0.62

Note: The means that do not share the same superscripts are significantly different, Tukey, $p < .05$.

between-subjects factors. As expected, the interaction between Information condition and Block was not significant, *Wilks' Lambda* = .99, $p = .66$, indicating that the implicit attitudes toward African-Americans were similar across experimental condition.[4] In sum, the IAT results were consistent with previous research. The IAT data indicated that participants, on average, held an implicit preference for White over Black. Furthermore these implicit attitudes did not vary as a function of experimental condition.

Explicit attitudes

A 3 (Information: Accurate, Inaccurate, No Information) $\times$ 2 (Order: IAT First, IAT Last) $\times$ 2 (Time: Pre-test, Post-test) mixed ANOVA was conducted to examine whether participants in the Accurate condition would display greater changes in MRS scores, relative to participants in the Inaccurate and No Information conditions. There was a main effect for order ($F(1, 106) = 4.8$, $p < .05$), with participants who completed the IAT before the MRS reporting higher MRS scores ($M = 2.62$) than those who completed the MRS before the IAT ($M = 2.31$). However, there were no other significant effects involving order.

More importantly, the expected Time $\times$ Information interaction was significant ($F(2, 106) = 21.6$, $p < .001$), indicating that the degree of change in MRS scores from pre-test to post-test varied as a function of the Information condition (see Table 1). Planned follow-up tests revealed that the pre-test MRS scores for participants in the No Information condition ($M = 2.02$) were not significantly different than their post-test scores ($M = 1.82$)

($t (33) = 1.5$, $p = ns$). Participants in the Inaccurate condition displayed a significant increase in their MRS scores from pre-test ($M = 2.13$) to post-test ($M = 2.96$) ($t (36) = 5.2$, $p < .05$).[5] Additionally, the MRS scores for participants in the Accurate condition were significantly higher at post-test ($M = 3.55$), relative to their pre-test scores ($M = 2.30$) ($t (40) = 7.3$, $p < .001$). Planned comparisons also revealed that participants in the Accurate condition displayed significantly higher post-test MRS scores, relative to participants in the Inaccurate condition ($t (76) = 2.4$, $p < .05$) and the No Information condition ($t (73) = 8.0$, $p < .001$).[6] Thus the hypothesis, that the degree of change in MRS scores would vary according to experimental condition, was supported; participants in the Accurate condition displayed the greatest increase in their MRS scores, participants in the Inaccurate condition displayed a significant but smaller change in their MRS scores, and participants in the No Information condition did not display any change in their MRS scores. Also, this crucial interaction effect was not qualified by the Order variable.

Relationship between implicit and explicit measures

In order to examine the relationship between implicit and explicit attitudes, the IAT effect was collapsed into a single composite score. To accomplish this goal for response latencies, the mean *white+good/black+bad* latency was subtracted from the mean *white+bad/black+good* latency. Similarly, the mean *white+good/black+bad* error rate was subtracted from the *white+bad/black+good* error rate. Next, the IAT effects for both latency and error rates were standardized

and these standardized scores were then averaged together. This average captured the IAT effect into a single composite score, where higher scores indicated more negative implicit attitudes toward Blacks.[7]

This IAT effect was then used as the dependent variable in a multiple regression analysis that was designed to test the hypothesis that the relationship between implicit and explicit measures would be stronger in the Accurate condition than in the Inaccurate and No Information conditions. To test this hypothesis, two contrast codes representing the Information condition were employed (see Aiken & West, 1991; Cohen & Cohen, 1983). The first contrast code compared the Accurate condition versus the Inaccurate and No Information conditions and the second contrast code compared the Accurate and Inaccurate conditions to the No Information condition. We also included a final contrast coded variable that represented the order in which participants completed the MRS and IAT. The three contrast coded variables, the composite IAT effect, and their corresponding interaction terms were then used to predict post-test MRS scores. It was predicted that the interaction between the IAT effect and the first contrast coded variable (that contrasted the Accurate versus the Inaccurate and No Information conditions) would be a significant predictor of MRS scores. Such a pattern of results would signify that the relationship between the IAT effect and MRS scores was significantly stronger in the Accurate condition than in the Inaccurate and No Information conditions.

Overall the regression model accounted for 49% of the variance in post-test MRS scores (see Table 2). In terms of individual variables, the composite IAT effect (beta = −.21, $p < .01$), the contrast comparing the Accurate condition to the Inaccurate and No Information conditions (beta = −.47, $p < .01$), and the contrast comparing the No Information condition to the

Table 2. Multiple regression analysis of the implicit–explicit relationship

Predictor variable	Beta	t	p value
Main effects			
IAT effect	−.21	2.91	<.01*
Order condition	.13	1.80	.08
Information condition			
Contrast One (Accurate vs. Inaccurate/No Info)	−.47	6.48	<.01*
Contrast Two (No Info vs. Pipeline/Inaccurate)	.37	5.17	<.01*
Two-Way Interactions			
IAT effect × Order condition	−.03	0.36	.72
IAT effect × Information condition			
IAT effect × Contrast One (Accurate vs. Inaccurate/No Info)	−.16	2.21	.03*
IAT effect × Contrast Two (No Info vs. Pipeline/Inaccurate)	.04	0.51	.62
Order condition × Information condition			
Order condition × Contrast One (Accurate vs. Inaccurate/No Info)	−.08	1.04	.30
Order condition × Contrast Two (No Info vs. Pipeline/Inaccurate)	.14	1.91	.06
Three-Way Interaction			
IAT effect × Order condition × Information condition			
IAT effect × Order condition × Contrast One	.05	0.63	.53
IAT effect × Order condition × Contrast Two	−.02	0.23	.82

$R^2 = .49$

Notes: Regression model predicting post-test MRS scores from the composite IAT effect, Order condition and Information condition. Asterisks indicate statistically significant predictors, $p < .05$.
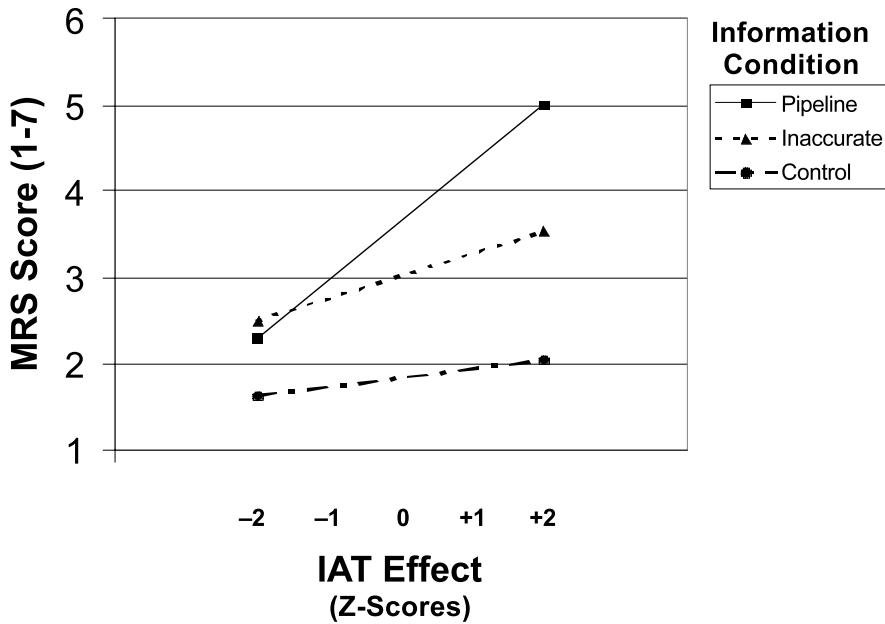
Figure 1. IAT effect × Information condition interaction. Lines represent simple regression lines predicting MRS scores from the composite IAT effect in each condition.

Accurate and Inaccurate conditions (beta = .37, $p < .01$), were all significant predictors of post-test MRS scores. The contrast that corresponded to the order in which participants completed the MRS and IAT was not significant, nor were any of the interactions involving this variable.[8]

Most importantly, the only significant interaction term was the interaction between the IAT effect and the contrast comparing the Accurate condition to the Inaccurate and No Information conditions (beta = −.16, $p < .05$), indicating that the relationship between the IAT effect and MRS scores was moderated by the information that participants received about the IAT. The exact nature of this interaction is depicted in Figure 1, which presents the simple regression lines predicting post-test MRS scores from the composite IAT effect in each condition. As the figure indicates, among participants in the Inaccurate and No Information conditions there is a very weak (and nonsignificant) relationship between MRS

scores and IAT scores. However, for participants in the Accurate condition there is a fairly strong positive relationship between MRS scores and IAT scores. This pattern of results suggests that as the motivation to accurately report one's racial attitudes increases, the relationship between MRS scores and IAT scores strengthens.[9]

A supplementary analysis was also performed in which the implicit–explicit relationship was examined using a correlational approach (see Table 3). Supportive of the regression analysis, the pattern of implicit–explicit correlations in each condition yielded a similar interpretation. The correlation between the composite IAT effect and post-test MRS scores was significant in the Accurate condition ($r = .51$, $p < .001$) and was not significant in either the Inaccurate condition ($r = .18$, $p = ns$) or the No Information condition ($r = .14$, $p = ns$). Similar to the results of the regression analysis, the implicit–explicit correlation in the Accurate condition was significantly stronger than the correlations

Table 3. Strength of implicit–explicit relationship as a function of Information condition

	Correlation between MRS and IAT effect		
Experimental condition	Composite IAT effect	IAT effect for latency	IAT effect for errors
Accurate	.51*	.39*	.46*
Inaccurate	.18	.25	.04
No Information	.14	.12	.10

Note: Asterisks denote statistically significant correlations, $p < .05$.

observed in the Inaccurate ($z = 1.65$, $p < .05$; one-tailed) and No Information ($z = 1.75$, $p < .05$; one-tailed). Thus the pattern of correlations mirrors the relationships observed in the multiple regression analysis.

Were there any other plausible interpretations of the results?

Although the predicted pattern of results was observed, there remained two possible alternative interpretations of the results. One possible alternative interpretation of the results revolves around the incidental feedback that participants receive when completing the IAT. When completing the IAT, participants are often aware that they fare poorly on the test. As a result, one possible interpretation of the strengthened implicit–explicit relationship is that participants adjust their explicit responses to 'fit' with their performance on the IAT. According to this interpretation, the explicit attitudes expressed in the Accurate condition would reflect their perceived performance on the IAT, rather than their actual racial attitudes. If this alternative interpretation was correct, then the crucial IAT effect × Information condition interaction should have been qualified by the three-way interaction effect involving the Order variable. However, the three-way interaction term that tested this effect did not approach significance (see Table 2). Also, the correlations between MRS scores and the IAT effect among participants in the Accurate condition were very similar in magnitude regardless of the order of completion (IAT first $r = .46$, IAT last $r = .53$).

Another alternative interpretation that could potentially explain the increased implicit–explicit relationship is that people simply reported more negative attitudes in the Accurate condition, regardless of their level of implicit prejudice. However, the data are not consistent with this interpretation. If all participants in the Accurate condition simply elevated their MRS scores to a similar degree, the relationship between post-test MRS scores and IAT scores would not have been any stronger in the Accurate condition. An increase in the implicit–explicit relationship requires, by definition, that those who had more negative IAT scores reported negative explicit attitudes to a greater degree than those who had more positive IAT scores. Therefore the stronger implicit–explicit relationship observed in the Accurate condition is not consistent with the notion that everyone simply reported more negative MRS scores in the Accurate condition.

Discussion

The current study was designed to demonstrate that the apparent dissociation between implicit and explicit measures of racial attitudes is exaggerated in most research, due to the conditions under which explicit attitudes are assessed. The results indicated that when participants believed their 'true attitudes' were being accurately assessed, there was a significant relationship between an implicit measure of racial attitudes (the IAT) and an explicit measure of racial attitudes (the MRS). When participants did not believe that their self-reported explicit attitudes could be accurately corroborated with an implicit measure, there was no association

between implicit and explicit attitudes. Thus the results suggest that as the motivation to report explicit attitudes that are consistent with implicit attitudes increases, the implicit–explicit relationship strengthens due to changes in self-reported explicit attitudes (implicit attitudes were not influenced by the information that participants received about the IAT). Furthermore, the results of the current study cannot be explained as an order effect, since the magnitude of implicit–explicit link did not vary as a function of the order in which the IAT and MRS were completed.

These results have implications for the *representational* explanation of the poor correspondence between implicit and explicit racial attitudes. Some theorists have argued that the dissociation between implicit and explicit attitudes is evidence that implicit and explicit attitudes are independent cognitive representations that are the result of different psychological processes (e.g. Devine, 1989; Wilson et al., 2000). However, the results of the current study indicate that this dissociation, evident in so many previous studies, is at least partially due to social desirability concerns. To conclusively demonstrate the true extent to which implicit and explicit racial attitudes are dissociated, social desirability motivations and other factors that are known to underestimate the magnitude of the implicit–explicit link, such as measurement error, must be minimized.

How strong might the implicit–explicit link be under ideal conditions that minimize the impact of factors that cause the implicit–explicit link to be underestimated? While the current study cannot definitively address this question, a supplementary analysis was performed in which the magnitude of the implicit–explicit relationship was corrected for the effect of measurement error, which is sometimes referred to as correction for attenuation (Cohen & Cohen, 1983). This analysis allows one to estimate the magnitude of a correlation between any two variables if the measurement error due to unreliability could be removed. This analysis is performed by dividing the original correlation coefficient by the square root of the product of the reliabilities of each

measure. In the current analysis, the original value of the implicit–explicit correlation in the Accurate condition ($r = .51$) was divided by the square root of the product of the reliabilities of each measure (MRS alpha = .86; IAT alpha = .50). This calculation yielded a strong implicit–explicit correlation ($r = .79$). While the results of this analysis must certainly be interpreted with caution, it nonetheless suggests that employing a measurement procedure that both minimizes the impact of error variance (such as structural equation modeling) and yields more forthright explicit responses (such as the bogus pipeline) may result in a strong implicit–explicit relationship that shows much less evidence of dissociation than has been commonly observed.

Furthermore, the results of the current study provide additional evidence for the utility of implicit measures. Despite the obvious value of implicit measures as being more resistant to social desirability forces (e.g. Banse, Seise, & Zerbes, 2001), some have questioned the value of measures like the IAT (Cameron et al., 2000). One criticism of implicit measures has been that they are, at best, only weakly related to explicit measures. Ironically, the current results suggest that the apparent poor convergent validity of implicit measures is partially due to a shortcoming of explicit measures. When the impact of social desirability concerns on explicit attitudes is minimized, the convergent validity of implicit measures markedly improves.

Future research could examine the behavioral correlates of the implicit–explicit relationship under varying conditions. Research thus far indicates that explicit attitudes are generally correlated with deliberative responses and that explicit measures are typically not correlated with subtle, spontaneous behaviors (Dovidio et al., 1997). Research could examine whether explicit attitudes that are assessed while participants are 'on the pipeline' are more predictive of subtle interracial behaviors. Furthermore, Plant, Devine, and Brazy (2003) demonstrated that there are important individual differences in the nature of the attitudes that Whites' report while on the bogus pipeline. Additional

research could examine whether these individual differences may moderate the magnitude of the implicit–explicit relationship under bogus pipeline conditions.

Notes

1. The time between completion of the pre-test and participation in the laboratory portion of the study ranged from 3 to 11 weeks. The time elapsed between the pretest and the laboratory session did not appear to influence participants' responses on the MRS; the time elapsed (in weeks) was not correlated with MRS scores reported in the laboratory session, or the change in MRS scores from pre-test to the laboratory session.

2. In their meta-analysis of the validity of the bogus pipeline procedure, Roese and Jamieson (1993) explicitly recommended that participants should be asked to predict the output of the pipeline device, rather than just being instructed to answer accurately. This recommendation was for several reasons. One, this response format yields larger effects than does asking participants to only respond accurately. Two, a substantial majority of past research with the bogus pipeline has employed this type of response format. Three, this response format does not appear to simply induce a demand to respond more extremely, but rather it yields responses that are more accurate (Arkin & Lake, 1983). For example, Quigly-Fernandez and Tedeschi (1978) demonstrated that participants who were asked to guess the output of a bogus EMG apparatus gave responses that more accurately reflected their actual experiences.

3. IAT scores were also calculated using alternative analytic procedures including Greenwald et al.'s (2003) revised scoring procedure as well as a natural logarithmic transformation. Our results were somewhat weaker, although still very similar, using these alternative scoring procedures.

4. We also performed two ANOVAs in which each IAT effect was analyzed separately. Nearly identical results were obtained using this approach.

5. This unexpected increase in MRS scores in the Inaccurate condition may have been due to the fact that participants were aware that their responses on the MRS would be compared to their IAT scores. It is also possible that when participants in the Inaccurate condition were asked 'what the computer would predict' their responses to be on the MRS, they may have guessed the computer would inaccurately label them as having relatively negative attitudes toward African-Americans.

6. These significant differences among post-test MRS scores remained significant even after statistically controlling for pre-test MRS scores using a covariance analysis, ps < .05.

7. The reliability of this composite measure (alpha = .50) was similar to the reliability of other implicit measures (Dovidio et al., 2001), which generally have lower reliabilities than traditional self-report measures of attitudes.

8. While no effects involving the Order variable were significant at the .05 level, the main effect for the order condition was marginally significant, indicating the MRS scores were slightly higher when participants completed the IAT first. Also the interaction between Order and the second Information condition contrast was marginally significant. This effect indicated that the elevation of MRS scores in the Inaccurate and Accurate condition was greater when the IAT was completed first. While these effects suggest that Order may have had some impact on MRS scores, neither of these two marginal effects qualified the major hypothesized effects.

9. Two separate regression analyses were performed to examine whether the latency effect and the error rate effect could predict MRS scores when each IAT effect was treated independently, instead of being combined into a composite measure. In each of these analyses the critical interaction effect was marginally significant ps < .10, one-tailed).

Acknowledgements

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interaction.* Newbury Park, CA: Sage.

Arkin, R. M., & Lake, E. A. (1983). Plumbing the depths of the bogus pipeline: A reprise. *Journal of Research in Personality, 17,* 81–88.

Banaji, M. R., & Greenwald, A. G. (2000, March 19). Pride and prejudice. In G. Moriba-Meadows (Producer), *Dateline NBC.* New York: National Broadcasting Corporation.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48*, 145–160.

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 359–374). Mahwah, NJ: Erlbaum.

Boniecki, K. A., & Jacks, J. Z. (2002). The elusive relationship between measures of implicit and explicit prejudice, *Representative Research in Social Psychology, 26*, 1–14.

Cameron, J. A., Alvarez, J. M., & Bargh, J. A. (2000). *Examining the validity of implicit and explicit measures of prejudice: Is there really a bonafide pipeline?* Paper presented at the first annual meeting of the Society for Personality and Social Psychology, Nashville, TN.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* New York: Erlbaum.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and change. *Psychological Science, 12*, 163–170.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316–328.

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

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 3–32). Newbury Park, CA: Sage.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology, Volume 4: Intergroup processes* (pp.175–197). Oxford, UK: Blackwell.

Dovidio, J. F., Kawakami, K., & Gaertner S. L. (2002). Implicit and explicit racial attitudes and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62–68.

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). The nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510–540.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 23*, 316–326.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*, 1013–1027.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology. 74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: An improved scoring system. *Journal of Personality and Social Psychology, 85*, 197–216.

Jones, J. M. (1997). *Prejudice and racism.* New York: McGraw-Hill.

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin, 27*, 212–225.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–126). Orlando, FL: Academic Press.

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19*, 395–417.

Plant, E. A., Devine, P. G., & Brazy, P. C. (2003). The bogus pipeline and motivations to respond without prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes and Intergroup Relations, 6*, 187–200.

Quigley-Fernandez, B., & Tedeschi, J. T. (1978). The bogus pipeline as a lie detector: Two validity studies. *Journal of Personality and Social Psychology, 36,* 247–256.

Roese N. J., & Jamieson, D. W. (1993). Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin, 114*, 363–375.

Schlenker, B. R., & Weingold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology, 43*, 133–168.

Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology, 18*, 247–255.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*, 101–126.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*, 262–274.