Persuasive message scrutiny as a function of implicit-explicit discrepancies in racial attitudes

India R. Johnson [a,*], Richard E. Petty [b], Pablo Briñol [c], Ya Hui Michelle See [d]

[a] Elon University, United States
[b] The Ohio State University, United States
[c] Universidad Autónoma de Madrid, Spain
[d] National University of Singapore, Singapore

HIGHLIGHTS

- We test a variation of the "Watchdog Hypothesis" based on the idea of implicit ambivalence.
- Those low in explicit prejudice and high in implicit prejudice demonstrated enhanced scrutiny of race-relevant messages.
- We also test if individuals high in explicit prejudice but low in implicit prejudice will engage in more processing.
- As evaluative discrepancies increased, we found enhanced persuasive scrutiny, regardless direction of discrepancy.

ARTICLE INFO

ABSTRACT

Past research has shown that individuals low in prejudice think more carefully when information is from or about stigmatized individuals than non-stigmatized individuals. One explanation for this effect is that the heightened scrutiny stems from a motivation to guard against potential prejudice toward stigmatized others (i.e. "watchdog motivation"). The present research tested a variation of the watchdog hypothesis based on the idea of implicit ambivalence. Specifically, we argue that among individuals low in explicit (i.e., deliberative) prejudice, it is those who are also relatively high in implicit (i.e., automatic) prejudice who will do the most processing in prejudice relevant contexts. The implicit ambivalence framework also makes a novel prediction that individuals who are relatively high in explicit prejudice but low in implicit prejudice would also engage in enhanced information processing. As predicted, people with racial implicit-explicit attitude discrepancies, regardless of the direction of discrepancy, were found to engage in greater of scrutiny of a message about the hiring of Black faculty (study 1), a message about a Black job candidate (study 2), and even when the Black concept was merely primed subliminally prior to reading a race-irrelevant message (study 3).

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The race of a person can influence how much scrutiny he or she receives from other people. In an initial series of studies on this topic, White and Harkins (1994) presented Caucasian participants with a persuasive message from a White or a Black source who advocated for instituting senior comprehensive exams. The message contained either strong or weak arguments. Across several replications, they consistently found that the impact of argument quality on attitudes was greater when the source was Black rather than White. Subsequent research

has shown that the enhanced scrutiny effect also applies to situations in which the target rather than the source of the message is from a stigmatized group (Fleming, Petty, & White, 2005). The present research examines whether the enhanced scrutiny of Blacks by Whites stems from a variation of the "watchdog hypothesis" that is consistent with the idea of Whites having implicit ambivalence toward Blacks (Petty, Briñol, & Johnson, 2012).

2. The "watchdog hypothesis" and enhanced scrutiny

In a series of studies (Fleming et al., 2005; Petty, Fleming, & White, 1999) the enhanced scrutiny of stigmatized sources and targets by Whites was attributed to a "watchdog motivation." The main idea was that White individuals might be processing messages from or about Blacks and other stigmatized groups more than the same messages

* Corresponding author at: Elon University, Department of Psychology, Campus Box 2337, Elon, NC 27244, United States.
E-mail addresses: ijohnson5@elon.edu (I.R. Johnson), petty.1@osu.edu (R.E. Petty), pablo.brinnol@uam.es (P. Briñol), psysyhm@nus.edu.sg (Y.H.M. See).

from or about Whites in order to guard against some possible prejudice they might display in the situation. For example, thinking about a message about a Black individual carefully would be one way of ensuring fairness. Petty et al. (1999) reasoned that if a watchdog motivation were operating, it should primarily be Whites who were relatively low in prejudice who would show the enhanced scrutiny effect since these individuals would be most disturbed by showing prejudice.

To examine this notion, prejudice toward Blacks was assessed with explicit self-report measures (e.g., Katz & Hass, 1988), and reactions to persuasive messages from Black and White sources were evaluated. In several studies White individuals who were very low in prejudice were the ones who processed messages more for Black than White sources. This enhanced scrutiny effect by low prejudiced individuals was replicated when the message was about a Black versus a White target individual rather than from a Black versus a White source (Fleming et al., 2005).

In sum, a number of experiments have shown White individuals who are low in explicit prejudice toward Blacks tend to scrutinize a message either from a Black source or about a Black target more carefully than when the same message is from or about a White person. The current research aims to explore the watchdog motivation further by proposing that if low prejudiced individuals are motivated to engage in enhanced scrutiny of Blacks to look out for their own possible prejudice, then it should mostly be those individuals who have some reason to suspect possible prejudice on their part who would engage in this scrutiny. That is, people who do not want to be prejudiced or who see themselves as unprejudiced (low explicit prejudice) but who harbor automatic negative reactions toward Blacks (high implicit prejudice) would experience conflict and be the most vigilant in guarding against prejudice (see also, Devine, Monteith, Zuwerink, & Elliot, 1991; Monteith, 1993; Monteith & Devine 1993).

One way to identify individuals who have some prejudice about which they could be concerned is to use a measure of automatic bias. Over the past decade, several techniques for assessing automatic prejudice have been developed (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998). The goal of each is to assess an evaluation that automatically comes to mind when an attitude object is encountered. If watchdog motivation stems from a desire to watch out for one's own possible prejudice and we are able to assess both deliberative (explicit) and automatic (implicit) prejudice, then it should be the case that among those who are relatively low in their explicit prejudice, it is individuals who are also relatively high in their implicit prejudice who are the most likely to show enhanced scrutiny of stigmatized others. Individuals who are low in explicit prejudice and also low in implicit prejudice should have nothing to fear with respect to their own prejudice (i.e., have nothing to watch out for), and thus should be less likely to engage in enhanced scrutiny. We test this implication of the watchdog hypothesis for scrutiny of messages from or about Blacks by White individuals across three studies.

3. Implicit ambivalence and information processing

In addition to examining a unique implication of the watchdog hypothesis, the current research also examines the implications of recent work on *implicit ambivalence* (Petty, Tormala, Briñol, & Jarvis, 2006), for the enhanced scrutiny effect (Petty & Briñol, 2009; Petty et al., 2012; Shoda, McConnell, & Rydell, 2014). Implicit ambivalence refers to a situation in which an endorsed reaction is contrary to an unendorsed or rejected automatic reaction that is linked to an attitude object (Petty & Briñol, 2006). Implicit ambivalence differs from explicit ambivalence which refers to conflict stemming from endorsed positivity and negativity (Kaplan, 1972; see van Harreveld, Nohlen, & Schneider, 2015). Put simply, if a person explicitly endorses both positive and negative evaluations of Blacks, that person is explicitly ambivalent and as in other cases of explicit ambivalence would report feeling mixed about the attitude object (Priester & Petty, 1996). However, if a person

endorses largely positive reactions to Blacks (low explicit prejudice) but has negative reactions automatically spring to mind whenever the minority group is mentioned (high implicit prejudice) the person does not report feeling mixed about the attitude object. In this sense, the ambivalence is implicit rather than explicit because the person does not report being ambivalent toward the attitude object (see Petty et al., 2006) even though some discomfort is associated with that object (e.g., Rydell, McConnell, & Mackie, 2008; see Petty et al., 2012, for further discussion).[1] Critically, the implicit ambivalence framework stipulates that implicit ambivalence can be assessed by a discrepancy in the valence of an attitude uncovered by an implicit versus explicit attitude measure (see Briñol, Petty, & Wheeler, 2006). Thus, people who are relatively low in explicit prejudice toward Blacks on a self-report measure but relatively high in implicit prejudice as measured by an assessment of automatic attitudes, would have racial implicit ambivalence.

Prior research in non-racial domains has shown that people who have relatively large discrepancies between their implicit and explicit evaluations of a variety of objects express more discomfort about the attitude object (e.g., Rydell et al., 2008) and are more likely to process information relevant to those objects than are people for whom implicit and explicit evaluations are relatively low in discrepancy. In one study (Briñol et al., 2006), for example, as implicit-explicit discrepancies in self-esteem increased, so too did processing of a message relevant to the self as indexed by a greater difference in responses to arguments that varied in their quality. The discomfort from this discrepancy is presumed to motivate the processing of information whenever the discrepancy is activated such as when confronted by information relevant to the discrepancy (Briñol et al., 2006; Johnson & McDonough-Caplan, 2016; Petty et al., 2006; Rydell et al., 2008). Thus, implicit ambivalence has a similar impact on information processing as does explicit ambivalence (e.g., Maio, Bell, & Esses, 1996). Interestingly, the implicit ambivalence hypothesis predicts the same information processing outcome as the watchdog hypothesis. In fact, the watchdog hypothesis can be viewed as a particular instantiation of the more general implicit ambivalence hypothesis. That is, the watchdog hypothesis focuses on a particular discrepancy between having relatively low explicit racial prejudice but relatively high automatic prejudice.

Importantly, in the present studies we also test a novel hypothesis from the implicit ambivalence perspective not anticipated or examined in prior research on racial prejudice. That is, not only should people who are relatively lower in their explicit prejudice than they are in their implicit prejudice engage in enhanced scrutiny of information from or about Blacks, but so too should individuals who are relatively higher in their explicit prejudice than they are in their implicit prejudice.[2] This idea follows directly from the implicit ambivalence notion because any discrepancy between automatic and deliberative evaluations should produce implicit ambivalence regardless of the direction of the discrepancy. In watchdog terminology, just as individuals who are favorable toward Blacks on deliberative measures may want to be sure that they are not unduly unfavorable in their reactions (and thus watch out for any automatic or gut negative reactions), so too might individuals who are high in explicit prejudice want to be sure that they are not unduly positive (and thus watch out for any automatic or gut positive reactions they might have). Put simply, the implicit ambivalence

---

[1] People might reject an automatic evaluative association for a number of reasons. They might realize that it stems from the culture (e.g., media exposure) and not from personal beliefs (e.g., Devine, 1989). Or, the association can represent the opinions of others that have been encoded (e.g., Han, Olson, & Fazio, 2006; Priester & Petty, 2001). In addition, the association can represent a previously accepted personal view that has more recently been discredited (e.g., Gregg, Seibt, & Banaji, 2006; Petty et al., 2006; for a review, see Petty & Briñol, 2009). In all of these cases, unendorsed automatic negativity can conflict with endorsed positivity.

[2] The original watchdog hypothesis (Petty et al., 1999) only anticipated the possibility that people low in their explicit prejudice might watch out for their implicit prejudice, and subsequent research on racial prejudice only examined one direction of discrepancy – the extent to which implicit prejudice exceeded explicit prejudice (e.g., Penner et al., 2010; Shoda et al., 2014), not the reverse direction of discrepancy.

notion argues that the larger the discrepancy between implicit and explicit evaluations, the more likely people are to process information about the stigmatized group carefully, compared to when the magnitude of discrepancy between implicit and explicit evaluations is low.

4. The current research

The present work investigates whether among those individuals who are relatively low in explicit prejudice, it is primarily those who are also relatively high in their implicit prejudice who are more likely to engage in enhanced processing of information from or about Blacks. Furthermore, we test the novel hypothesis that among those who are relatively high in explicit racial prejudice, it is those who are also relatively low in their implicit prejudice who are more likely to engage in enhanced processing. Across three studies, we varied how the persuasive communication or context related to Blacks. Additionally, White students' attitudes toward Blacks were assessed using both automatic (implicit) and deliberative (explicit) measures in each study. In order to capture the magnitude of discrepancy, an index was formed as the absolute value of the difference between the standardized explicit and implicit measures of racial attitudes. We chose to operationalize implicit ambivalence in this manner for several reasons. Foremost, as noted earlier, this approach is what has been used in prior research on implicit ambivalence and thus allows us to compare our results to prior work (Briñol et al., 2006; Johnson & McDonough-Caplan, 2016; Rydell & Durso, 2012; Rydell et al., 2008). Second, this approach allows a clear test of the hypothesis that the overall *magnitude* of the discrepancy between implicit and explicit racial attitudes should predict the extent of scrutiny of information from or about Blacks. Just as various conceptualizations of explicit ambivalence focus on the magnitude of conflict rather than the positive or negative evaluations themselves, our approach allows us to do the same with respect to implicit ambivalence. Finally, indexing the extent of discrepancy in this manner also allows us to most directly examine our prediction that *direction* of discrepancy does not matter for information processing (i.e., implicit score more prejudiced than explicit or vice versa). Thus, in all analyses we also code for the direction of discrepancy to see if it moderates the results.

In addition to completing the implicit and explicit measures of racial attitudes, all of the participants in each study were exposed to a persuasive message containing strong or weak arguments. As in past research, the strong arguments were designed to elicit primarily favorable thoughts if people thought about them, whereas the weak arguments were designed to elicit mostly unfavorable thoughts when scrutinized (see Petty & Cacioppo, 1986). The degree to which participants think carefully about the message information was assessed by examining the extent to which the quality of the arguments affected post-message attitudes toward the advocacy (Petty, Wells, & Brock, 1976). If thinking about the message information presented is relatively high, as it should be when implicit-explicit attitude discrepancies are relatively high, argument quality should have a larger impact on attitudes than if thinking about the message information presented is relatively low, as it should be when implicit-explicit attitude discrepancies are relatively low. In statistical terms, our prediction is that magnitude of discrepancy should interact with argument quality to predict attitudes, such that argument quality is more likely to impact attitudes as discrepancies become larger. To the extent that this interaction is obtained, it would support the idea that implicit ambivalence is a contributor to the enhanced processing of information from or about minority groups identified in prior research. Furthermore, in terms of the "watchdog hypothesis," if this two-way interaction is not further moderated by the direction of discrepancy, it would suggest that people are watching out to resolve their own racial implicit ambivalence regardless of whether their implicit prejudice is greater than their explicit or vice-versa.

5. Study 1: Attitudes toward a Black faculty hiring program

5.1. Method

5.1.1. Participants and design

Sixty-eight White undergraduate psychology students (49 females, 19 males; $M_{age} = 19.56$, $SD = 1.26$) at the Ohio State University (OSU) participated in partial fulfillment of a course requirement. Data collection began within a few weeks of the close of the term; thus, we adopted a time-based stopping rule, choosing to end data collection at the close of the semester. Students were randomly assigned to receive either strong or weak arguments in favor of a proposal to hire more Black faculty at their university. Additionally, implicit and explicit measures of racial attitudes were assessed for all participants so we could form an index of implicit-explicit discrepancy. The independent variables thereby constituted an Argument Quality (strong vs. weak) × Implicit-Explicit Discrepancy (continuously scored) × Direction of Discrepancy (higher prejudice on explicit or implicit measure) design. Finally, in this and all subsequent studies, all exclusions, as well as manipulations and measures germane to our hypotheses are reported.

5.1.2. Procedure

Upon arrival, participants were seated at individual computer stations and were presented with all materials on the computer using MediaLab software (Jarvis, 2000). First, participants completed the automatic attitude measure toward Blacks, a racial Implicit Association Test (IAT; Greenwald et al., 1998). This task was introduced as a research project in which participants were to classify different words into different combinations of categories. After the IAT task, participants were told that due to extra time remaining in the session, they would also be participating in a research study to assess their opinions toward a new policy at OSU. Participants then received a persuasive message containing strong or weak arguments. After reading the message, participants reported their opinions toward the proposed program and completed the Anti-Black Scale (Katz & Hass, 1988), our measure of explicit prejudice.

5.1.3. Independent variables

5.1.3.1. Argument quality.
The strong and weak messages were borrowed from previous research and were pre-tested, such that the strong version produced mostly favorable thoughts about the proposal whereas the weak version produced mostly counterarguments (see Briñol, Petty, & McCaslin, 2009). An example strong argument was that since the number and quality of professors would increase with this program (without any tuition increase), the number of students per class could be reduced by 25%. In contrast, an example weak argument was that implementing the program would allow current professors to have more free time for themselves (see Briñol et al., 2009).

5.1.3.2. Explicit measure of prejudice.
In this and all subsequent studies, participants completed the Anti-Black Scale (Katz & Hass, 1988), rating 10 items on 6-point scales ranging from *strongly disagree* (0) to *strongly agree* (5). An example item is "On the whole, Black people don't stress education and training." Ratings on items were inter-correlated ($\alpha = 0.85$), so they were averaged to form one overall attitude index for each participant. In addition, the quality of the message did not influence participants' responses on the measure, $F(1, 66) = 1.85$, $p = 0.18$, $\eta^2 = 0.03$.

5.1.3.3. Implicit measure of prejudice.
The automatic measure was an implicit association test (IAT) in which five stereotypically Black names (i.e., Tyrone, LaToya, Lamar, Malik, Jamal) and five White names (i.e., Andrew, Katie, Josh, Brandon, Colleen) were paired with five positive (i.e., gold, joy, smile, peace, paradise) and five negative (i.e., poison, abuse, pain, death, filth) words (see Greenwald et al., 1998, for the

scoring procedure and rationale). The stimuli appeared within a centered white window. Reminder labels were positioned on top of the stimuli on the left and right side. These reminders read "Black" and "White" for single target-classification blocks, "pleasant" and "unpleasant" for single attribute-classification blocks. Mixed target + attribute blocks were also accompanied by appropriate labels (e.g., "pleasant or White" and "unpleasant or Black"). Incorrect classifications were followed by error feedback (i.e., the word "ERROR"). Summary feedback was provided at the end of each block informing participants about their average response latencies and percentage of errors for that block. All practice tasks in the IAT were administered in five blocks. In this and all subsequent studies, data-collection tasks consisting of combined target + attribute classifications were administered in four blocks. Within each block, stimuli were randomly selected without replacement and no more than two consecutively presented stimuli belonged to the same category.

The measure of relative automatic preference for White over Black was calculated using the response latencies for: (Black + pleasant and White + unpleasant) minus (Black + unpleasant and White + pleasant; i.e., the IAT effect). To correct for anticipatory responses and momentary inattention, latencies <300 ms and >3000 ms were recorded as 300 and 3000 ms respectively (e.g., see Greenwald et al., 1998). The latencies were log transformed to normalize the distribution. In addition, although the message was presented to participants before the IAT, as expected, the quality of the message did not influence participants' IAT responses, $F(1, 66) = 0.00$, $p = 1.00$, $\eta^2 \leq 0.001$.

5.1.3.4. Implicit-explicit discrepancy. The explicit and implicit measures of attitudes were unrelated to each other ($r = 0.03$, $p = 0.79$), a common finding in the literature on racial prejudice (e.g., Correll, Park, Judd, & Wittenbrink, 2002; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1995). An index of the magnitude of the implicit-explicit discrepancy was formed as the absolute value of the difference between the standardized explicit and implicit measures of racial attitudes. The discrepancy index considers where people fall within the distribution of participants in the study on the implicit versus explicit measures, and thus is a relative measure of discrepancy. A zero on the index indicates that the person's place in the distribution is exactly the same on the implicit and explicit measures (e.g., high in the distribution on both, low in the distribution on both, middling on both, etc.). Discrepancies can be in either direction; people can be higher in prejudice in the sample distribution on the explicit measure than the implicit measure (a positive discrepancy) or they can be lower in the distribution on the explicit measure than the implicit measure (a negative discrepancy). We calculated the absolute value of the difference between the two standardized measures as our key index of implicit-explicit discrepancy (see also, Briñol et al., 2006; Durso & Rydell, 2012; Johnson & McDonough-Caplan, 2016).

5.1.3.5. Direction of discrepancy. To code for the direction of discrepancy, we subtracted the standardized implicit score from the standardized explicit score. Difference scores greater than zero were coded as a positive discrepancy (i.e. explicit score > implicit score), whereas difference scores less than zero were coded as a negative discrepancy (i.e. implicit score > explicit score). Difference scores ranged from −2.75 to 3.82, and no participant had a difference score equivalent to zero. Thirty-five participants were coded as having a negative discrepancy and thirty-three participants were coded as having a positive discrepancy.

5.1.4. Dependent variable: Attitudes toward the hiring program

Our key dependent measure used to gauge degree of scrutiny was attitudes toward the new Black faculty hiring program. Attitudes toward the program were assessed using three 9-point (1–9) semantic differential scales (i.e., bad-good, against-in favor, harmful-beneficial). Ratings on these items were highly inter-correlated ($\alpha = 0.90$), and were averaged to form one overall attitude index for each participant.

5.2. Results and discussion

Attitudes toward the proposal were submitted to a hierarchical regression analysis, with (1) magnitude of implicit-explicit discrepancy (centered continuous variable), (2) direction of the discrepancy (effects coded; −1 = implicit prejudice > explicit prejudice vs. 1 = explicit prejudice > implicit prejudice), and (3) argument quality (effects coded; −1 = weak vs. 1 = strong) as the IVs. In this and all subsequent studies, main effects were interpreted in the first step, two-way interactions in the second step, and the three-way interaction in the third step. Furthermore, in all cases where the expected Magnitude of Discrepancy × Argument Quality interaction emerged, simple slope analyses were conducted where low and high discrepancy scores were centered at one standard deviation below and above the mean, respectively (see Cohen, Cohen, West, & Aiken, 2003).

Results revealed a main effect of argument quality, such that strong arguments ($M = 7.12$, $SD = 1.65$) produced more positive attitudes than weak arguments, ($M = 6.41$, $SD = 1.84$), $B = 0.92$, $t(64) = 2.09$, $p = 0.04$, 95% CI: 0.041, 1.79. Of most importance, the predicted two-way interaction between the magnitude of participants' discrepancy and argument quality emerged, $B = 1.07$, $t(61) = 2.18$, $p = 0.03$, 95% CI: 0.09, 2.05. Simple slope analyses examining effects within high discrepancy and low discrepancy individuals confirmed that the argument quality effect was greater among high than low discrepancy individuals. Among participants with high implicit-explicit discrepancies in racial attitudes, evaluations of the proposal to hire more Black faculty were significantly impacted by argument quality such that more positive evaluations followed the strong rather than weak message, $B = 2.01$, $t(61) = 3.07$, $p = 0.003$, 95% CI: 0.71, 3.32. Thus, high discrepancy participants were processing information regarding the hiring program carefully. In contrast, attitudes among participants with low implicit-explicit discrepancies were not significantly influenced by argument quality, $B = 0.04$, $t(61) = 0.07$, $p = 0.95$, 95% CI: −1.15, 1.22, thus suggesting that low discrepancy participants were not processing the information as carefully (see top panel of Fig. 1).

No other significant effects emerged, ps > 0.21. Most importantly, the three-way interaction of argument quality, implicit-explicit discrepancy, and direction of discrepancy was not significant, $B = −0.28$, $t(60) = −0.55$, $p = 0.59$, 95% CI: −1.31, 0.75 (see Supplemental File A for full statistics for Direction). The absence of a three-way interaction suggests that high discrepancy participants scrutinized the message to a greater extent than low discrepancy participants held regardless of the direction of the discrepancy. These results indicate that among those who were relatively low in their explicit prejudice, it was primarily those who were relatively high in implicit prejudice who engaged in greater scrutiny of a message about a program favoring Blacks. However, among those who were relatively high in explicit prejudice, it was those who were relatively low in implicit prejudice who engaged in the greatest scrutiny.

6. Study 2: Attitudes toward a Black person

Study 1 provided initial support for our hypothesis about processing of discrepancy-relevant information based on the notion of implicit ambivalence. Our second study was designed to accomplish several objectives. The first goal was to provide a conceptual replication of the first study with a different paradigm. In Study 1, participants read a proposal to increase the hiring of Black faculty and thus the attitude was about the category of Black professors. In Study 2, however, participants were asked to evaluate one particular Black individual. We highlight this shift in the target of judgment for two reasons. First, past research in the domain of stereotyping and prejudice has demonstrated different evaluations have emerged when Whites have judged racial categories (e.g., Blacks) versus individuals (e.g., Tyrone) (e.g., see Fiske, Neuberg, Beattie, & Milberg, 1987; Brewer, 1988). Specifically, judgments made at the category level often tend to be consistent with the negative
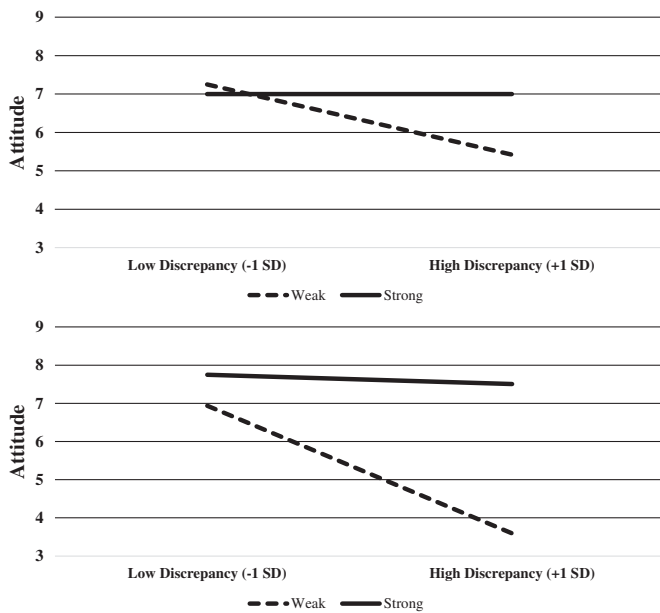
Fig. 1. Top panel: Attitudes as a function of argument quality and racial implicit-explicit discrepancies (one standard deviation below and above the mean) (Study 1). Bottom panel: Attitudes as a function of argument quality and racial automatic-deliberative discrepancies (one standard deviation below and above the mean) (Study 2).

stereotype of that group, but when people are asked to focus on a particular individual, evaluations tend to be more reflective of the person's abilities rather than negative associations tied their category membership. Thus, the effects of implicit-explicit discrepancies on information processing might only apply when the message to be evaluated deals with a racial category, as in Study 1, and not a particular individual. On the other hand, the notion of implicit ambivalence suggests that implicit-explicit discrepancies in racial attitudes would trigger enhanced scrutiny whenever the discrepancy is activated which should be the case whether the information received is about a category of Blacks or an individual Black person. Consequently, we expected to replicate the findings from Study 1 although the information was about a particular person rather than a class of individuals.

A second reason to move to assessment of an individual rather than a racial category is a practical one. People spend time interacting with individuals of different races, they do not interact with racial categories. Although it is potentially important to understand how implicit and explicit racial attitudes affect processing of information about race-linked issues or categories (e.g., affirmative action), it may be even more important to understand how implicit and explicit racial attitudes affect scrutiny of particular individuals. In Study 2, participants reviewed the curriculum vitae of one Black job candidate for a position in their own University's psychology department. Participants were randomly assigned to read a vita containing strong or weak credentials and attitudes toward the job candidate were assessed. Finally, participants' racial attitudes were again measured using both implicit and explicit measures. However, in this study, both of these measures followed presentation of the job candidate information. By placing the implicit measure before the message in Study 1, it might have sensitized participants to any racial discomfort they might have had. Thus, in Study 2, both measures of racial attitudes followed presentation of the information to be scrutinized.

In another departure from Study 1, in Study 2 the IAT was comprised of pictures of Black and White individuals rather than names in order to demonstrate convergent validity across different measurement variations. We not only switched to a picture IAT but we also improved our scoring of the IAT to be consistent with more recent recommendations

regarding the computation of D-scores (Greenwald, Nosek, & Banaji, 2003).[3]

6.1. Method

6.1.1. Participants and design

Fifty-nine White psychology students (39 females, 20 males; $M_{age} = 19.08$, $SD = 1.51$) at Ohio State University (OSU) participated in exchange for partial course credit. Data collection took place the final week of the fall semester and thus we again adopted a time-based stopping rule, choosing to end data collection at the close of the term. Participants were randomly assigned to receive either a vita containing strong or weak credentials. Additionally, implicit and explicit measures of racial attitudes were assessed for all participants so that an index of implicit-explicit discrepancy could be formed. The independent variables thereby constituted a Vita Quality (strong vs. weak) × Implicit-Explicit Discrepancy (continuously scored) × Direction of Discrepancy (higher on explicit or implicit measure) design.

6.1.2. Procedure

Participants were escorted to individual computer stations and presented with all materials on the computer using MediaLab software (Jarvis, 2000). They were instructed today's study involved their evaluation of a prospective job candidate for the department of psychology. Participants' were informed OSU wanted to learn students' opinions about potential faculty candidates prior to hiring them. In addition, participants were told OSU was committed to hiring the 'best and the brightest' and the best predictors of a job candidate's future excellence 'is the amount and quality of previous research and teaching experience in the field of Psychology.'

Participants' were then informed they had been randomly selected to evaluate a job candidate, and all participants viewed the vita of a candidate named Tyrone Edwards (cf., Petty et al., 2006). Participants viewed a vita containing either strong or weak attributes and then reported their attitudes toward Tyrone as a potential faculty member. Participants' were then told it was necessary for them to complete a 'personality questionnaire' to help organize their responses, which included the racial Implicit Association Test (IAT; Greenwald et al., 1998) and the Anti-Black scale (Katz & Hass, 1988), the same measures of racial attitudes as used in Study 1, though the IAT used racial pictures instead of race-identified first names.

6.1.3. Independent variables

6.1.3.1. Vita quality.
Half of the participants viewed a vita which contained strong information implying that Tyrone would be well qualified for the position in Psychology, whereas the remaining half reviewed a vita containing weak information suggesting that Tyrone would be poorly qualified to fill the position. The two vitas were pretested, such that the strong version produced mostly favorable thoughts about the job candidate whereas the weak version of the vita elicited mostly unfavorable thoughts (adapted from Petty et al., 2006).

The vita containing strong arguments indicated that the job candidate had earned his Ph.D. from Stanford University and had been the recipient of several national awards for both his research and teaching ability, clearly indicating that Tyrone was well qualified for the position. In contrast, the weak vita indicated that Tyrone had yet to defend his dissertation and to date 'had only published two articles in journals of medium quality.' Thus, the weak vita plainly indicated that Tyrone was not as well-suited as the strong job candidate for the faculty position.

---

[3] The new scoring method was not used for Study 1 because the data were collected prior to 2003 and the raw reaction time data are no longer available. That is, the IAT data for Study 1 were recorded and saved using the original recommendations for scoring IAT data and thus were analyzed using that method (see Greenwald et al., 1998).

6.1.3.2. Explicit measure of prejudice. The explicit measure and scoring were identical to Study 1 (Katz & Hass, 1988), and the ratings of items on the scale were highly inter-correlated ($\alpha = 0.84$). In addition, the presentation of the differing vitas was not related to participants' responses on the measure, $F(1, 57) = 2.30$, $p = 0.13$, $\eta^2 = 0.04$.

6.1.3.3. Implicit measure of prejudice. The racial IAT (Greenwald et al., 1998) was again used to assess participants' level of implicit prejudice. However, in Study 2, photos of Black and White faces were paired with evaluative stimuli. Evaluative stimuli consisted of ten positive (i.e. freedom, health, family, peace, cheer, friend, heaven, loyal, pleasure, gentle) and ten negative (i.e. abuse, crash, filth, stink, assault, disaster, pollute, divorce, jail, ugly) words. Outside of these adjustments, administration of the task was consistent with Study 1.

Each participants' D-score was calculated (see Greenwald et al., 2003), eliminating trials with response latencies faster than 300 ms and slower than 3000 ms, and all participants whose overall accuracy was <80%. Larger D-scores reflect relatively greater implicit negativity against Blacks. Finally, the presentation of the differing vitas was not related to participants' IAT scores, $F(1, 57) = 0.26$, $p = 0.61$, $\eta^2 = 0.005$.

6.1.3.4. Implicit-explicit discrepancy. The measures of explicit and implicit racial attitudes were unrelated to one another ($r = 0.11$, $p = 0.40$). An index of implicit-explicit discrepancy was formed using the same procedure as Study 1.

6.1.3.5. Direction of discrepancy. We again coded for direction of discrepancy. Difference scores ranged from −2.92 to 1.65 and no participant's difference score was equivalent to zero; twenty-seven participants were coded as having a positive discrepancy while thirty-one were coded as having a negative discrepancy.

6.1.4. Dependent variable: Attitudes toward the job candidate

Attitudes toward the job candidate were assessed using six 9-point (1–9) semantic differential scales (i.e., bad-good, against-in favor, harmful-beneficial, foolish-wise, negative-positive, unfavorable-favorable) on which they rated the quality of Tyrone as a potential faculty member. We added three additional attitude items in this study to the ones used in Study 1 in order to have an even more reliable and complete measure that included both cognitive and affective features in ratings of the job candidate (see Supplemental File B for analyses in which our dependent variables were the three attitude measures from Study 1). Ratings on these items were highly inter-correlated ($\alpha = 0.94$), so they were averaged to form one overall attitude index for each participant.

6.2. Results and discussion

Analyses were conducted consistent with that of Study 1. Attitudes toward the job candidate were submitted to a hierarchical regression analysis, with (1) magnitude of the implicit-explicit discrepancy (centered continuous variable), (2) direction of the discrepancy (effects coded; −1 = implicit prejudice > explicit prejudice vs. 1 = explicit prejudice > implicit prejudice), and (3) argument quality (effects coded; −1 = weak vs. 1 = strong) as the IVs. Results revealed a main effect of argument quality, such that the strong vita ($M = 6.80$, $SD = 0.85$) produced more positive attitudes toward the job candidate than the weak vita ($M = 4.88$, $SD = 1.43$), $B = 0.926$, $t(54) = 6.13$, $p < 0.0001$, 95% CI: 0.62, 1.23.

Critically, the predicted two-way interaction between the magnitude of participants' implicit-explicit discrepancies and argument quality emerged, $B = 0.59$, $t(51) = 3.23$, $p = 0.002$, 95% CI: 0.22, 0.95. Simple slope analyses confirmed that among participants with relatively high implicit-explicit discrepancies, attitudes were significantly impacted by argument quality such that more positive attitudes followed the strong rather than weak vita, $B = 1.55$ $t(51) = 6.75$, $p < 0.0001$,

95% CI: 1.08, 2.01. Thus, high discrepancy participants were scrutinizing the information contained in the job candidate's vita carefully. In contrast, the impact of argument quality on attitudes among participants with relatively low implicit-explicit discrepancies was smaller, $B = 0.46$, $t(51) = 2.25$, $p = 0.029$, 95% CI: 0.05, 0.87, suggesting that low discrepancy participants were not processing the information as carefully as high discrepancy participants (see bottom panel of Fig. 1).

Consistent with Study 1 and the implicit ambivalence framework, no main effects or interactions of direction of discrepancy emerged and the critical three-way interaction of discrepancy, argument quality and direction was non-significant, $B = 0.003$, $t(50) = -0.001$, $p = 0.99$, 95% CI: −0.61, 0.61 (see Supplemental File C for full statistics for Direction). The absence of moderation by direction suggests that regardless of the nature of the divergence between implicit and explicit racial attitudes, as discrepancy between these measures increased, participants' engaged in greater information processing. Additionally, this study revealed that the enhanced scrutiny that results from implicit-explicit discrepancy does not just occur when the information is about the general category relevant to the discrepancy (as shown in Study 1), but also when the information is about a specific member of that category.

7. Study 3: Processing a persuasive message when race is primed

The present work thus far provides evidence that as discrepancies between one's implicit and explicit racial attitudes increase, greater processing of a Black-relevant message also occurs. Additionally, consistent with previous work examining implicit ambivalence, implicit-explicit discrepancy was not moderated by direction of one's discrepancy. The goals for our final study were two-fold.

First, one criticism of the previous studies is we may have failed to uncover the moderation by direction due to a lack of power as a result of our limited sample size. Consequently, to ensure this was not the case, for Study 3 we significantly increased our sample size. Specifically, we utilized G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) to determine that a sample size of at least 327 was needed in order to ensure our study achieved a minimum power of 80%, assuming an effect size for the discrepancy × argument quality interaction consistent with that of study 1, the smaller effect size of studies 1 & 2 (i.e. $f^2 = 0.06$ & 0.20, respectively).

Second, one important issue to address is that in our first two studies, all participants received a message relevant to Blacks in some way. Because of this, it is possible that people with discrepancies in their implicit-explicit attitudes might generally be prone to engage in information processing, such that even if we gave them material not relevant to Blacks in any way, they would still process it more intently. In contrast, if implicit ambivalence is involved, the mere existence of a discrepancy in one domain should not result in the indiscriminate processing of any information present. Enhanced thinking is expected only if the discrepancy is activated in some way as it would be if the message was relevant to the dimension (e.g., see Briñol et al., 2006). If the discrepancy is not activated, the mere presence of an implicit-explicit discrepancy on a given topic is not expected to enhance information processing.

To address this issue, our third study included a manipulation in which either the Black concept or a control concept (Buddhist Monk) was subliminally primed prior to reading a persuasive message unrelated to race. Additionally, the choice to activate the discrepancy-relevant concept in such a subtle manner was intentional. In Studies 1 & 2, the discrepancy-relevant concept was made salient via the persuasive message itself. This has also been the standard procedure in prior research on the scrutiny effect in which the race of the source or target to be evaluated was manipulated in a blatant way (e.g., Fleming et al., 2005; Petty et al., 1999; White & Harkins, 1994). One open question, then, is whether an explicit focus on the Black concept is necessary for the enhanced processing to occur (see Supplemental File D for details on a pilot experiment examining an explicit focus on the Black versus White concept and the enhanced processing effect).

Based on the notion of ambivalence being *implicit* when discrepancies arise between explicit and implicit racial attitudes, an explicit focus on the topic of the discrepancy may not be necessary. In the previous studies, it is likely that the presence of a message relevant to Blacks activated any discrepancy in implicit versus explicit racial attitudes and the resulting discomfort led people to be more vigilant in their information processing. However, if merely activating the discrepancy is sufficient to produce the discomfort that is presumed to lead to enhanced information processing, then even more subtle prompts to the discrepancy relevant concept could be effective in activating the discrepancy and producing enhanced information processing. That is, as long as the discrepancy is activated by some cue relevant to Blacks within the persuasion context, people could feel uncomfortable and may therefore process whatever information is before them in an attempt to address or understand this discomfort (cf., Bless, Bohner, Schwarz, & Strack, 1990; Rydell et al., 2008). Indeed, the entire notion of implicit ambivalence hinges on the assumption that people may not be clear about the precise source of their discomfort (Petty et al., 2006), yet they feel uncomfortable nonetheless. It is for this reason that the discomfort from implicit ambivalence, like cognitive dissonance (e.g., Zanna & Coooper, 1974), can be misattributed to something else (see Rydell & Durso, 2012). Thus, if the discomfort is not linked to any obvious cue in the current environment, then it might lead people to more carefully attend to and process whatever information is available in the immediate context, even if that information is not particularly relevant to the discrepancy. The question of whether explicit cues to race from the message environment (source, target, topic) are necessary for racial attitude discrepancies to enhance information processing or whether subtler activation of race is sufficient is examined in Study 3.

Thus, in Study 3, participants who varied in their implicit-explicit racial attitude discrepancies were exposed to a persuasive message on the race irrelevant-topic of senior comprehensive exams and were asked to report their attitudes toward the proposal in the message. Prior to reading the message, however, participants were subliminally primed with either the Black concept or Buddhist Monk concept. The former prime should presumably activate any racial attitude discrepancies but do so more subtly than in the previous two studies or prior research on this topic. Additionally, the Buddhist Monk prime choice was intentional as a critic may argue that activating the concept of White is still relevant to race and one's automatic and/or deliberate racial attitudes. The Buddhist Monk concept, however, is low in relevance to racial attitude discrepancies. Thus Study 3 extends prior work on implicit ambivalence as it is the first study to examine the impact of activating the discrepancy-related dimension outside of the persuasive message.

7.1. Method

7.1.1. Participants and design

Three hundred and thirty-five White undergraduates (168 females, 167 males; $M_{age} = 19.25$, $SD = 2.14$) were recruited at both The Ohio State University (OSU) and Elon University.[4] OSU participants were compensated with partial course credit whereas Elon participants took part in exchange for course credit or a $25 Target gift card. One hundred and thirty-one participants were recruited from OSU; this was the number attained after the first four weeks of the semester, our initial a priori

stopping point. Because the OSU data collection failed to reach a sample size sufficient to attain the desired power, 204 additional participants were recruited from Elon across two semesters. Participants were treated identically across the two locations and results were similar after controlling for the effect of institution. Thus, the two samples were combined to ensure maximum power.

In each location participants first completed a task in which they were subliminally primed with either the Black concept or the Buddhist monk concept. Participants were then randomly assigned to read a persuasive message unrelated to race, containing either strong or weak arguments. Consistent with the previous studies, implicit and explicit measures of racial attitudes were assessed for all participants in order to form an index of implicit-explicit discrepancy. Consequently, the independent variables constituted a Prime (Black concept vs. Buddhist Monk concept) × Argument Quality (strong vs. weak) × Implicit-Explicit Discrepancy (continuously scored) × Direction of Discrepancy (higher on explicit or implicit measure) design.

7.1.2. Procedure

Participants were seated at individual cubicle stations and MediaLab (Jarvis, 2000) was again used to administer all elements of the experiment. Participants were presented with a brief introduction stating they were going to be completing two unrelated experiments, packaged together for convenience. The first experiment, sponsored by the "Cognitive Sciences Department" was a word recognition task, our means of subliminally priming either the Black or Buddhist Monk concept. Participants then moved on to the second experiment and viewed a proposal in favor of implementing senior comprehensive exams at their university, containing either strong or weak arguments. Finally, participants rated their attitudes toward the proposal and completed the same racial IAT and Anti-Black scale items used in study 2.

7.1.3. Independent variables

7.1.3.1. Subliminal priming.
To activate the Black concept, participants were subliminally primed using a lexical decision task from previous research (see DeMarree & Loersch, 2009). Specifically, participants were instructed to indicate whether a presented letter string was a word (e.g. walnut) or a non-word (e.g. nuwalt). Participants were to hit the "Z" key if the presented letter string was a non-word and the "?" key if the letter string was a word. Prior to each letter string presentation, the primed concept was first presented for 12 ms, followed by a mask (XXXXX) for 225 ms. For those participants randomly assigned to the Buddhist Monk conditions, the word "Buddhist" or "Monk" was presented subliminally before each letter-string combination. For those participants randomly assigned to the Black concept conditions, the word "Black" or "African American" was presented subliminally before each letter-string condition. Participants completed a total of 30 trials.

7.1.3.2. Argument quality.
Participants also received either strong or weak arguments regarding a proposal in favor of implementing senior comprehensive exams. The strong message in favor of the proposal contained such arguments as Ivy League schools and several Big Ten universities have adopted the exams to maintain their record of excellence. In contrast, the weak message in favor of senior comprehensive exams included such arguments as one student claimed that the history of such exams can be traced to the ancient Greeks and was therefore a tradition that should continue. These messages have been used widely in previous research (see Petty & Cacioppo, 1986) and were pre-tested so that strong arguments elicited mostly favorable thoughts and the weak arguments elicited mostly unfavorable thoughts when people were instructed to think carefully about them.

---

[4] Data collection proceeded first with the OSU sample and then a few years later with the Elon sample due to the change in institutions of the first author. A preliminary analysis after the OSU data collection was conducted and although the key predicted three-way interaction was significant in this sample, a decision was made to increase sample size to attain more power to detect a possible moderating effect of direction of discrepancy. Most importantly, our key predicted finding, the three-way interaction of extent of discrepancy × argument quality × prime, was significant for both our Elon sample, $B = 0.587$, $t(192) = 5.30$, $p < 0.0001$, 95% CI: 0.37, 0.81 and the OSU sample, $B = 0.403$, $t(115) = 2.15$, $p = 0.03$, 95% CI: 0.03, 0.77. In essence, then, the Elon sample can be considered as an independent replication of the OSU sample.

7.1.3.3. Explicit measure of prejudice. The explicit prejudice measure and scoring were identical to Studies 1 and 2 (Katz & Hass, 1988). Once again, ratings of scale items were highly inter-correlated ($\alpha = 0.80$). Unlike the previous studies, there was an unexpected significant impact of the prime on the Anti-Black measure, $F(1 331) = 4.41$, $p = 0.03$, $\eta^2 = 0.01$, such that those subliminally primed with the African American concept expressed more negative attitudes toward Blacks ($M = 3.46$, $SD = 0.83$) relative to those primed with the Buddhist Monk concept ($M = 3.27$, $SD = 0.82$). Both the argument quality manipulation and, of most importance, the interaction of prime and argument quality were non-significant, $F(1 331) = 0.55$, $p = 0.46$, $\eta^2 = 0.002$, and $F(1 331) = 2.15$, $p = 0.14$, $\eta^2 = 0.006$, respectively.[5]

7.1.3.4. Implicit measure of prejudice. The racial Implicit Association test (IAT; Greenwald et al., 1998) was again used to assess participant's level of implicit prejudice and was completed after reading the message. Scoring was identical to that of study 2. Neither the prime, $F(1, 331) = 0.61$, $p = 0.44$, $\eta^2 = 0.005$, nor the argument quality manipulation, $F(1, 331) = 0.72$, $p = 0.39$, $\eta^2 = 0.002$, affected this measure, and there was no interaction either, $F(1, 331) = 0.09$, $p = 0.76$, $\eta^2 \leq 0.0010$.

7.1.3.5. Implicit-explicit discrepancy. Measures of explicit and implicit racial attitudes were again unrelated to one another ($r = 0.033$, $p = 0.54$). An index of implicit-explicit discrepancy was formed using the same procedure as in the prior studies.

7.1.3.6. Direction of discrepancy. We again coded for direction of discrepancy. Difference scores ranged from -3.46 to 3.77 and no participant's score was equivalent to zero. Among them, 169 were coded as having a negative discrepancy while the remaining 166, a positive discrepancy.

7.1.4. Dependent variable: Attitudes toward the proposal

Attitudes toward senior comprehensive exams were assessed using the same six 9-point (1–9) semantic differential scales utilized in study 2. Ratings on these items were highly inter-correlated ($\alpha = 0.94$), so they were averaged to form one overall attitude index for each participant.

7.2. Results and discussion

Attitudes toward comprehensive exams were submitted to a hierarchical regression analysis, with (1) magnitude of implicit-explicit discrepancy (centered continuous variable), (2) direction of the discrepancy (coded; -1 = implicit prejudice > explicit prejudice vs. 1 = explicit prejudice > implicit prejudice), (3) argument quality (coded; -1 = weak vs. 1 = strong), and (4) subliminal prime (coded: -1 = Buddhist Monk vs. 1 = Black) as the IVs, with the four-way interaction interpreted in the final step (Cohen et al., 2003)
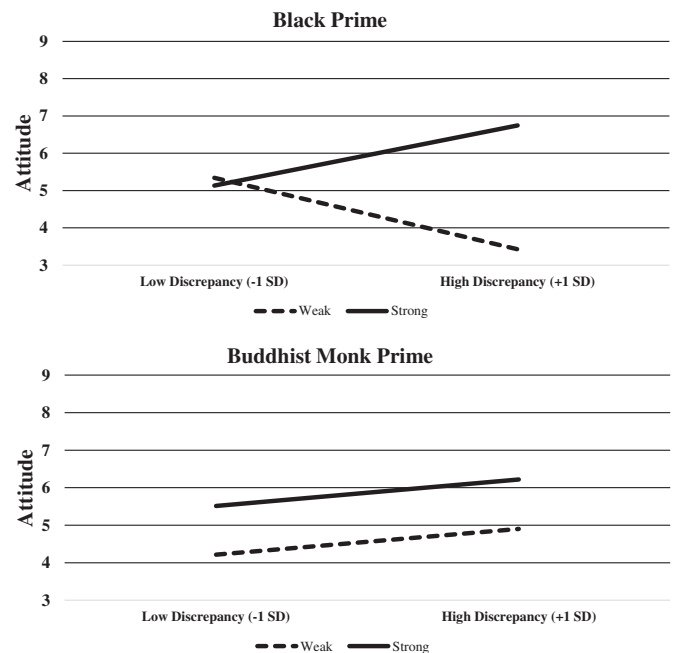
Results revealed a main effect of argument quality, such that strong arguments ($M = 7.22$, $SD = 1.42$), produced more positive attitudes toward the senior comprehensive exam proposal than weak arguments ($M = 5.71$, $SD = 1.48$), $B = 0.848$, $t(330) = 10.45$, $p < 0.0001$, 95% CI: 0.69, 1.01. A main effect of prime also emerged, such that those primed with the Buddhist Monk concept ($M = 5.71$, $SD = 1.62$) produced more positive attitudes toward the proposal than those primed with the Black concept ($M = 5.16$, $SD = 1.80$), $B = -0.268$, t

$(330) = -3.29$, $p = 0.001$, 95% CI: -0.428, -0.108. In addition, the two-way interactions of magnitude of discrepancy and argument quality, $B = 0.520$, $t(325) = 5.445$, $p < 0.0001$, 95% CI: 0.333, 0.708, discrepancy and prime, $B = -0.203$, $t(325) = -2.06$, $p = 0.04$, 95% CI: -0.398, -0.009, and argument quality and prime, $B = 0.224$, $t(325) = 2.89$, $p = 0.004$, 95% CI: 0.072, 0.376, were all significant.

However, of most importance, these two-way interactions were all qualified by the predicted three-way interaction between the magnitude of participants' Discrepancy, Argument Quality, and Prime, $B = 0.587$, $t(322) = 6.24$, $p < 0.0001$, 95% CI: 0.40, 0.77. We decomposed the three-way interaction as a function of prime, and found that the two-way interaction between the magnitude of participants' discrepancy and argument quality was significant for those subliminally primed with the Black concept, $B = 1.08$, $t(161) = 9.11$, $p < 0.0001$, 95% CI: 0.84, 1.31 (see top panel of Fig. 2), replicating the previous studies.

Simple slope analyses were conducted to examine effects in high and low discrepancy individuals in the Black prime condition and revealed that among Black prime participants who had high implicit-explicit discrepancies, attitudes were impacted by argument quality such that more positive attitudes followed the strong rather than weak message, $B = 1.66$, $t(161) = 14.69$, $p < 0.0001$, 95% CI: 1.43, 1.88. Thus, high discrepancy participants were processing information regarding the proposal carefully. In contrast, among Black Prime participants who had low implicit-explicit discrepancies, attitudes were not impacted by argument quality, $B = -0.102$, $t(161) = -0.644$, $p = 0.521$, 95% CI: -0.42, 0.21, suggesting that low discrepancy participants were not processing the information as carefully. Finally, consistent with the prior studies, within the Black prime condition, the three-way interaction of Discrepancy, Argument Quality, and direction was non-significant, $B = -0.17$, $t(161) = -1.33$ $p = 0.19$, 95% CI: -0.41, 0.08. In fact, no main effects or interactions of Direction of discrepancy emerged for those primed with the Black concept ($ps > 0.56$; see Supplemental File E for full statistics).

In addition, there was no interaction between argument quality and discrepancy in the Buddhist monk prime condition, $B = 0.006$, $t(162) = 0.042$, $p = 0.97$, 95% CI: -0.26, 0.27 (see bottom panel of Fig. 2). The only effects to emerge for the Buddhist monk prime participants was a main effect of argument quality, $B = 0.65$, $t(164) = $

---

[5] Because our subliminal prime manipulation unexpectedly influenced Anti-Black Scores, we also examined the impact of our independent variables on our implicit-explicit discrepancy index. Results indicated that neither the main effect of argument quality, $F(1 331) = 1.28$, $p = 0.26$, $\eta2 = 0.004$, prime, $F(1 331) = 0.14$, $p = 0.71$, $\eta2 = 0.000$, nor the interaction of the two, $F(1 331) = 0.03$, $p = 0.85$, $\eta2 = 0.000$, were-significant, suggesting that the prime manipulation did not impact our key predictor of enhanced processing (i.e. the discrepancy index). Additionally, given that the pattern of results for Study 3 is consistent with our previous studies in which the Anti-Black scale was unaffected by any manipulations, we felt confident that use of our use of the discrepancy index in this study to predict enhanced scrutiny was not compromised.

Fig. 2. Attitudes as a function of prime, argument quality and racial implicit-explicit discrepancies (one standard deviation below and above the mean) (Study 3).

5.69, $p < 0.0001$, 95% CI: 0.43, 0.88, such that strong arguments produced more positive attitudes toward the proposal than weak arguments, in addition to a main effect of Discrepancy, $B = 0.39$, t (164) = 2.83, $p = 0.005$, 95% CI: 0.12, 0.66, which indicated that higher discrepancies were associated with more positive attitudes toward the proposal.

These results suggest that individuals who are implicitly ambivalent with respect to their racial attitudes– high in implicit prejudice, low in explicit or the reverse – engage in greater scrutiny of information when race was activated as evidenced by a greater impact of argument quality on attitudes. When race was not activated by the prime, however, the magnitude of discrepancy had no impact on information processing.

In sum, study 3 conceptually replicated our previous findings by showing that as implicit-explicit discrepancies in racial attitudes increase, people become more likely to scrutinize a message as long as the Black concept is activated. However, the present study extends past work by illustrating that the discrepancy-relevant concept need not be activated directly by some feature of the persuasive message. Even when activated outside of the persuasive message, those who are more discrepant in their racial implicit and explicit attitudes engage in greater processing. As in Studies 1 and 2, this conclusion was supported by the finding that the attitudes of relatively discrepant individuals were more reflective of the quality of the information they received than were the attitudes of less discrepant individuals. The fact that enhanced information processing with larger implicit-explicit discrepancies only occurred when the Black concept was activated and not when the Buddhist monk concept was activated, is consistent with the idea that the information processing is not necessarily a conscious strategy but may instead result from the diffuse discomfort that the activated discrepancy produces (cf., Petty et al., 2012; Rydell & Durso, 2012).

8. General discussion

As noted in introducing our research, investigators initially uncovered a phenomenon in which White individuals tended to engage in greater scrutiny of information when it was presented by a Black rather than a White source (see White & Harkins, 1994). Subsequent research showed that this enhanced processing of Black over White sources extended to greater processing of Black over White targets (Fleming et al., 2005) and was most likely to occur among individuals who were low in their explicit prejudice. It was assumed that because these individuals would be concerned about being prejudiced when assessing information from or about Blacks, they would guard against this possible prejudice by processing the information very carefully (Petty et al., 1999). Contemporary research suggests that many White individuals who score low in explicit prejudice might harbor automatic negative reactions to Blacks and that they might wish to overcome these negative reactions in order to act in an unprejudiced way (Monteith, 1993; Plant & Devine, 1998). These automatic negative reactions are captured in contemporary implicit measures of racial attitudes such as the IAT (Greenwald et al., 1998).

The goal of the current research was to provide a clear empirical test of whether the enhanced scrutiny of Black sources and targets by Whites who are relatively low in their explicit prejudice could stem from a desire to watch out for the possibility that they might act in a prejudiced way. Consistent with this reasoning, we found that among individuals who were relatively low in explicit prejudice, it was those who were also relatively high in implicit prejudice who were more likely to process messages about Blacks (Studies 1 and 2) or about a race irrelevant topic if the Black concept was subtly activated outside of the message (Study 3). If both explicit and implicit prejudice were relatively low or high, then there was no enhanced scrutiny of information. These results are consistent with the notion that people low in explicit prejudice may be processing messages from and about Blacks in an attempt

to guard against their own implicit racial prejudice, though this effort may not be consciously or deliberately invoked.

In addition, however, the present work also uncovered a result that was not anticipated by the original watchdog hypothesis. That is, the direction of implicit-explicit discrepancy did not matter for information processing. An increasing discrepancy between explicit and implicit attitudes in either direction was sufficient to enhance information processing as long as the discrepancy-relevant concept was activated in some way. Thus, individuals who were relatively high in explicit prejudice but relatively low in implicit prejudice were also more likely to process information when Black was activated compared to individuals who were more consistent in their explicit and implicit prejudice. Our statistical comparison involved including a factor of direction of discrepancy in our analysis of magnitude of discrepancy. In no case did direction of discrepancy moderate the results which should have occurred if one direction of discrepancy was more responsible than the other direction for the two-way interaction between magnitude of discrepancy and argument quality.

Overall, then, individuals who were discrepant in their degree of implicit and explicit prejudice were more likely to process information when the Black race was salient than were individuals who had low discrepancies. This result suggests the parsimonious conclusion that the enhanced information processing with increasing implicit-explicit attitude discrepancies may result from the discomfort that accompanies implicit ambivalence (see Petty & Briñol, 2009; Petty et al., 2006, 2012; Rydell et al., 2008). This is analogous to prior work showing that people engage in greater processing to reduce the discomfort from explicit ambivalence (e.g., Maio, Bell, & Essess, 1996) and from cognitive dissonance (e.g., Cooper, 2007). It also resonates with the view that negative affect more generally can signal that there is something problematic in the environment that may require enhanced information processing to address (e.g., Bless et al., 1990). Our final study is especially informative in this regard in that enhanced information processing occurred among individuals with implicit-explicit racial discrepancies in a situation in which the Black concept was merely primed and nothing about the persuasive message itself was relevant to race.

8.1. Absolute versus relative discrepancies

Although our research showed that direction of discrepancy did not moderate the results, it is important to note that direction of discrepancy was calculated on a *relative* basis, which is the standard practice in this domain (e.g., Briñol et al., 2006; Rydell et al., 2008). As noted previously, we adopted this *relative* approach because it allows us to directly examine the impact of the magnitude of discrepancy and the direction of discrepancy along with their interaction. By standardizing both explicit and implicit attitude measures and then using the absolute value of the difference between them, we created a relative magnitude of discrepancy index for each participant and also coded for direction of discrepancy. Increasing discrepancies in favor of the explicit measure means that people were progressively higher in the prejudice distribution in the sample tested on the explicit than the implicit measure. In contrast, increasing discrepancies in favor of the implicit measure means that people were progressively higher in the prejudice distribution in the sample tested on the implicit than the explicit measure. Because of this procedure, we had comparable numbers of individuals on each side of this discrepancy (implicit > explicit and explicit > implicit), and we were able to show that direction of the relative discrepancy did not moderate the effects of amount of discrepancy on information processing.

There is an alternative way to use our relative discrepancy index in an analysis and that is to examine the index without taking the absolute value. Without taking the absolute value, the discrepancy index includes direction; whereas, negative discrepancies indicate that implicit attitudes were more prejudiced than explicit attitudes and positive

discrepancies indicate that explicit attitudes were more prejudiced than implicit attitudes. To examine the data using the simple difference in standardized attitude scores (i.e., no absolute values taken) we conducted an analysis in which we used the difference in standardized attitude scores and tested for curvilinear argument quality effects across levels of discrepancy. According to our conceptualization, in this alternative analysis, both high positive and high negative discrepancy individuals should process the race-relevant message more (i.e., show a larger argument quality effect) than people who have little discrepancy. That is, since the direction of discrepancy is not predicted to matter, both high positive and high negative discrepancies should look similar. When analyzed in this manner, the results from our experiments look as predicted. Fig. 3 shows the significant interaction between discrepancy (quadratic term) and argument quality collapsing all three studies, including study as a factor. For this analysis, Studies 1 and 2 along with the Black prime conditions for study 3 were included. The figure shows that the argument quality effect on attitudes is smallest with the lowest discrepancy scores and then increases as discrepancy increases in either direction for both those who received strong arguments, $B = 0.067$, $t (158) = 3.02$, $p = 0.003$, and weak arguments, $B = -0.25$, $t (129) = -6.39$, $p < 0.0001$.

A final methodological point to consider is that on the surface, the present findings may appear to diverge from past research examining Whites' implicit and explicit attitudes toward Blacks. Specifically, a variety of research examining implicit and explicit attitudes toward Blacks has found, in general, White individuals typically display automatic negativity toward Blacks on implicit measures, but endorse more positive attitudes toward Blacks on explicit measures (Baron & Banaji, 2006; Dovidio et al., 1997; Turner, Hewstone, & Voci, 2007). In other words, much of previous research has found that Whites tend to be high in implicit prejudice, but low in explicit prejudice. However, in the present work our sample of participants consisted of both combinations – individuals who were relatively high in implicit prejudice and low in explicit as well as individuals who were relatively low in implicit prejudice and high in explicit. Since the latter combination seems unique, the present work raises the question of what type of person is low in implicit prejudice, yet high in implicit prejudice?

First, it is critical to mention that in the present research, whether a person was classified as "high" or "low" in implicit or explicit prejudice

Fig. 3. Standardized attitudes as a function of argument quality and difference scores (i.e. Standardized Explicit Minus Standardized Implicit) across studies (Studies 1–3; only Black prime conditions for Study 3). Difference scores in the studies range from −3.00 to 3.82.

was done on a *relative* basis. Methodologically, although our relative discrepancy index does a good job of capturing people's relative standing in the distribution (as in prior research), our index can only deal with the range of scores observed in the sample. Thus, by an *absolute* criterion, it could be the case that all of the participants are relatively high in both explicit and implicit prejudice (e.g., everyone in the sample could have attitudes on the negative side of the scale), or be positive on the explicit prejudice measure and negative on the implicit prejudice measure, and so forth. If this were true, however, it would still be the case that some people were more negative or positive than others on each measure and that the relative discrepancies our index captures is successful in predicting the extent of information processing. Nevertheless, to explore this issue further, we approximated an "absolute" rather than a relative discrepancy index in each study by using the middle point of the explicit scale (e.g., 3.5 on a 0 to 5-point scale) and the zero point in the IAT to approximate absolute positive/negative evaluative differences between participants high and low on both of the racial attitude dimensions. Using an absolute criterion revealed that across all 3 studies only 32 participants out of 462 demonstrated a pattern whereby the implicit measure suggested absolute positivity but the explicit measure suggested absolute negativity.

Although few participants demonstrated low implicit but high explicit prejudice in an absolute terms, we offer some insight on the type of individual who might fit this combination. One possibility is that the implicit positivity could reflect personal associations, whereas the negativity on an explicit measure reflects explicit norms. For instance, media coverage and television programming featuring positive representations of Blacks has increased in recent years; and, thus one potential source of implicit positivity among our young student sample may be based on repeated exposure to positive exemplars of Blacks in the media (e.g. President Barack Obama). In contrast, an individual's deliberative attitude might be reflective of those in the person's immediate peer group or parents. Future work adopting an implicit measure that assesses normative evaluations (see Yoshida, Peach, Zanna, & Spencer, 2012) or personal beliefs (Olson & Fazio, 2004) could potentially explore this possibility. Although other ways in which this combination might come about can be speculated upon, we suggest that future research examine this potentially interesting group of individuals.

Of most importance, at the conceptual level, our hypothesis is that the magnitude of the discrepancy matters more than the direction of the discrepancy. For example, an individual with a slightly positive attitude toward Blacks on an explicit measure and slightly negative associations with Blacks on an implicit measure would be categorized as discrepant in absolute terms, whereas an individual with a slightly negative attitude toward Blacks on an explicit measure and an extremely negative association with Blacks on an implicit measure would not, even though the magnitude of difference between the implicit and explicit scores of the latter individual in the sample distribution would presumably be much larger and potentially more consequential in our view. Nevertheless, future research would benefit from exploring implicit and explicit discrepancies in other domains (e.g., sexual orientation, mentally ill, elderly, etc.) so that the issue of absolute versus relative discrepancies can be examined more systematically.

8.2. Explicit versus implicit prejudice

In the present work, we examined how the magnitude of individuals' relative implicit-explicit attitude discrepancies with respect to Blacks impacted their information processing. As discrepancies increased, so too did information processing as long as Black was made salient in some way. As a methodological note, an alternative statistical analysis is to look at the effects of implicit and explicit prejudice separately. Within the prejudice literature, there are numerous studies documenting a role for both implicit and explicit measures of prejudice. The most common findings are: (1) that implicit and explicit measures of prejudice predict different outcomes such as implicit measures
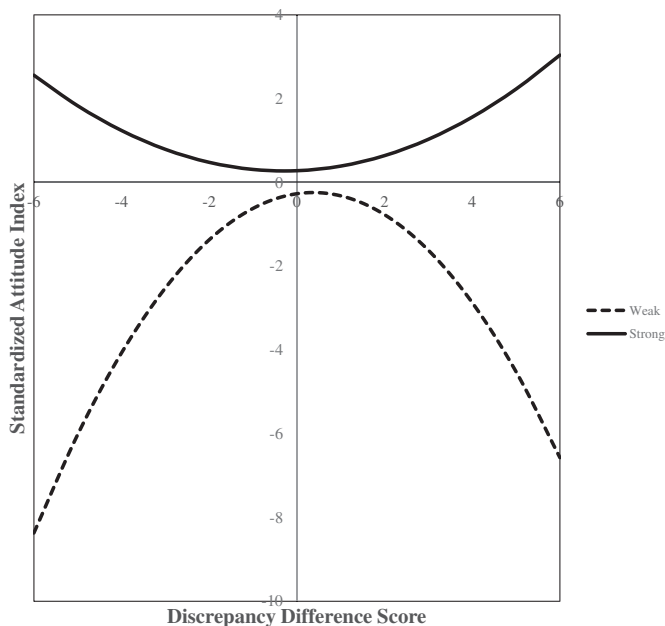
predicting automatic behaviors and explicit measures predicting deliberative behaviors (e.g., Dovidio et al., 1997) and (2) explicit and implicit measures predict unique main effect variance in the same outcome (e.g., see Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

In contrast, the current hypothesis is rather unique within the prejudice literature – that implicit and explicit measures of prejudice will interact to predict an outcome – information processing. To test this interaction hypothesis more directly, we conducted an Explicit Attitude × Implicit Attitude × Argument Quality regression analysis including Study as a factor. This analysis was conducted using combined data from Studies 1 and 2 along with the Black prime conditions from Study 3. As expected, we found a significant 3-way interaction of Explicit Attitude × Implicit Attitude × Argument Quality, $B = -0.26$, $t (285) = -3.57$, $p \leq 0.0001$, 95% CI: -0.40, -0.12, with no main effects or interactions with the study variable. This interaction is depicted in Fig. 4 where the top panel shows that among those who are relatively low in explicit prejudice, as implicit prejudice increases, argument quality has a larger impact on attitudes, $B = 0.33$, $t (285) = 3.66$, $p < 0.001$, 95% CI: 0.15, 0.50. The bottom panel shows that among those who are relatively high in explicit prejudice, as implicit prejudice decreases, argument quality has a larger impact on attitudes, $B = -0.18$, $t (285) = -2.00$, $p = 0.04$, 95% CI: -0.38, -0.003.

This analysis, comparable to the Magnitude of Discrepancy × Direction of Discrepancy × Argument Quality interaction reported earlier in each study, demonstrates in an alternative manner that discrepancies in either direction (implicit > explicit or explicit > implicit) lead to more information processing. We preferred our analysis because it provided a specific statistical test of whether direction of discrepancy matters whereas the alternative analysis does not. In both cases, however, the results suggest a more general phenomenon of people who are relatively high in implicit ambivalence processing more than those who are relatively low in implicit ambivalence and provides further evidence
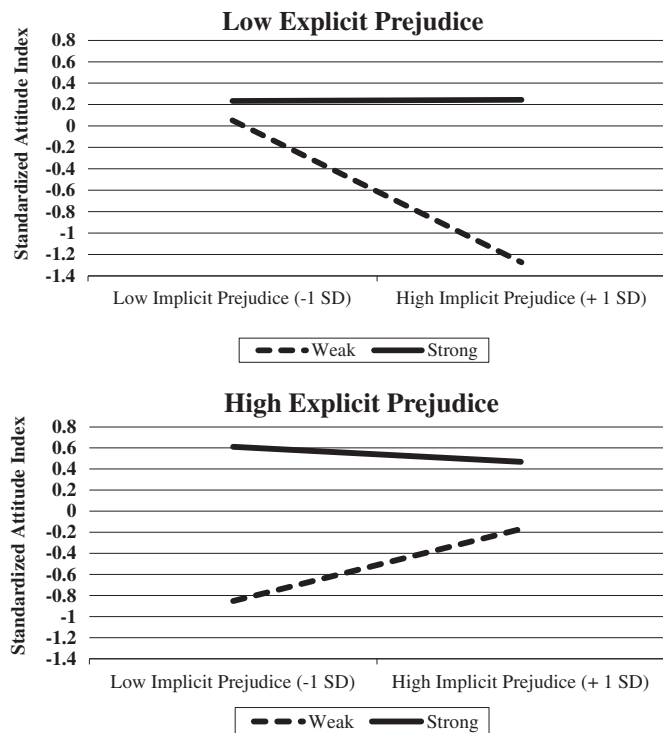
for the utility of assessing both explicit and implicit forms of prejudice. To the extent that explicit attitudes indicate the attitude that a person consciously endorses, any indication that this was not the reaction they were actually having (e.g., experiencing automatic negativity toward Blacks when positivity is endorsed or vice-versa), would lead to discomfort. This discomfort, like the discomfort that comes from explicit ambivalence and cognitive dissonance, would lead people to process relevant information more in an attempt to either reduce the discomfort (Rydell et al., 2008) or address the general negative feeling that makes the situation seem problematic (e.g., Bless et al., 1990).[6]

8.3. Implicit ambivalence vs. theories of prejudice

The present work examines the information processing consequences of implicit ambivalence in a racial domain. Although an individual can potentially be implicitly ambivalent toward any attitude object, because the present work focuses on racial implicit ambivalence one important question to address is how the implicit ambivalence framework fits with other theories of prejudice. One prominent theory relevant to the present work is aversive racism (Dovidio & Gaertner, 2004; Gaertner & Dovidio, 1981, 1986), which stipulates that strong egalitarian norms in the United States motivate White Americans to evaluate Black Americans positively; however, deep seeded biases can often lead to negative evaluation and treatment. Critically, the aversive racism framework stipulates that valuing egalitarianism while simultaneously possessing negative attitudes toward Black Americans can lead to feelings of ambivalence or discomfort stemming from holding conflicting explicit beliefs. Thus, examining the present work through the lens of aversive racism, one could argue that those high in implicit-explicit discrepancy were motivated to process the persuasive message carefully not because of the discomfort stemming from racial implicit ambivalence but rather from valuing egalitarian norms, while also harboring negative feelings toward Black Americans.

Despite the surface similarity, we believe the aversive racism account is insufficient to account for all of our data for two reasons. First, the aversive racism account seems especially suited to explain what happens when explicit attitudes toward Blacks are more favorable (e.g., from egalitarian values) than implicit attitudes (i.e., from negative automatic feelings about Blacks). However, the aversive racism framework is less well suited to account for the situation in which explicit attitudes are less favorable than implicit attitudes. Yet, our findings for both of these situations were the same. Second, the aversive racism account seems contingent on race being explicitly relevant. Although race is explicitly related to the persuasive message in our Studies 1 and 2, this is not the case for Study 3. Specifically, in Study 3, we chose to activate the concept of African American/Black subliminally and outside of the persuasive communication. Additionally, the persuasive message itself was unrelated to race (i.e. favoring senior comprehensive exams). Consequently, it seems unlikely that the enhanced scrutiny for those high in implicit-explicit discrepancy was due to discomfort stemming from harboring conflicting beliefs in an explicitly race-relevant situation but rather from the discomfort associated with implicit ambivalence itself.

Low Explicit Prejudice

High Explicit Prejudice

Fig. 4. Top panel: Standardized attitudes as a function of argument quality and implicit prejudice (one standard deviation below and above the mean) for low explicit prejudice (one standard deviation below) (Studies 1–3; only Black prime conditions for Study 3). Bottom panel: Standardized attitudes as a function of argument quality and implicit prejudice (one standard deviation below and above the mean) for high explicit prejudice (one standard deviation above) (Studies 1–3; only Black prime conditions for Study 3).

---

[6] In addition to completing the Anti-Black scale, participants also completed the Pro-Black Scale Katz & Hass, 1998) in all studies. Thus, the Anti and Pro black scale measures could be used to calculate a measure of explicit ambivalence using the most commonly employed ambivalence formula – the similarity intensity model (see Thompson, Zanna, & Griffin, 1995). Although explicit ambivalence did not interact with argument quality in any study, $B = -0.008$, $t (285) = -0.082$, $p = 0.93$, 95% CI: -0.20, 0.19, to ensure that the critical interaction of implicit-explicit discrepancy and argument quality predicted message processing above and beyond explicit ambivalence, we combined all the Black-relevant conditions from Studies 1-3 and entered both explicit ambivalence and the explicit ambivalence × argument quality interaction as additional predictors, also controlling for the effect of study. Critically, the key implicit-explicit discrepancy × argument quality interaction remained significant after controlling for both explicit ambivalence and its interaction with argument quality, $B = 0.93$, $t (285) = 7.92$, $p < 0.0001$, 95% CI: 0.70, 1.17. Thus, the impact of implicit ambivalence on information processing is above and beyond any possible impact from explicit ambivalence.

Nonetheless, future research is warranted to better delineate under what conditions an aversive racism versus an implicit ambivalence framework is applicable.

In accord with the meta-cognitive model of attitudes (MCM; see Petty, Briñol, & DeMarree, 2007), the present research also offers a unique perspective on various theories postulating that people are often motivated to correct for their internalized prejudice. According to several formulations (e.g., Dovidio & Gaertner, 2004; Dunton & Fazio, 1997), some White individuals have automatic negative reactions to Blacks, but have egalitarian values or desires not to be prejudiced that cause them to discount their activated negativity and report positivity on explicit measures. This analysis assumes that the causal sequence is that people have pre-existing automatic attitudes that are negative, some pre-existing motive to control these reactions, and these interact to determine a *constructed* (and potentially false or socially desirable) deliberative positive attitude. Although this is perfectly plausible and certainly can occur, our studies suggest another possibility – that motives can follow from pre-existing positive and negative associations, with one of them negated. That is, some people may recognize that they have both existing positive and negative associations, with the latter being unwanted or at least not endorsed. Because they find the latter to be inappropriate or wrong, they develop a motive to control these negative reactions. Conversely, some individuals may have an automatic positive association, and reject that in an attempt to form a more accurate or correct impression, and thus endorse a relatively more negative evaluation. Thus, rather than a positive constructed attitude following from the interaction of negative automatic attitudes and a motive to control them, it could be that a motive to control negative reactions follows from the presence of both positive and negative associations to a minority group with one of the two being rejected while the oppositely-valenced association is endorsed. Based on this possibility, future research should explore the role of motives to inhibit prejudice and how these motives potentially fit with the implicit ambivalence framework.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2016.11.007.

References

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*(1), 53–58. http://dx.doi.org/10.1111/j.1467-9280.2005.01664.x.

Bless, H., Bohner, G., Schwarz, N., & Strack, F. (1990). Mood and persuasion: A cognitive response analysis. *Personality and Social Psychology Bulletin*, *16*, 331–345. http://dx.doi.org/10.1177/0146167290162013.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull, & R. S. Wyer (Eds.), *A dual process model of impression formation* (pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum.

Briñol, P., Petty, R. E., & Wheeler, S. C. (2006). Discrepancies between explicit and implicit self-concepts: Consequences for information processing. *Journal of Personality and Social Psychology*, *91*, 154–170. http://dx.doi.org/10.1037/0022-3514.91.1.154.

Briñol, P., Petty, R. E., & McCaslin, M. J. (2009). *Automatic and deliberative attitude change from thoughtful and non–thoughtful processes. Attitudes: Insights from the new implicit measures.* Mahwah, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Cooper, J. (2007). *Cognitive dissonance: 50 years of a classic theory.* London, England: Sage Publications.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314–1329. http://dx.doi.org/10.1037/0022-3514.83.6.1314.

DeMarree, K. G., & Loersch, C. (2009). Who am I and who are you? Priming and the influence of self versus other focused attention. *Journal of Experimental Social Psychology*, *45*, 440–443. http://dx.doi.org/10.1016/j.jesp.2008.10.009.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. http://dx.doi.org/10.1037/0022-3514.56.1.5.

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, *60*, 817–830. http://dx.doi.org/10.1037/0022-3514.60.6.817.

Dovidio, J., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology. 36.* (pp. 1–52).

Dovidio, J., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510–540. http://dx.doi.org/10.1006/jesp.1997.1331.

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. http://dx.doi.org/10.1177/0146167297233009.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. -G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

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. http://dx.doi.org/10.1037/0022-3514.69.6.1013.

Fiske, S. T., Neuberg, S. L., Beattie, A. E., & Milberg, S. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, *23*, 399–427. http://dx.doi.org/10.1016/0022-1031(87)90038-2.

Fleming, M. A., Petty, R. E., & White, P. (2005). Stigmatized targets and evaluation: Prejudice as a determinant of attribute scrutiny and polarization. *Personality and Social Psychology Bulletin*, *4*, 496–507. http://dx.doi.org/10.1177/0146167204271585.

Gaertner, S. L., & Dovidio, J. F. (1981). *Racism among the well-intentioned. Pluralism, racism, and public policy: The search for equality, 208–222.*

Gaertner, S. L., & Dovidio, J. F. (1986). *The aversive form of racism.* Academic Press.

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. http://dx.doi.org/10.1037/0022-3514.74.6.1464.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197. http://dx.doi.org/10.1037/0022-3514.85.2.197.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41. http://dx.doi.org/10.1037/a0015575.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*(1), 1. http://dx.doi.org/10.1037/0022-3514.90.1.1.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the implicit association test. *Journal of Experimental Social Psychology*, *42*(3), 259–272. http://dx.doi.org/10.1016/j.jesp.2005.04.006.

Jarvis, W. B. G. (2000). *MediaLab (version 2006) [computer software].* Columbus, OH: Empirisoft.

Johnson, I. R., & McDonough-Caplan, H. (2016). Implicit-explicit evaluative discrepancies in depression and information seeking. *Social Cognition*, *34*(3), 238–253.

Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic-differential technique. *Psychological Bulletin*, *77*, 361–372.

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, *55*, 893–905. http://dx.doi.org/10.1037/0022-3514.55.6.893.

Maio, G. R., Bell, D. W., & Essess, V. M. (1996). Ambivalence in persuasion: The processing of messages about immigrant groups. *Journal of Experimental Social Psychology*, *32*, 513–536. http://dx.doi.org/10.1006/jesp.1996.0023.

Monteith, M. J. (1993). Self-regulation of prejudice responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*, 469–485. http://dx.doi.org/10.1037/0022-3514.65.3.469.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed versus other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology*, *64*, 198–210. http://dx.doi.org/10.1037/0022-3514.64.2.198.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the implicit association test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*(5), 653.

Penner, L. A., Dovidio, J. F., West, T. V., Gaertner, S. L., Albrecht, T. L., Dailey, R. K., & Markova, T. (2010). Aversive racism and medical interventions with Black patients: A field study. *Journal of Experimental Social Psychology*, *46*, 436–440.

Petty, R. E., & Briñol, P. (2006). A metacognitive approach to "implicit" and "explicit" evaluations: Comment on Gawronski and Bodenhausen (2006). *Psychological Bulletin*, *132*, 740–744. http://dx.doi.org/10.1037/0033-2909.132.5.740.

Petty, R. E., & Briñol, P. (2009). Implicit ambivalence: A meta-cognitive approach. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 119–164). New York, NY: Psychology Press.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Petty, R. E., Wells, G. L., & Brock, T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, *34*, 874–884. http://dx.doi.org/10.1037/0022-3514.34.5.874.

Petty, R. E., Fleming, M. A., & White, P. H. (1999). Stigmatized sources and persuasion: Prejudice as a determinant of argument scrutiny. *Journal of Personality and Social Psychology*, *76*, 19–34. http://dx.doi.org/10.1037/0022-3514.76.1.19.

Petty, R. E., Tormala, Z. L, Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, *90*, 21–41. http://dx.doi.org/10.1037/0022-3514.90.1.21.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: Implications for attitude measurement, change, and strength. *Social Cognition*, 25, 657–686. http://dx.doi.org/10.1521/soco.2007.25.5.657.

Petty, R. E., Briñol, P., & Johnson, I. (2012). Implicit ambivalence. In B. Gawronski, & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 178–201). New York, NY: Guilford Press.

Plant, A. E., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811–832. http://dx.doi.org/10.1037/0022-3514.75.3.811.

Priester, J. R., & Petty, R. E. (1996). The gradual threshold model of ambivalence: Relating the positive and negative bases of attitudes to subjective ambivalence. *Journal of Personality and Social Psychology*, 71, 431–449. http://dx.doi.org/10.1037/0022-3514.71.3.431.

Priester, J. R., & Petty, R. E. (2001). Extending the bases of subjective attitudinal ambivalence: Interpersonal and intrapersonal antecedents of evaluative tension. *Journal of Personality and Social Psychology*, 80(1), 19–34. http://dx.doi.org/10.1037/0022-3514.80.1.19.

Rydell, R. J., & Durso, G. R. O. (2012). Can I borrow a feeling? Spillover of negative arousal from inconsistent information during attitude formation diminishes perceptions of well-being. *Journal of Experimental Social Psychology*, 48, 575–578. http://dx.doi.org/10.1016/j.jesp.2011.10.018.

Rydell, R. J., McConnell, A. R., & Mackie, D. M. (2008). Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. *Journal of Experimental Social Psychology*, 44, 1526–1532. http://dx.doi.org/10.1016/j.jesp.2008.07.006.

Shoda, T. M., McConnell, A. R., & Rydell, R. J. (2014). Having explicit-implicit evaluation discrepancies triggers race-based motivated reasoning. *Social Cognition*, 32(2), 190–202. http://dx.doi.org/10.1521/soco.2014.32.2.190.

Thompson, M. M., Zanna, M. P., & Griffin, D. W. (1995). *Let's not be indifferent about (attitudinal) ambivalence. Attitude strength: Antecedents and consequences.* Mahwah, NJ: Erlbaum Associates.

Turner, R. N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology*, 93(3), 369. http://dx.doi.org/10.1037/0022-3514.93.3.369.

van Harreveld, F., Nohlen, H. U., & Schneider, I. K. (2015). The ABC of ambivalence: Affective, behavioral, and cognitive consequences of attitudinal conflict. *Advances in Experimental Social Psychology*, 52, 285–324. http://dx.doi.org/10.1016/bs.aesp.2015.01.002.

White, P., & Harkins, S. G. (1994). Race of source effects in the elaboration likelihood model. *Journal of Personality and Social Psychology*, 67, 790–807. http://dx.doi.org/10.1037/0022-3514.67.5.790.

Yoshida, E., Peach, J. M., Zanna, M. P., & Spencer, S. J. (2012). Not all automatic associations are created equal: How implicit normative evaluations are distinct from implicit attitudes and uniquely predict meaningful behavior. *Journal of Experimental Social Psychology*, 48, 694–706. http://dx.doi.org/10.1016/j.jesp.2011.09.013.

Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, 29, 703–709. http://dx.doi.org/10.1037/h0036651.