

Caught in the Middle: Defensive Responses to IAT Feedback Among Whites, Blacks, and Biracial Black/Whites

Social Psychological and
Personality Science
1-9

© The Author(s) 2014
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1948550614561127
spps.sagepub.com



Jennifer L. Howell¹, Sarah E. Gaither², and Kate A. Ratliff¹

Abstract

This study used archival data to examine how White, Black, and biracial Black/White people respond to implicit attitude feedback suggesting that they harbor racial bias that does not align with their self-reported attitudes. The results suggested that people are generally defensive in response to feedback indicating that their implicit attitudes differ from their explicit attitudes. Among monoracial White and Black individuals, this effect was particularly strong when they learned that they were implicitly more pro-White than they indicated explicitly. By contrast, biracial Black/White individuals were defensive about large discrepancies in either direction (more pro-Black or more pro-White implicit attitudes). These results pinpoint one distinct difference between monoracial and biracial populations and pave the way for future research to further explore how monoracial majority, minority, and biracial populations compare in other types of attitudes and responses to personal feedback.

Keywords

multiracial identity, automatic/implicit processes, defensive processes, attitudes, racial identity

Research suggests that the majority of Americans personally identify as racially egalitarian (Crandall, Eshleman, & O'Brien, 2002; McConahay, 1986; O'Brien et al., 2010) and that they will engage in a variety of cognitive strategies to preserve this view (O'Brien et al., 2010). For example, White individuals expecting that they would soon be confronted with feedback about their own racism sought out racially biased comparison targets, helping them to maintain a view that they are less biased than other people (O'Brien et al., 2010). Participants in another study derogated a measure of implicit attitudes when it suggested that they were more implicitly biased than they indicated explicitly (e.g., that they were more pro-straight, or associated Black individuals with weapons; Howell & Ratliff, 2014). But to date, research has yet to compare how these findings differ across racial groups, and we know very little about how racial minorities may respond. In this study, we use archival data to examine how White, Black, and biracial Black/White individuals respond to implicit attitude feedback suggesting that they harbor racial bias that does not align with their self-reported attitudes.

and White) and two evaluative concepts (e.g., good and bad). Perhaps the most common source of data using the IAT is the Project Implicit website (<https://implicit.harvard.edu>), which allows participants to complete an IAT, receive feedback about their implicit attitudes (e.g., your results suggest a strong automatic preference for White individuals compared to Black individuals), and answer questions about their opinions of the IAT. Despite often explicitly endorsing egalitarianism, rarely do visitors receive feedback indicating that they are, in fact, egalitarian (Axt, Ebersole, & Nosek, 2014).

So how do people react to IAT feedback that is potentially threatening to an egalitarian self-view? In one study, White participants who were led to believe the IAT might show they harbor unconscious pro-White biases often chose to avoid receiving IAT feedback altogether (Howell et al., 2013). Furthermore, when White participants learned that the IAT measured racial bias (compared to knowledge of a stereotype), they also displayed increased defensive outcomes such as higher pro-White bias on the measure (indicating

The Implicit Association Test (IAT) and Feedback for White Participants

One popular measure of racial attitudes is the IAT (Greenwald, McGhee, & Schwartz, 1998), which uses reaction times to assess associations between two target categories (e.g., Black

¹ University of Florida, Gainesville, FL, USA

² University of Chicago, Chicago, IL, USA

Corresponding Author:

Jennifer L. Howell, University of Florida, Psychology Building, P.O. Box 112250, Gainesville, FL 32611, USA.

Email: jenny.howell@ufl.edu

stereotype threat; Frantz, Cuddy, Burnett, Ray, & Hart, 2004) and higher levels of implicit (but not explicit) self-esteem, suggesting they automatically buffered their self-image prior to feedback (Rudman, Dohn, & Fairchild, 2007). Additionally, prior to receiving IAT feedback, White participants are generally able to detect that the IAT will indicate pro-White biases, prompting increased negative affect and external attributions for bias (e.g., errors in measurement, color patterns, rather than personal bias; Monteith, Voils, & Ashburn-Nardo, 2001). Taken together, these outcomes suggest that White individuals engage in defensive strategies when their egalitarian identity is threatened by IAT feedback. Although we know something about how White individuals react when the IAT might threaten their egalitarian identity, we know little about how racial minorities react. In the following section, we discuss how Black and biracial Black/White individuals may respond to nonegalitarian IAT feedback.

Racial Minorities and Egalitarianism

It is possible that Black/White egalitarianism is not as desirable for Black individuals as it is for White individuals. For instance, certain historical movements (e.g., the Black Pride movement) have suggested that Black people should prefer other Black individuals over others (e.g., White individuals; Sniderman & Piazza, 2002). Moreover, many Black parents believe that they should instill pro-Black attitudes in children (Thomas & Speight, 1999) and interventions aimed at reducing risky behavior among Black adolescents (e.g., smoking, drinking, and unprotected sex) support such pro-Black socialization (Murry et al., 2005). Consistently, Black individuals often endorse ethnocentric attitudes more so than do White individuals (Judd, Park, Ryan, Brauer, & Kraus, 1995; Ryan, Hunt, Weible, Peterson, & Casas, 2007).

Alternatively, because Black individuals are faced with discrimination on a daily basis (Klonoff & Landrine, 2000) and desire greater social equality (Pew, 2013), they may find Black/White egalitarianism particularly desirable. Indeed, some Black parents feel it is important to teach their children to be egalitarian (Bowman & Howard, 1985) and such messages may help buffer youth against the harsh influences of inevitable discrimination (Neblett, Philip, Cogburn, & Sellers, 2006). So the question remains: How do Black individuals respond if they learn they have either pro-White or pro-Black implicit bias?

Although both monoracial Black and White individuals may be threatened by IAT feedback indicating they have pro-White bias—though possibly for different reasons—biracial Black/White individuals further complicate this issue because their racial in-groups are comprised of both Blacks and Whites. Biracial individuals may identify only as a member of their minority (i.e., Black) or their majority (i.e., White) racial group (Rockquemore, Brunsma, & Delgado, 2009) or as multiracial (Kerwin, Ponterotto, Jackson, & Harris, 1993). Moreover although biracial individuals have flexible racial identities (Bonam & Shih, 2009; Gaither, Sommers, & Ambady, 2013;

Renn, 2008), they often experience increased social isolation (Brackett et al., 2006; Gaskins, 1999), face unwanted pressure to “choose sides” (Herman, 2004; Renn, 2008), and often do not receive social recognition of their multiracial status (Rockquemore & Brunsma, 2002), despite desiring such recognition (Remedios & Chasteen, 2013; Renn, 2008). As a result, biracial Black/White individuals who identify as biracial may respond differently to IAT feedback compared to their monoracial counterparts—they may be particularly motivated to prefer White and Black individuals equally because an egalitarian outlook complements their dual heritage.

Consequences of Implicit–Explicit Attitude Discrepancy

Although currently we do know how race may influence responses to IAT feedback, some research has examined how people respond, more broadly, to having discrepant implicit and explicit attitudes. Holding discrepant implicit and explicit attitudes, even without being aware of that discrepancy, can have tangible consequences. For instance, research suggests that to the extent that their implicit attitudes are discrepant from their explicit attitudes, people experience cognitive dissonance (Rydell, McConnell, & Mackie, 2008) which in turn leads them to process new discrepancy-relevant information more thoroughly and elaborately (Briñol, Petty, & Wheeler, 2006; Rydell et al., 2008). When it comes to racial attitudes, such cognitive dissonance may cue people who value egalitarianism to strongly avoid expressing prejudice (Monteith, 1993; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002). Nevertheless, research demonstrates that White participants who value White individuals over Black individuals more implicitly than they do explicitly will engage in behavior that justifies White superiority (e.g., judging arguments more harshly when they come from a Black author than a White author; Shoda, McConnell, & Rydell, 2014). Thus, holding discrepant implicit and explicit attitudes may have a variety of consequences both for information processing and for racial perceptions.

Importantly, research also suggests that educating people about the discrepancy between their implicit and explicit attitudes may help them regulate their biased reactions (Correll, Park, Judd, & Wittenbrink, 2002; Monteith et al., 2002). Therefore, exposing people to their own bias through IAT feedback could be an incredibly useful prejudice-reduction tool (Hillard, Ryan, & Gervais, 2013). Nevertheless, defensive reactions to IAT feedback (Howell et al., 2013) may undermine the usefulness of this tool. As such, it is important to understand when it is that people will respond defensively to IAT feedback. Moreover, because different racial groups may value interracial egalitarianism more or less than others, it is important to know whether these responses may also be moderated by race to ensure the generalizability of these findings.

Overview of the Present Study

In this study, we examine the reactions to IAT feedback of over 1 million White, Black and biracial Black/White participants

who completed the Black–White/Good–Bad IAT. Although a variety of work suggests that White individuals will respond defensively to feedback indicating they are not egalitarian, this study expands prior research in three important ways. First, it builds on research examining how the discrepancy between explicit attitudes and implicit feedback influences defensive responses. Second, it represents the first large-scale investigation of how people react to actual Black–White IAT feedback. In doing so, it informs research on defensiveness and the IAT and contributes to the discourse on using the Black–White IAT as an educational tool to improve interracial interactions (Casad, Flores, & Didway, 2013).

Finally, this study extends research on Black–White attitudes and defensiveness to Black and biracial Black/White individuals. Indeed, despite being among the fastest growing populations in the United States (U.S. Census Bureau, 2012), no study to date has examined both the implicit and explicit racial attitudes of biracial Black/White individuals (see Nosek et al., 2007 for one study investigating implicit and explicit attitudes of multiracial individuals generally—with no direct focus on biracial Black/White individuals nor defensiveness). Therefore, this study informs research both about the reactions of IAT-relevant minorities (i.e., Black and biracial Black/White participants) and allows us to understand how a potentially conflicting social identity (i.e., Black/White) influences defensiveness, informing earlier work on biracial identity conflicts.

Method

Participants

Participants were 1,129,991 volunteers (664,071 women, 461,811 men, and 4,109 unreported) who completed the Black–White/Good–Bad IAT on the Project Implicit website between September 9, 2006, and December 31, 2012 (ages 18–89, mean $[M]_{\text{age}} = 30.1$ years, $SD_{\text{age}} = 12.0$).¹ Visitors typically come to the website via Internet outlets such as “blogs, . . . personal recommendations, search engines, topically relevant sites that provided a link, [or] as a class or work recommendation or assignment,” (Nosek et al., 2007, p. 7). Nevertheless, we have no information concerning the location in which participants completed the task (e.g., at home, in a library, or in a computer lab). Additionally, because this is an archival data set, we have access only to the information available in the public-use data set.

Participants explicitly identified their race by selecting one of nine racial options (White, Black or African American, more than one race—Black/White, more than one race—other, American Indian/Alaska Native, East Asian, South Asian, Native Hawaiian or other Pacific Islander, other or unknown). We restricted analysis to three groups to which the IAT was relevant, “White” ($n = 929,424$), “Black or African American” ($n = 180,654$), and “more than one race—Black/White” ($n = 19,933$). We did not exclude any participants based on ethnicity. Of all participants, 2.2% of Black, 4.0%

of White, and 14.5% of biracial Black/White participants indicated that they were Hispanic or Latino/Latina.² Data were retrieved from <https://osf.io/52qx1/> and are available for public use (Xu, Nosek, & Greenwald, 2014).

Material and Procedure

Overall procedure. After consenting to participation, participants completed, in a counterbalanced order, a demographic questionnaire, the Black–White/Good–Bad IAT, and measures of their explicit attitudes. The Black–White/Good–Bad IAT requires participants to quickly categorize pictures of faces as Black or White and positive (e.g., happy) or negative (e.g., awful) words as good or bad. Next, participants received IAT feedback and subsequently answered a series of questions about the feedback.

Implicit attitudes. The IAT assessed associations between two target categories (Black people and White people) and two evaluative attributes (good and bad). The IAT consisted of seven trial blocks and was scored with the D_1 algorithm (Greenwald, Nosek, & Banaji, 2003). Response latencies <300 ms were removed, and trial latencies were calculated from the beginning of the trial until the time of a correct response, regardless of whether an error was made prior to the correct response. We used this D score in all calculations related to *implicit attitudes*. A positive D score indicated an implicit preference for the White people relative to Black people. Based on this score, participants received feedback indicating that they had a *Slight*, *Moderate*, or *Strong preference for White people compared to Black people* or *for Black people compared to White people* or that they had *no automatic preference*.³ We used this 7-point scale in all of our calculations involving *implicit feedback*. Participants did not receive feedback if they had an overall error rate greater than 30% or if they responded to more than 10% of trials in less than 300 ms.

Explicit attitudes. Participants self-reported their preference for Black versus White individuals on a scale ranging from 1 (*I strongly prefer Black people to White people*) to 7 (*I strongly prefer White people to Black people*). Thus, participants reported their explicit attitudes on a scale that used the same anchor points as did implicit feedback.

Defensiveness. The primary outcome of interest was defensiveness. In line with earlier research (Howell & Ratliff, 2014), we used a 3-item index of defensiveness ($\alpha = .70$) which included the items: “The IAT does not reflect anything about my thoughts or feelings unconscious or otherwise,” “Whether I like my IAT score or not, it captures something important about me” (reverse coded); and “The IAT reflects something about my automatic thoughts and feelings concerning this topic” (reverse coded). Scale options were *Strongly Disagree*, *Disagree*, *Agree*, *Strongly Agree*. We coded

Table 1. Distribution of Direction of Feedback.

Participant Race	Explicit Pro-White Bias > Implicit Pro-White Bias (Percentage)	Explicit Pro-White Bias = Implicit Pro-White Bias (Percentage)	Implicit Pro-White Bias > Explicit Pro-White Bias (Percentage)
White	17.8	19.4	62.8
Black	24.4	19.4	56.2
Biracial Black/White	24.1	20.9	55.0
Average across participants	18.9	19.4	61.7

Table 2. Descriptive Statistics: Explicit Attitudes, Implicit Attitudes, Explicit-Feedback Correlations, IAT and Explicit Correlations, and Discrepancy Scores.

Participant Race	IAT D-Score M (SD)	Explicit M (SD)	Feedback-Explicit Correlations r [CI _{95%}]	IAT D-Score Explicit Correlations r [CI _{95%}]	Feedback-Explicit Discrepancy M (SD)
White	0.41 (0.41)	4.60 (0.94)	.20 [.20, .20]	.16 [.15, .16]	0.90 (1.56)
Black	-0.05 (0.45)	2.94 (1.32)	.16 [.15, .16]	.26 [.24, .27]	0.87 (1.99)
Biracial Black/White	0.16 (0.45)	3.89 (1.07)	.25 [.23, .26]	.21 [.21, .22]	0.70 (1.75)
Average	0.33 (0.45)	4.33 (1.17)	.19 [.19, .20]	.21 [.20, .21]	0.89 (1.64)

Note. IAT = implicit association test.

responses on a scale ranging from 1 (*Strongly Disagree*) to 4 = (*Strongly Agree*).

Calculating discrepancy magnitude and discrepancy direction. To compute implicit feedback–explicit attitude discrepancy, we subtracted explicit attitude from IAT feedback. We then calculated the *magnitude* of this discrepancy by taking the absolute value of the implicit–explicit discrepancy. We square root transformed this value to account for positive skew that results from taking the absolute value of a variable distributed on a zero point (Freeman & Tukey, 1950) and grand-mean centered it for each group. We also coded the *direction* of explicit–implicit discrepancy as -1 for participants who received feedback indicating they implicitly preferred Black to White targets more than they did explicitly, 0 if implicit feedback and explicit attitudes aligned, and $+1$ for participants who received feedback indicating they implicitly preferred White to Black targets more than they did explicitly. Table 1 shows the distribution of the direction variable among White, Black, and biracial Black/White participants.

Analysis and interpretation. We conducted three separate sets of analyses, namely, one for White participants, one for Black participants, and one for biracial Black/White participants. We first examined the descriptive statistics (e.g., M , SD) of each of the variables.⁴ Next, we conducted regressions where we predicted defensiveness (i.e., feedback derogation) from the following: (a) the magnitude of the discrepancy between self-reported attitude and IAT feedback, (b) the direction of the discrepancy between self-reported attitude and IAT feedback, and (c) their interaction. The final model was as follows: $\text{Defensiveness} = b_{00} + b_{01}(\text{Magnitude}^5) + b_{02}(\text{Direction}) + b_{03}(\text{Magnitude}^5 \times \text{Direction})$.

The effects of each of these predictors, in the absence of any other significant predictors, can be interpreted as follows: A positive main effect of magnitude (b_{01}) would indicate that participants are more defensive to the extent that their implicit and explicit attitudes differed, a positive main effect of direction (b_{02}) would indicate that participants are more defensive when they learned they preferred White people more implicitly than they indicated explicitly. A significant interaction (b_{02}) term would indicate that the effect of magnitude on defensiveness (e.g., greater magnitude of discrepancy relating to increased defensiveness) depends on the direction of feedback.

Results

Attitudes

Table 2 shows the average implicit and explicit attitudes, the correlations between explicit attitudes and implicit attitudes and feedback, and the discrepancy between implicit feedback and explicit attitudes for each group as well as across the sample. An examination of this table reveals explicit in-group bias among both White, one sample $t(808,155) = 573.36$, $p < .001$, $d = 0.64$, and Black participants, one sample $t(152,826) = -313.80$, $p < .001$, $d = 0.81$. Biracial Black/White participants demonstrated a very slight, almost negligible, pro-Black bias, one sample $t(17,145) = -13.41$, $p < .001$, $d = 0.09$. Comparing participant groups, White participants showed more pro-White bias than did biracial Black/White participants, $t(17,704.4776) = -85.67$, $p < .001$, $d = 0.71$, who showed more pro-White bias than did Black participants, $t(22,446.7957) = -107.075$, $p < .001$, $d = 0.79$.

When examining implicit attitudes, White participants demonstrated clear in-group bias, one sample $t(855,136) = 915.74$, $p < .001$, $d = 0.99$. Black participants demonstrated

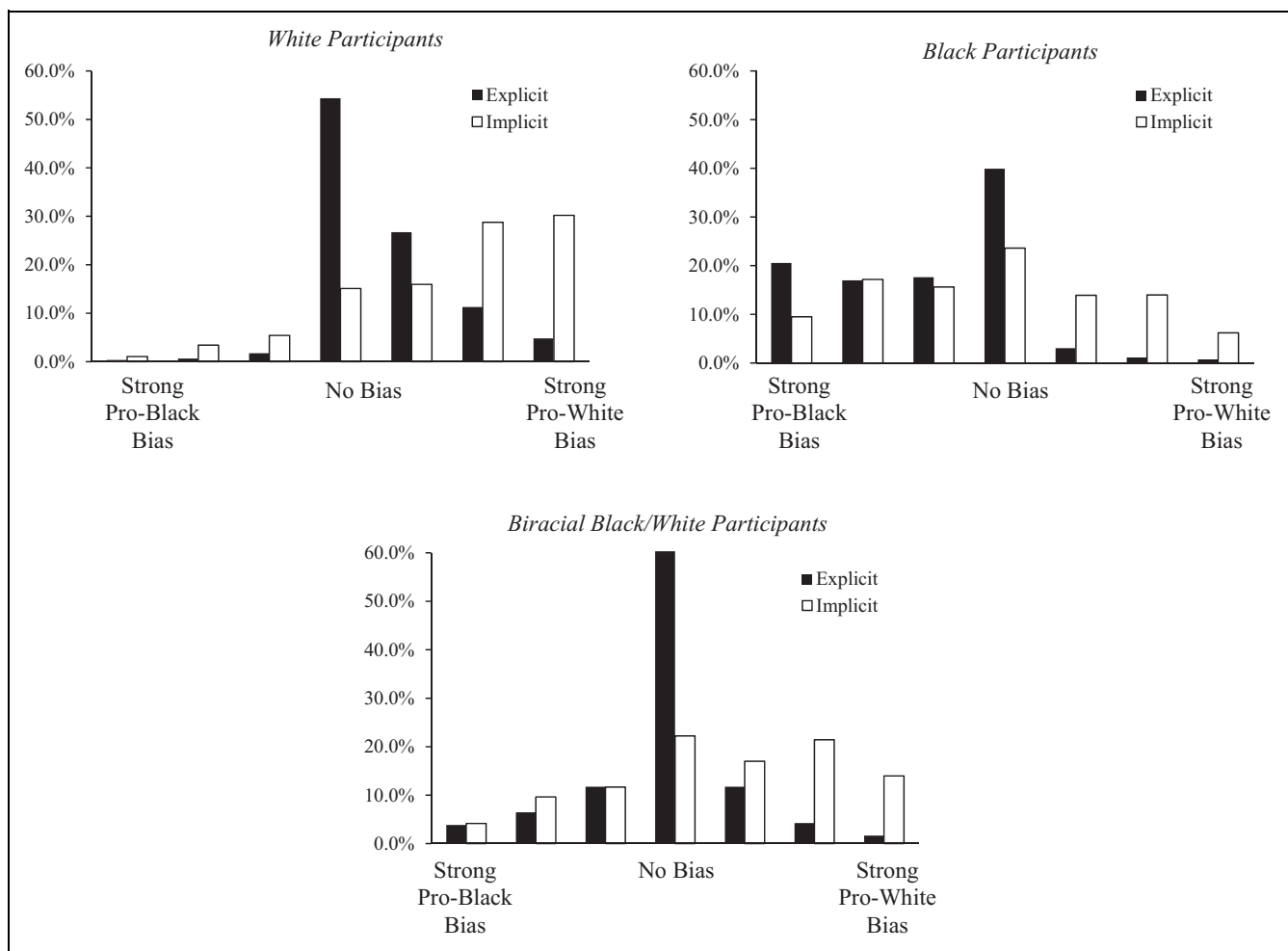


Figure 1. Distribution of explicit attitudes and implicit feedback among each group of participants.

very slight, implicit pro-Black bias, one sample $t(162,122) = -43.74, p < .001, d = 0.11$. Biracial Black/White participants demonstrated moderate implicit pro-White bias, one sample $t(18,140) = 47.44, p < .001, d = 0.35$. Similar to explicit attitudes, White participants showed more implicit pro-White bias than did biracial Black/White participants, $t(18,787.1048) = -74.59, p < .001, d = 0.58$, who showed more implicit pro-White bias than did Black participants, $t(22,446.8439) = -58.91, p < .001, d = 0.46$.

Feedback

All participants disproportionately received feedback indicating they had more implicit pro-White bias than they indicated explicitly, $\chi^2s(2, ns > 16,254) > 3,400, ps < .001$. Moreover, χ^2 analyses of the percentage of participants receiving each type of feedback across races indicate that the percentages of each type of feedback did not differ significantly between races $\chi^2s(2, 200) < 2.90, ps > .24$.

White participants had the greatest discrepancy between their implicit and explicit attitudes and were second highest in correspondence between their explicit attitudes and IAT feedback.

Black participants had the second highest explicit–feedback discrepancy and the lowest in explicit–feedback correspondence. By contrast, biracial Black/White participants had both the lowest discrepancy, and the greatest correspondence between their explicit attitudes and their IAT feedback, $r = .25, 95\% \text{ CI} = [.23, .26]$. Figure 1 shows the distribution of explicit attitudes and implicit feedback among White (Panel 1), Black (Panel 2), and biracial Black/White (Panel 3) participants and shows that the majority of individuals, particularly biracial Black/White participants indicated they were egalitarian explicitly but far fewer received feedback indicating such.

Defensiveness

Overall, Black participants were the most defensive ($M = 2.38$ out of 4, $SD = 0.74$), followed by biracial Black/White participants ($M = 2.33, SD = 0.73$) who were followed by White participants ($M = 2.31, SD = 0.68$). Nevertheless, the difference between monoracial Black and White participants (the highest and lowest groups) was very small, suggesting that all three groups were about equally defensive, $t(167,465.04)^{10} = 29.09, p < .001, d = 0.10$.

The overall model significantly predicted defensive responses among White $F(3, 549,344) = 12,835.49, p < .001, R^2 = .06$, Black, $F(3, 108,909) = 1,287.77, p < .001, R^2 = .03$, biracial Black/White participants, $F(3, 11,529) = 212.33, p < .001, R^2 = .0511$. A small, but significant interaction between magnitude and direction emerged among White, $b = .12$, standard error (SE) = .005, $t = 26.84, r_{\text{partial}} = .04, p < .001$, and Black participants, $b = .05, SE = .009, t = 5.31, r_{\text{partial}} = .02, p < .001$, qualifying both main effects. As such, we examined the simple main effects of magnitude for both directions of feedback.

For White participants, increased implicit–explicit discrepancy was related to increased defensiveness more when their IAT feedback indicated that they were more pro-White than they indicated explicitly, $b = .54, SE = .004, t = 122.44, r_{\text{partial}} = .16, p < .001$, than when their IAT feedback indicated that they were more pro-Black than they indicated explicitly, $b = .29, SE = .006, t = 44.27, r_{\text{partial}} = .06, p < .001$.

A similar pattern emerged for Black participants such that increasing implicit–explicit discrepancy was related to increased defensiveness more when their IAT feedback indicated that they were more pro-White than they indicated explicitly, $b = .37, SE = .008, t = 44.60, r_{\text{partial}} = .13, p < .001$, than when their IAT feedback indicated that they were more pro-Black than they indicated explicitly, $b = .27, SE = .014, t = 19.25, r_{\text{partial}} = .06, p < .001$. In sum, among monoracial White and Black participants, increased discrepancy between implicit and explicit attitudes prompted increased defensiveness most when the feedback indicated that they were more pro-White implicitly than they indicated explicitly.

The interaction between magnitude and direction was not significant among biracial Black/White participants, $b = .00, SE = .03, t = -.02, r_{\text{partial}} = .00, p = .99$. However, the main effects of both magnitude, $b = .40, SE = .03, t = 244.86, r_{\text{partial}} = .13, p < .001$, and direction, $b = .06, SE = .009, t = 6.95, r_{\text{partial}} = .06, p < .001$, emerged. Specifically, biracial Black/White participants responded more defensively when their feedback indicated that they were more pro-White than they indicated explicitly. Nevertheless, defensiveness increased as discrepancy increased in either direction.

Discussion

The results of this large-scale archival investigation of explicit attitudes, implicit feedback, and defensiveness revealed two major findings. First, although the majority of White, Black, and biracial Black/White participants explicitly endorsed egalitarian attitudes, race clearly influenced both attitudes and defensiveness. Indeed, both White and Black participants showed explicit and implicit bias (Axt et al., 2014; Judd et al., 1995). By contrast, biracial Black/White participants showed almost no explicit bias but moderate pro-White implicit bias. Furthermore, although biracial Black/White participants' implicit and explicit attitudes were more consistent

and less discrepant than the other two groups, their defensiveness scores still fell between Black (most defensive) and White (least defensive) participants' defensiveness scores.

We consistently found a main effect of magnitude of discrepancy, suggesting that that implicit feedback that was discrepant from their explicit attitudes prompted defensiveness in all three participant groups. Among both Black and White participants, an interaction between magnitude and direction of implicit–explicit discrepancy suggested that the effect of magnitude of discrepancy was stronger when participants received feedback indicating that their implicit attitudes were more pro-White than their explicit attitudes. This is consistent with the notion that White individuals want to avoid appearing racist (O'Brien et al., 2010) and that Black individuals value pro-Black bias (Sniderman & Piazza, 2002). Among biracial Black/White participants, no such interaction emerged, suggesting that biracial individuals may value Black/White egalitarianism more than their monoracial counterparts.

One possibility for why biracial Black/White participants reacted defensively to large discrepancies in either direction of implicit–explicit discrepancy (i.e., a main effect of magnitude not qualified by an interaction with direction) is that they may identify with both their White and Black in-groups (Renn, 2008; Rockquemore & Brunnsma, 2002). Therefore, the in-group bias explicitly endorsed by their monoracial counterparts (Axt et al., in press) is instead expressed as egalitarianism. In line with this reasoning, biracial Black/White individuals were the least likely to explicitly endorse prejudice in either direction. Thus, when biracial Black/White participants received feedback indicating they were either pro-White or pro-Black, such feedback likely challenged their other racial identity (Rollins & Hunter, 2013) and their explicitly egalitarian views, causing them to respond defensively to greater implicit–explicit discrepancies in either direction. Future work should examine these outcomes with other types of biracial individuals (i.e., Asian/White) and whether biracial individuals endorse egalitarianism overall more than other racial groups or if this egalitarian view and response to IAT feedback is specific to one's racial in-groups.

Theoretical Implications

These results contribute to a growing body of work examining defensive reactions to feedback about racial attitudes. The results suggest that there is not a universal formula for understanding defensive reactions to IAT feedback. Indeed, relevant demographics may influence whether people are motivated to receive feedback that paints them, or their in-group, in a positive light (Dunning, 2007) or that aligns with their expectations (Swann, 1990).

To our knowledge, this study represents the first large-scale examination and comparison of the implicit and explicit racial attitudes of biracial Black/White individuals. These data add to a limited understanding regarding conflicting racial identities for the biracial population (Renn, 2008; Rockquemore &

Brunsmas, 2002) and extend prior work by demonstrating that having multiple racial identities can influence reactions to different types of social feedback for the biracial population.

Finally, this research contributes to existing knowledge about implicit–explicit attitude discrepancy. As stated previously, research suggests that holding discrepant implicit and explicit attitudes can affect cognitive dissonance (Rydell et al., 2008) and information processing (Briñol et al., 2006), and it may help regulate prejudiced thoughts and behavior (Monteith et al., 2002). Nevertheless, research also suggests that such a discrepancy can also increase prejudiced perspectives (Shoda et al., 2014). Although we do not have data for these other outcomes, our results demonstrate that making people aware of the discrepancy between their implicit and explicit attitudes may not be a panacea for prejudice, as they may react defensively, derogating the message. However, this is a question worthy of future investigation.

Practical Applications

Research suggests that defensive strategies can help people maintain positive self-views, which, in turn, can promote their overall well-being (Taylor & Brown, 1988). Still, other research suggests that such defensiveness may promote unhealthy behavior or inadvertently hinder self-improvement (McQueen, Vernon, & Swank, 2013; Taylor, Neter, & Wayment, 1995). This study demonstrates an additional potential pitfall of using defensiveness to maintain a positive, but potentially inaccurate, illusion: it may undermine the benefits of IAT feedback. Indeed, although the Project Implicit website may be useful in educating people about their personal implicit bias (Casad et al., 2013; Hillard et al., 2013), those individuals who potentially benefit most from learning about implicit bias—those whose implicit attitudes differ greatly from their explicit attitudes—were also the most likely to be defensive. As such, those interested in educating people about their implicit attitudes should consider defensiveness-reducing strategies such as affirmation (Cohen & Sherman, 2014).

Limitations and Future Directions

The archival data from Project Implicit was limited in various ways. First, participants chose to complete the study knowing they would learn their implicit attitudes. Although this contributes to the study's ecological validity, it is possible that participants who did not choose to complete this IAT may somehow be different. Second, we were unable to assess the extent to which people identified with their race. Because our sample reflects biracial individuals who self-identified as biracial Black/White, we do not know whether these results may differ for biracial individuals who self-identify in other ways (e.g., as monoracially Black or White). As such, future studies should include more sensitive measures of racial identification.

Conclusion

The results of this study suggest that people are defensive in response to feedback indicating that their implicit attitudes differ from their explicit attitudes. Among monoracial White and Black individuals, this effect was particularly strong when they learned that they were implicitly more pro-White than they indicated explicitly. By contrast, biracial Black/White individuals were defensive about large discrepancies in either direction (e.g., more pro-Black or more pro-White implicit attitudes). These results pinpoint one distinct difference between monoracial and biracial populations and pave the way for future research to further explore how different populations compare in attitudes and reactions to self-relevant feedback.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This article is based on work supported by a National Science Foundation Graduate Research Fellowship (under Grant No. DGE-0802270) awarded to Jennifer Howell, by a University of Chicago Provost's Postdoctoral Scholarship awarded to Sarah Gaither, and by a grant from Project Implicit awarded to Kate Ratliff. Kate Ratliff is a consultant with Project Implicit, Inc., a non-profit organization that includes in its mission "to develop and deliver methods for investigating and applying phenomena of implicit social cognition, including especially phenomena of implicit bias based on age, race, gender or other factors."

Notes

1. All measures appeared beginning September 9, 2006.
2. Including Hispanic/Latino/Latina participants does not change the pattern of results presented here.
3. The cutoffs for each type of feedback were as follows: $d \leq -0.65$ = strong pro-Black preference; $d = -0.64$ to -0.35 = moderate pro-Black preference; $d = -0.34$ to -0.15 = slight pro-Black preference; $d = -0.14$ to 0.15 = no preference; $d = 0.16$ to 0.34 = slight pro-White preference; 0.35 to 0.64 = moderate pro-White preference; $d \geq 0.65$ = strong pro-White preference.
4. Because the cell sizes were uneven, we imposed the Welch–Satterthwaite degree of freedom penalty to correct for heteroscedasticity in all between-group comparisons (Satterthwaite, 1946; Welch, 1947).
5. Group mean centered.
6. Degree of freedom not whole because equal variances were not assumed, $F = 452.63$, $p < .001$.
7. Degree of freedom not whole because equal variances were not assumed, $F = 6,054.49$, $p < .001$.
8. Degree of freedom not whole because equal variances were not assumed, $F = 314.42$, $p < .001$.
9. Degree of freedom not whole because equal variances were not assumed, $F = 3.81$, $p = .05$.

10. Degree of freedom not whole because equal variances were not assumed, $F = 1,641.33$, $p < .001$.
11. Adding gender, political orientation, and Hispanic/Latino/Latina status to the model improved overall prediction among Black, $\Delta R^2 = .01$, biracial Black/White, $\Delta R^2 = .02$, and White, $\Delta R^2 = .01$, participants but did not change the pattern of results for any of the predictors. As such, we do not discuss these results here.

References

- Axt, J. R., Ebersole, C. R., & Nosek, B. A. (2014). The rules of implicit evaluation by race, religion, and age. *Psychological Science*, Advance online publication. doi:10.1177/0956797614543801.
- Bonam, C. M., & Shih, M. (2009). Exploring multiracial individuals' comfort with intimate interracial relationships. *Journal of Social Issues*, *65*, 87–103.
- Bowman, P. J., & Howard, C. (1985). Race-related socialization, motivation, and academic achievement: A study of black youths in three-generation families. *Journal of the American Academy of Child Psychiatry*, *24*, 134–141.
- Brackett, K. P., Marcus, A., McKenzie, N. J., Mullins, L. C., Tang, Z., & Allen, A. M. (2006). The effects of multiracial identification on students' perceptions of racism. *The Social Science Journal*, *43*, 437–444.
- 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.
- Casad, B. J., Flores, A. J., & Didway, J. D. (2013). Using the implicit association test as an unconsciousness raising tool in psychology. *Teaching of Psychology*, *40*, 118–123.
- Cohen, G., & Sherman, D. K. (2014). The psychology of change: Self-affirmation theory and values affirmation interventions. *Annual Review of Psychology*, *65*, 333–371.
- 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. doi:10.1037/0022-3514.83.6.1314
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*, 359–378. doi:10.1037/0022-3514.82.3.359
- Dunning, D. (2007). Self-image motives: Further thoughts and reflections. *Journal of Consumer Psychology*, *17*, 258–260. doi:10.1016/s1057-7408(07)70036-0
- Frantz, C. M., Cuddy, A. J., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30*, 1611–1624.
- Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, *21*, 607–611.
- Gaither, S. E., Sommers, S. R., & Ambady, N. (2013). When the half affects the whole: Priming identity for biracial individuals in social interactions. *Journal of Experimental Social Psychology*, *49*, 368–371.
- Gaskins, P. F. (1999). *What are you?: Voices of mixed-race young people*. New York: Macmillan.
- 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. doi: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*, 197–216.
- Herman, M. (2004). Forced to choose: Some determinants of racial identification in multiracial adolescents. *Child Development*, *75*, 730–748.
- Hillard, A. L., Ryan, C. S., & Gervais, S. J. (2013). Reactions to the implicit association test as an educational tool: A mixed methods study. *Social Psychology of Education*, *16*, 495–516.
- Howell, J. L., Collisson, B., Crysel, L., Garrido, C. O., Newell, S. M., Cottrell, C. A., . . . Shepperd, J. A. (2013). Managing the threat of impending implicit attitude feedback. *Social Psychological and Personality Science*, *4*, 714–720. doi:10.1177/1948550613479803
- Howell, J. L., & Ratliff, K. A. (2014). *Implicit-explicit attitude discrepancy prompts defensive responding to IAT feedback* (Unpublished manuscript). University of Florida, Gainesville, FL.
- Judd, C. M., Park, B., Ryan, C. S., Brauer, M., & Kraus, S. (1995). Stereotypes and ethnocentrism: Diverging interethnic perceptions of African American and White American youth. *Journal of Personality and Social Psychology*, *69*, 460–481.
- Kerwin, C., Ponterotto, J. G., Jackson, B. L., & Harris, A. (1993). Racial identity in biracial children: A qualitative investigation. *Journal of Counseling Psychology*, *40*, 221–231.
- Klonoff, E. A., & Landrine, H. (2000). Is skin color a marker for racial discrimination? Explaining the skin color, Åihypertension relationship. *Journal of Behavioral Medicine*, *23*, 329–338.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). New York: Academic Press.
- McQueen, A., Vernon, S. W., & Swank, P. R. (2013). Construct definition and scale development for defensive information processing: An application to colorectal cancer screening. [Research Support, Non-U.S. Gov't]. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *32*, 190–202. doi:10.1037/a0027311
- Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*, 469.
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*, 1029.
- Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition*, *19*, 395–417. doi:10.1521/soco.19.4.395.20759
- Murry, V. M., Brody, G. H., McNair, L. D., Luo, Z., Gibbons, F. X., Gerrard, M., & Wills, T. A. (2005). Parental involvement

- promotes rural African American youths' self-pride and sexual self-concepts. *Journal of Marriage and Family*, 67, 627–642.
- Neblett, E. W., Philip, C. L., Cogburn, C. D., & Sellers, R. M. (2006). African American adolescents' discrimination experiences and academic achievement: Racial socialization as a cultural compensatory and protective factor. *Journal of Black Psychology*, 32, 199–218.
- Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Greenwald, A. G. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18, 36–88.
- O'Brien, L. T., Crandall, C. S., Horstman-Reser, A., Warner, R., Alsbrooks, A., & Blodorn, A. (2010). But I'm no bigot: How prejudiced White Americans maintain unprejudiced self-images. *Journal of Applied Social Psychology*, 40, 917–946. doi:10.1111/j.1559-1816.2010.00604.x
- Pew. (2013). King's dream remains an Elusive goal; Many Americans see racial disparities. Retrieved from <http://www.pewsocialtrends.org/2013/08/22/kings-dream-remains-an-elusive-goal-many-americans-see-racial-disparities/>.
- Remedios, J. D., & Chasteen, A. L. (2013). Finally, someone who "gets" me! Multiracial people value others' accuracy about their race. *Cultural Diversity and Ethnic Minority Psychology*, 19, 453–460.
- Renn, K. A. (2008). Research on biracial and multiracial identity development: Overview and synthesis. *New Directions for Student Services*, 2008, 13–21.
- Rockquemore, K. A., & Brunnsma, D. L. (2002). Socially embedded identities: Theories, typologies, and processes of racial identity among Black/White biracials. *The Sociological Quarterly*, 43, 335–356.
- Rockquemore, K. A., Brunnsma, D. L., & Delgado, D. J. (2009). Racing to theory or retheorizing race? Understanding the struggle to build a multiracial identity theory. *Journal of Social Issues*, 65, 13–34.
- Rollins, A., & Hunter, A. G. (2013). Racial socialization of biracial youth: Maternal messages and approaches to address discrimination. *Family Relations*, 62, 140–153.
- Rudman, L. A., Dohn, M. C., & Fairchild, K. (2007). Implicit self-esteem compensation: automatic threat defense. *Journal of Personality and Social Psychology*, 93, 798–813.
- Ryan, C. S., Hunt, J. S., Weible, J. A., Peterson, C. R., & Casas, J. F. (2007). Multicultural and colorblind ideology, stereotypes, and ethnocentrism among Black and White Americans. *Group Processes & Intergroup Relations*, 10, 617–637.
- 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.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114.
- Shoda, T. M., McConnell, A. R., & Rydell, R. J. (2014). Having explicit-implicit evaluation discrepancies triggers race-based motivated reasoning. *Social Cognition*, 32, 190–202.
- Sniderman, P. M., & Piazza, T. L. (2002). *Black pride and black prejudice*. Princeton, NJ: Princeton University Press.
- Swann, W. B., Jr. (1990). To be adored or to be known? The interplay of self-enhancement and self-verification. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 408–448). New York, NY: Guilford Press.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210. doi:10.1037/0033-2909.103.2.193
- Taylor, S. E., Neter, E., & Wayment, H. A. (1995). Self-evaluation processes. *Personality and Social Psychology Bulletin*, 21, 1278–1287.
- Thomas, A. J., & Speight, S. L. (1999). Racial identity and racial socialization attitudes of African American parents. *Journal of Black Psychology*, 25, 152–170.
- U.S. Census Bureau. (2012). The two or more races population: 2010. *2010 Census Brief*. Retrieved May 1, 2014, from <http://www.census.gov/prod/cen2010/briefs/c2010br-13.pdf>
- Welch, B. L. (1947). The generalization of student's problem when several different population variances are involved. *Biometrika*, 34, 28–35.
- Xu, K., Nosek, B. A., & Greenwald, A. G. (2014). Data from the race implicit association test on the Project Implicit demonstration website. *Journal of Open Psychology Data*, 2, e3. doi:10.5334/jopd.ac

Author Biographies

Jennifer L. Howell is a doctoral candidate at the University of Florida in the Department of Psychology. She earned her BA in psychology from the Southwestern University in Georgetown, TX, and her MS from the University of Florida in Gainesville, FL. Her research focuses broadly on psychological-threat management.

Sarah E. Gaither earned her BA in social welfare from UC Berkeley and her MS and PhD in social psychology from the Tufts University. She is currently a Provost's Postdoctoral Scholar in the Psychology Department and Fellow at the Center for the Study of Race, Politics, and Culture at the University of Chicago. Her research focuses on interracial interactions, racial categorizations, and biracial identity flexibility and experiences.

Kate A. Ratliff is an assistant professor in the Department of Psychology at the University of Florida and director of Research and Education at Project Implicit. Much of her research focuses on implicit prejudice and stereotyping; she is also working with her graduate students to understand the role of implicit attitudes in a wide variety of domains such as romantic relationships, group identification, health decision making, environmental concerns, and moral reasoning.