Exploring Racial Bias in a European Country with a Recent History of Immigration of Black Africans

Patricia M Power[1] , Colin Harte[2], Dermot Barnes-Holmes[2], and Yvonne Barnes-Holmes[2]

[1]Department of Psychology, National University of Ireland Maynooth, Ireland

[2]Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium

Corresponding Author:

Colin Harte

Department of Experimental, Clinical and Health Psychology

Ghent University

Henri Dunantlaan, 2

9000 Ghent

Belgium

Email: Colin.Harte@UGent.be

Abstract

The current study examined levels of racial bias among black and white individuals residing in Ireland using the Implicit Relational Assessment Procedure (IRAP) and a range of questionnaire measures. The IRAP required participants to respond quickly and accurately on a computer-based task. On some blocks of trials participants were required to respond in a pro-white and anti-black manner, whereas on other blocks responding in the opposite direction was required (anti-white/pro-black). The difference in response latencies between these two types of trials provided an index of racial bias. Performance on the IRAP (i) revealed in-group/out-group bias for the white but not the black participants; (ii) substantively increased the predictive validity of a range of questionnaire-based measures; and (iii) provided the best prediction of racial group. The results support the utility of the IRAP as a measure of racial bias, and indicate that this bias differed between black and white Irish residents.

Keywords: Racial bias, IRAP, Relational Frame Theory, known-groups

The study of derived stimulus relations has been used widely in behavior analysis as a way of analyzing human language and cognition and relatively early, the paradigm was used to study social categorization and prejudice. The first study in this regard examined social categorization in Northern Ireland, where family names and sectarian symbols are often associated exclusively with either Catholic *or* Protestant communities (Watt, Keenan, Barnes, & Cairns, 1991). The study involved training participants in a series of matching-to-sample tasks that were designed to generate derived equivalence relations between Catholic names and Protestant symbols, that would be inconsistent with the verbal/social histories of participants who resided in Northern Ireland. The results showed that some Northern Irish residents did indeed demonstrate difficulty in forming these equivalence relations, whereas individuals from outside Northern Ireland did not. Numerous studies since have reported broadly similar outcomes in which participants with specific pre-experimental histories appear to show difficulty forming derived relations that are inconsistent with those histories (e.g., Barnes, Lawlor, Smeets, & Roche, 1996; Dixon, Rehfeldt, Zlomke, & Robinson, 2006; Leslie, et al., 1993; Merwin & Wilson, 2005).

The general strategy of comparing patterns of responding that are consistent versus inconsistent with the pre-experimental histories of participants carried through to more recent efforts to develop behavior-analytic procedures that may be used to assess verbal relations occurring in the natural environment (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008). The currently most widely used method in this regard is the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010), which was based explicitly on Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), an account that aims to bring together the study of derived stimulus relations and human language and cognition (for a detailed treatment of the theoretical development of the IRAP see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010).

The IRAP presents pairs of stimuli (e.g., words, pictures, statements) on each trial and participants are required to confirm or disconfirm the relation between the pairs. Corrective feedback is presented after each response. In general, the feedback is designed to be consistent on some blocks with participants' verbal histories, and on other blocks to be inconsistent. For example, an IRAP might require responding "True" to a picture of a flower and the word "pleasant" (history consistent) on one block, and "False" (history inconsistent) on another block. The basic logic of the IRAP is that, all things being equal, participants should show a tendency to respond more quickly on history-consistent, relative to history-inconsistent, blocks. This difference in latencies across the two types of blocks is often referred to as the IRAP effect or a positive or negative response bias, depending on whether the effect is above or below zero. It is important to understand that the term IRAP effect, or the concept of response bias, should not be interpreted as a proxy for a mental construct or implicit attitude in a cognitive or social psychological sense. Instead, these terms simply denote a tendency to respond in one particular direction over another on the IRAP. There are now over 50 published studies on the IRAP and the number of domains of interest has increased steadily, with a recent meta-analysis in the clinical domain yielding a relatively high level of predictive validity (Vahey, Nicholson, & Barnes-Holmes, 2015).

One of the earliest IRAP studies examined the response patterns of white participants toward pictures of black and white individuals (Barnes-Holmes, Murphy et al., 2010). Specifically, participants were presented with one of two label stimuli ("Safe" and "Dangerous") on each trial with a picture of a white or a black man holding a gun as a target stimulus. The IRAP required responding in a pro-white and anti-black pattern on some blocks of trials (e.g., pressing a key for “True” rather than “False” when “Safe” appeared with a picture of a white man). On other blocks of trials, responding in a pro-black and anti-white pattern was required (e.g., pressing a key for “True” rather than “False” when “Safe” appeared with a picture of a black man). The IRAP revealed pro-white and anti-black biases, although the anti-black effect was restricted to one trial-type. That is, participants responded “True” more quickly than “False” when presented with “Dangerous” and pictures of black men holding guns; when the pictures were of white men holding guns, participants responded “False” more quickly than “True”.

In discussing the results of this study, the authors noted that the IRAP effects may have been influenced by historical factors above and beyond those associated with so-called racially biased tendencies; “it is possible…. that a bias toward responding “True” over “False”, per se, interacted with the socially loaded stimulus relations presented in the IRAP” (Barnes-Holmes, Murphy et al., 2010, p.62). In this context, the absolute value of an IRAP effect for a particular trial-type should be interpreted with caution. Comparing IRAP effects between groups that are known to differ in some important or relevant way may be less prone to misinterpretation. This is due to the fact that any between-group difference that emerges should be the result, at least in part, of that difference, rather than some generic tendency to respond “True” more quickly than “False”. The same logic would apply to any other procedural variables that may influence IRAP performances, such as the particular instructions or response options employed in a given IRAP (e.g., Finn, Barnes-Holmes, Hussey, & Graddy, 2016; Maloney & Barnes-Holmes, 2016).

With the above in mind, it is interesting that only two studies have been published that have attempted to compare the responses of white with black individuals on an IRAP. The first of these, conducted by Drake et al. (2010) compared a sample of black and white participants with an IRAP that presented the label stimuli “white” and “black” with evaluative target stimuli. Results showed that both black and white participants responded with a pro-white and anti-black bias, however only the white-positive trial-type was statistically significant from 0. The second of these studies (Drake et al., 2015) presented black and white participants with two race IRAPs in a row. These IRAPs were comprised of the same stimuli as was used in the previous 2010 study and the results showed positive in-group biases but not necessarily negative out-group biases. Both of these studies involved relatively small samples of black participants (*N* = 4, *N* = 10, respectively), who were university undergraduates in the US. The data from both studies suggested that the patterns of responding on the IRAP by the black participants differed from those of the white participants, in a manner broadly consistent with known-group differences. Although the limitations of these studies (especially the small sample sizes and the reliance upon undergraduate samples) render it difficult to draw firm conclusions, there is now some preliminary evidence of racial known-groups differences using the IRAP.

The purpose of the present study was to conduct a known-groups analysis of racial bias using the IRAP in an Irish context with black and white participants. Specifically, the study was conducted in 2009 when Ireland was experiencing an economic recession and levels of immigration were falling across all groups. By this stage, therefore, even many recent immigrants had left Ireland because of the sharp downturn in employment opportunities. Critically, participants were not currently university undergraduates in either group, but were members of the general population. At the time of writing, this was the first study to be conducted with black Irish residents, using the IRAP or any measure of so-called implicit attitudes, and thus a precise prediction was difficult. Indeed, Ireland has a very short history of significant black immigration with past censuses showing, for example, that the number of black African nationals living in Ireland increased almost ten-fold from 4,867 in 1996 to 42,764 in 2006 (http://www.cso.ie/census/default.htm). As such, Ireland, especially in 2009, presented an unusual social and cultural context, relative to countries in which black minorities have resided for decades if not centuries. Furthermore, many black residents in Ireland came seeking asylum from various forms of persecution in their indigenous countries, and thus may not be directly comparable to previous samples of black participants employed in non-Irish studies of racial bias. Given this rather unusual historical context, there are insufficient grounds on which to make specific predictions concerning the differences that may emerge between white and black people on the IRAP or other such measures. In this sense, therefore, the current study is largely exploratory in nature.

One criticism of the original Barnes-Holmes, Murphy et al. (2010) study could be that the IRAP targeted only one specific dimension of racial bias, specifically safe versus dangerous. Given the common portrayal of black males in the North American and British media (the latter is widely available in Ireland) as violent gun-carrying gang members, it could be argued that the resulting IRAP effects were hardly surprising. In the current study, therefore, participants were asked to respond to the labels “I think Black People are” and “I think White People are”, and a range of negative-versus-positive attributes (e.g., “Stupid” versus “Clever”). If the anti-black and pro-white effects reported by Barnes-Holmes et al. are replicated, this would indicate that the IRAP could provide a more general measure of racial bias, rather than one that is restricted to a particular dimension.

Another important feature of the current study is that it sought to test the predictive validity of the IRAP using a known-groups approach. That is, logistical regression analyses are employed to determine if the IRAP data accounts for additional variance beyond that provided by self-report measures of racial bias. In addition, discriminant analyses are used to determine if the IRAP and the self-report measures independently predict the race of the participants.

Method

Participants

Twenty-two black participants aged 17 to 26 years (*M* = 22), attending adult education classes in an inner-city Dublin school, completed the experiment individually. All participants were born in Nigeria but had been resident in Ireland for at least 5 years. The data from 6 of these participants were excluded because they failed to achieve or maintain the performance criteria on the IRAP (explained below), leaving *N* = 16 for analysis. Eighteen white participants who had been born and lived in Ireland for most of their lives successfully completed the study; they all resided in the Dublin area. They were aged 18 to 28 years (*M* = 23) and all completed the experiment individually. An exact record of the total number of white participants who were recruited for the study was not available in 2017 (8 years after the data were collected), but no more than 5 were excluded because they failed to achieve or maintain the performance criteria on the IRAP (explained below). The current data were collected in 2009 before we had access to the findings of Drake et al. (2010, 2015). No formal power analysis were conducted for the current study, but a recent meta-analysis of the IRAP in the clinical domain (Vahey et al., 2015) indicates that the current sample size (*N* = 34) is roughly in the region required to achieve the benchmark statistical power of .80 (see Cohen, 1988) for simple between-group comparisons and first-order correlational analyses.

Materials and Apparatus

Discrimination and Diversity Scales. The Discrimination Scale (DS) and the Diversity Scale (DV) were both created by Wittenbrink, Judd, and Park (1997). The DS consisted of 10 statements concerning beliefs about discrimination within Irish society (e.g., *These days, reverse discrimination against Whites is as much a problem as discrimination against Blacks itself*) and has reported an alpha reliability coefficient of .885. The DV comprised 4 beliefs about the value of ethnic diversity within society (e.g., *There is a real danger that too much emphasis on cultural diversity will tear Ireland apart*) and has reported an alpha reliability coefficient of .672. All items required participants to indicate agreement or disagreement with the statements on a 5- point scale from 1 = strongly agree and 5 = strongly disagree. Lower scores of 1-2 indicate pro-white/anti-black racial discrimination, while 4-5 indicate anti-white/pro-black racial discrimination, and 3 indicates no discrimination.

Semantic Differential Scales (SDSs). The study involved 6 7-point SDSs. Each scale ranged from -3 to +3 and had an oppositional adjective at each end (e.g., one scale was anchored at -3 with friendly and +3 with hostile) and participants selected one number along this line. The six oppositional adjective pairs were identical to those presented in the IRAP. Participants were told that the scales were used to assess their attitudes to two specific groups of people, black people and white people. The instructions explicitly encouraged them to record their immediate reaction to each group, rather than trying to figure out a "right answer". Participants were assured that all of their responses were anonymous and confidential. The first scale extended from *friendly at -3 to hostile at +3;* the second extended from *honest at -3 to deceitful at +3;* the third from *lazy at -3 to hardworking at +3;* the fourth from *peaceful at -3 to violent at +3;* the fifth from *bad at -3 to good at +3;* and the sixth from *stupid at -3 to clever at +3*. Each scale was presented twice, one of which assessed attitudes to Black People (presenting the words *Black People* with each scale), the other assessed attitudes to White People (presenting the words *White People* with each scale). In order to generate a measure of racial stereotyping for black versus white participants toward black versus white people, the individual participant ratings of the 6 scales that referred to white people were summed, as were the individual ratings of the 6 scales that referred to black people. An individual average score for white versus black people was then calculated, followed by a calculation of the mean scores for white versus black for each of the two groups of participants. For white participants, the mean score for white people was subtracted from the mean score for black people, in order to provide a measure of racial stereotyping. Similarly, for black participants, the mean score for white people was subtracted from the mean score for black people. Thus, a positive score indicated pro-black stereotyping and a negative score indicated pro-white stereotyping.

Feeling Thermometers. Two identical Feeling Thermometers, presented pictorially as visual analog scales, assessed favorability toward white and black people, from 0º (cold or unfavorable) to 99º (warm or favorable), with 10º intervals. Participants were asked to rate how they felt about white people on one thermometer and how they felt about black people on the other thermometer. In response, they marked a position on one of the intervals along each of the two pictorial thermometers.

Implicit Relational Assessment Procedure (IRAP). All participants completed the IRAP on a standard personal computer. The IRAP software (2008 version programmed in

Visual Basic 6) presented the stimuli and recorded participant responses. Each trial presented the label statement; “I think BLACK people are” or “I think WHITE people are”. One of 12 target stimuli was also presented, 6 stereotypically positive words (“Friendly”, “Honest”, “Hardworking”, “Peaceful”, “Good”, “Clever”) and 6 negative (“Hostile”, “Deceitful”, “Lazy”, “Violent”, “Bad”, “Stupid”). Each trial presented the two response options, “True” and “False”. Based on the various sample-target combinations, the IRAP comprised 4 trial-types; *White People-Positive, Black People-Negative, Black People-Positive,* and *White People-Negative* (see Figure 1).

INSERT FIGURE 1 HERE

Procedure

IRAP. The IRAP program began with a set of instructions, which described the task by illustrating the layout of the screen and explaining the response options. The instructions informed participants that on each trial one of two statements, “I think BLACK people are” or “I think WHITE people are”, would appear at the top of the screen along with a target word in the center of screen. Participants were also told that the response options “True” and “False” would appear at the bottom of the screen, and they were required to choose one of these options on each trial; they were told that the left-right positions of these response options would switch randomly from trial-to-trial. The instructions also informed participants that correct responses would allow them to progress to the next trial, but incorrect responses would produce a red ‘X’ in the middle of the screen, which could only be removed by pressing the correct key. In addition, participants were informed that if they took longer than 2000 ms on any IRAP trial, the phrase “Too Slow!” would be presented on the screen. It is important to note that no specified rules for responding were provided at any point, hence, participants learned the accurate pattern of responding on each block via the feedback contingencies. This is the primary purpose of the practice blocks, although corrective feedback for incorrect responding is retained even during the test blocks.

The IRAP task consisted of a minimum of two practice blocks and a fixed set of six test blocks. Each block presented 24 trials as four different trial-types (see Figure 1). The first block of the IRAP was consistent with pro-white/anti-black stereotyping (e.g., I think WHITE people are–Positive–True; I think BLACK people are–Positive–False; I think WHITE people are-Negative–False; I think BLACK people are–Negative–True). The feedback contingencies alternated from block to block. Thus, in the second block of the IRAP, correct responses were consistent with anti-white/pro-black stereotyping (e.g., I think WHITE people are–Positive–False; I think BLACK people are–Positive–True; I think WHITE people are–Negative–True; I think BLACK people are–Negative–False). Before each new block, participants were informed that the previously correct and wrong answers would be reversed. The order in which the IRAP blocks (i.e., consistent versus inconsistent) were presented was not counterbalanced, because previous research conducted at around the same time as the current data were collected had indicated that this variable did not significantly influence IRAP effects (e.g., McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009; Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009).

For the first two practice blocks, participants were informed that it was a practice phase and errors were expected. Participants were required to reach a standard of >/=80% correct responses, and a median response time of </=2000ms. Participants were allowed three attempts (a total of six practice blocks) to achieve the practice criteria, and if they failed to do

so, they were thanked, debriefed, and their data were discarded (six participants were removed from the study on this basis). Participants who did achieve the practice criteria proceeded to the six test blocks. No performance criteria were applied during the test blocks in order to proceed, but if a participant's performance fell below 80% accuracy for any test block the data for that participant were discarded (one participant was removed from the study on this basis). When all six test blocks had been completed, participants reported to the researcher.

Self-report measures. After the IRAP, participants completed the 4 self-report measures: the DS, DV, SDS, and the Feeling Thermometer. All participants completed the experiment in a single session that lasted approximately 20-30 minutes.

Results

IRAP

Data preparation. The primary datum was response latency (i.e., time in ms between trial onset and a correct response). In accordance with previous IRAP studies, response latency data were transformed into *D*-IRAP scores (see Nicholson & Barnes-Holmes, 2012). The data transformation yielded positive *D*-scores for positive bias, and negative scores for negative bias (i.e., the *D*-scores for the two black trial-types were inverted). A separate overall *D*-IRAP score was calculated, without inverting the black trial-type scores, with positive scores indicating a pro-white/anti-black bias and negative scores indicating a anti-white/pro-black bias.

Trial-type analyses. The *D*-IRAP scores for the four trial-types for black and white participants are presented in Figure 2. The black participants showed positive bias (toward black and white people) across all four trial-types. The white participants also showed a positive bias on the two white trial-types and on the Black-Positive trial-type, but they showed a relatively strong *negative* bias on the Black-Negative trial-type. Note also, that the positive bias by black participants on the *Black-Positive* trial-type was stronger than for white participants.

INSERT FIGURE 2 HERE

A mixed repeated measures 2 x 4 ANOVA was conducted on the *D*-IRAP scores, with race of participant as the between-participant variable and trial-type as the within-participant variable. There was a significant main effect for trial-type, $F(3, 32) = 6.31$, $p < .0006$, $\eta_p^2 = .16$, and for race $F(1, 32) = 11.9$, $p < .001$, $\eta_p^2 = .27$, and a significant interaction, $F(3, 32) = 7.65$ $p < .001$, $\eta_p^2 = .19$. Between-group post-hoc analyses revealed significant differences between black and white participants' performances on the two black trial-types (*p*s < .02), but not on the white trial-types (*p*s > .2).

Eight one-sample *t*-tests indicated that three trial-type effects for black participants were significantly different from zero (*p*s < .001); and the remaining *White-Negative* effect approached significance ($p = .06$). For white participants, *White-Positive* ($p <.0001$) and *Black-Negative* ($p <.03$) were significant (remaining *p*s > .2).

Split-half reliability. To assess the internal consistency of the IRAP, an overall split-half reliability score was calculated for both white and black participants. For the white participants, the overall *D*-IRAP measure produced a strong and significant split-half correlation, $r = .803$, $n = 18$, $p < .001$, but for the black participants it was weak and non-significant ($p = .6$).

Self-Report Measures

DS and DV Scales. The overall means for the DS scales showed pro-black racial discrimination (i.e., mean scores above 3) for white (*M* = 3.76, *SD* = .67) and black participants (*M* = 3.31, *SD* = .27), although a one-way ANOVA indicated that white participants' responses were significantly more positive, *F*(1, 32) = 6.129, *p* < .01, η_p^2 = .16. The overall means for the DV scales also revealed a pro-black bias for both white (*M* = 3.46, *SD* = .73) and black participants (*M* = 3.73, *SD* = .8), although a one-way ANOVA was non-significant (*p* > .3).

SDSs. Four overall means were calculated for the SDSs (white participants/Black People, *M* = .8, *SD* = 1.14; black participants/Black People, *M* = 1.6, *SD* = .6; white participants/White People, *M* = .87, *SD* = 1.14; black participants/White People, *M* = 1.23, *SD* = .55), and all revealed a positive bias (> 0). A 2 x 2 mixed repeated measures ANOVA found a significant main effect for race of participant *F*(1, 32) = 4.32, *p* < .04, η_p^2 = .12, but no other main or interaction effects (*p*s > .09). Follow up tests revealed that black participants rated black people significantly more positively than white participants, *F*(1, 32) = 6.768, *p* < .01, η_p^2 = .17. While black participants also rated white people more positively than white participants, this difference was not significant (*p* > .26).

Feeling thermometers. The overall means obtained on the Feeling Thermometers showed that white participants were more positive about white people than black people were about white people (White, *M* = 74.3, *SD* = 21.3; Black, *M* = 66.5, *SD* = 19.8). In contrast, black participants were more positive about black people than white people were about black people (White, *M* = 74.4, *SD* = 14.1; Black, *M* = 77.5, *SD* = 12.4). A 2 x 2 mixed repeated measures ANOVA yielded no significant main effects (*p*s > .1), but a significant interaction *F*(1, 32) = 13.125, *p* < .001, η_p^2 = .29. Two between-participant follow-up ANOVAs yielded one effect that approached significance; black participants rated black people more positively than white participants rated black people, $F(1, 32) = 3.620$, $p < .07$, $\eta_p^2 = .1$; the rating of white people by black and white participants did not differ significantly ($p > .9$). Two within-participant follow-up ANOVAs indicated that white participants rated white people significantly more positively than they rated black people $F(1, 17) = 9.686$, $p < .006$, $\eta_p^2 = .36$, and black participants rated black people more positively than they rated white people, but only at a level that approached significance, $F(1, 15) = 4.310$, $p < .06$, $\eta_p^2 = .2$.

Correlations Between the IRAP and Self-Report Measures

A correlation matrix of the IRAP and self-report measures was calculated across black and white participants. This involved correlating the four trial-type and overall *D*-IRAP scores with each of the six self-report measures. Out of the 30 correlations, six were significant and two approached significance (all other *p*s > .1), and these are presented in Table 1. For each of the eight correlations, the IRAP effect was consistent with the self-report measure. For example, increased pro-white bias on the *White-Positive* trial-type predicted lower ratings on the black feeling thermometer, whereas increased pro-black bias on the *Black-Positive* trial-type predicted higher ratings on this thermometer. Note also that a negative overall *D*-IRAP score indicated an anti-white/pro-black bias, and thus the negative correlation with the black semantic differential is consistent with the other correlations.

INSERT TABLE 1 HERE

Predictive Validity

A series of hierarchical logistic regression analyses were conducted to determine if one or more of the IRAP measures increased the predictive validity of each of the six self-report measures. The strategy adopted here involved determining if the IRAP measures increased the prediction of group status (black versus white) over and above the self-report measures. If the IRAP did not account for additional variance in this regard, it could be argued that employing such a measure in future research may be of limited value. For illustrative purposes, consider the first regression analysis reported in Table 2. The DS was entered as a predictor of race (i.e., white or black participant) in the first step of the model, and this proved to be weak but significant, $\beta = 1.82$, $p = .03$, accounting for 13% of the variance. The *White-Positive D*-IRAP scores were entered in the second step of the model and this produced virtually no increment in predictive validity, $\beta = 1.35$, $p = .32$, accounting for 15% of the variance (R^2 change = .02). A further four separate models were then created in which the DS was entered as the first step and the remaining IRAP measures were entered as second steps. The *Black-Positive*, *Black-Negative,* and overall *D*-IRAP measures significantly increased the predictive validity of the DS, with the Black-Negative measure yielding the largest increment (R^2 change = .41). The same general strategy was then applied to the remaining five self-report measures (see Table 2) and a similar pattern of results was obtained for these except, that the *Black-Positive* measure did not significantly increase predictive validity for the black semantic differential and black feeling thermometer. In short, the *Black-Negative* and Overall *D*-IRAP measures each significantly increased the predictive validity of each of the six self-report measures. The *Black-Negative* measure, in particular, produced large increases in the percentage of variance accounted for, adding between 36 to 44% to the self-report measures.

INSERT TABLE 2 HERE

Discriminant Analysis

A series of discriminant analyses were performed to determine the extent to which each of the IRAP and self-report measures predicted whether a participant was black or white.

For illustrative purposes, consider the first discriminant analysis reported in Table 3. The value of the discriminant function for the *White-Positive* IRAP measure was not significantly different for black and white participants, $\chi 2(1, 32) = 1.41$, $p = .23$, with the overall function successfully predicting outcome for 67.6% of cases, with accurate predictions being made for 62.5% of the black group, and 72.2% of the white group. This indicated a 37.5% false negative misclassification of the black group, and a 27.8% false positive classification of the white group. The remaining discriminant analyses indicated that three of the IRAP measures (*Black-Positive, Black-Negative,* and Overall *D*-IRAP) and two of the self-report measures (DS and black semantic differential) were significant predictors (the black feeling thermometer approached significance). The best predictor of group status was the *Black-Negative* IRAP measure, predicting outcome for 82.4% of cases.

INSERT TABLE 3 HERE

Discussion

The results from the IRAP revealed an anti-black bias, on the *Black-Negative* trial-type, for the white participants, which contrasted starkly with a pro-black bias for the black participants. A limited number of correlations (6 out of 30) were obtained between the IRAP and the self-report measures, which suggests that there was some functional overlap in the verbal behaviors targeted by the two types of measures. Critically, however, the IRAP provided increased predictive validity over and above the self-report measures. The results for the white participants replicated the earlier study, also conducted in Ireland, reported by Barnes-Holmes, Murphy et al. (2010). Interestingly, the data for the black participants indicated a relatively strong in-group bias, which contrasts with previous research that employed a widely used reaction-time based measure, the Implicit Association Test (IAT).

Specifically, some IAT studies have indicated that black participants fail to produce strong pro-black biases (e.g., Nosek, Banaji, & Greenwald, 2002). On balance, other reaction-time based measures, such as priming and the personalized IAT, have yielded in-group biases for black participants (e.g., Olson**,** Crawford, & Devlin, 2009). Unlike a regular IAT, a personalized IAT involves presenting “I like” versus “I dislike” as label stimuli rather than generic descriptors such as “Pleasant” and “Unpleasant”. Given that the current IRAP involved presenting the labels “I think ….” it could be seen as closer to the personalized IAT and thus the current results are in fact consistent with previous research on racial bias conducted in North America.

As noted earlier, the current study was conducted in 2009 before the publication of the only two other IRAP studies that have employed black participants (Drake et al., 2010; 2015). Any direct comparison between the current work and the results reported by Drake et al. must be made with caution because the latter studies were conducted in North America. Furthermore, there were many procedural differences across the studies. For example, Drake et al. (2010) required participants to maintain an accuracy criterion of 65% during the test blocks whereas this was set at 80% in the current study. Nevertheless, the findings across the three studies do overlap to some extent, but there are some differences, particularly at the trial-type level of analysis. In the current study, the IRAP produced positive bias scores among the black participants for both black and white people; the bias scores for the white participants were more variable with a relatively strong positive bias on the *White-Positive* trial-type and negative bias on the *Black-Negative* trial-type. In the Drake et al. (2010) study, the pattern was broadly similar in that the black participants produced positive bias scores across all four trial-types, whereas the white participants did not. In the Drake et al. (2015)

study, however, both white and black participants produced bias scores that varied across the four trial-types.

The current findings suggest that black people residing in Ireland do not show a negative bias toward white people on the IRAP. As noted in the introduction, black people living in Ireland perhaps differ considerably from black people residing in many other countries, in that those recently immigrated to and living in Ireland in some cases may have more recent experiences of persecution and/or imperial oppression. It would thus be interesting to repeat this study in, for example, the United Kingdom with black participants who have resided there for two or more generations. Indeed, future research might seek to determine if the children of the participants in the current study continue to show positive white bias. The fact that white participants showed negative bias (particularly on the *Black-Negative* trial-type) suggests that black people living in Ireland might be subjected to various forms of racial discrimination over the coming years and perhaps the positive bias shown by black participants will suffer as a result. In any case, the current data are important because they provide a record of racial bias in Ireland using both a reaction-time based measure and a range of self-report instruments at a particularly interesting time in Ireland's cultural evolution.

In general, there was limited evidence of between-group effects indicative of racial discrimination or stereotyping obtained from the self-report measures in the current study, although there were some exceptions. For example, the feeling thermometers indicated that white participants were significantly more positive when rating white people than when rating black people; in addition, black participants were more positive when rating black people than when rating white people, although this effect only approached significance. Thus, the IRAP was not the only measure to reveal between-group differences consistent with in- versus out-group response biases. On balance, the starkest contrast was observed for the *Black-Negative* trial-type on the IRAP, and indeed the regression and discriminant analyses bore this out. This finding indicates that it may well be useful when studying racial discrimination to include reaction-time based measures, such as the IRAP, in order to capture additional sources of variance beyond those provided by self-report measures.

It is interesting to note that the current data replicated a perhaps counter-intuitive result for the IRAP, in that white participants showed a positive bias (albeit non-significant) for black participants on the *Black-Positive* trial-type, but a relatively strong negative bias on the *Black-Negative* trial-type. As noted in the introduction, scores for individual trial-types on the IRAP should be interpreted with caution because IRAP effects may be moderated by a range of variables, including generic verbal biases inherent in natural language. On balance, a simple explanation in terms of such generic biases in the current study is problematic because it was only observed for the white participants. An explanation would thus seem to require identifying a relevant difference between the two racial groups. Perhaps, white participants are more prone to producing anti-black response biases when presented with negatively valenced stimuli on the IRAP because they are frequently subjected to negative portrayals of black people through the popular media. However, when presented with positively valenced stimuli (on the *Black-Positive* trial-type), responding was controlled more by the history of positive exemplars of black people that are also presented in the media (e.g., Barack Obama, Nelson Mandela, Morgan Freeman). Of course, one might ask why black people living in Ireland did not show a similar contrasting pattern for the out-group (i.e., positive bias on the *White-Positive* trial-type and negative bias on the *White-Negative* trial-type), given that the black participants would also have been exposed to both positive and negative examples of white people through the popular media. One explanation is that white people, *as a distinct racial group*, are not *strongly* stereotyped either positively or negatively in the Western media (but see also, for example, Conley, 2012; Conley, Rabinowitz, & Rabow, 2010; Conley & Ramsey, 2011). Furthermore, and as noted earlier, black residents in Ireland may have some sense of political and economic sanctuary, at least in the short to medium term, relative to their home nation.

Although the current findings replicate and extend previous research on racial discrimination in Ireland, a number of limitations should be noted. First, the black and white participants were not specifically matched for a range of demographic variables, such as levels of education, socio-economic status, and language ability. Thus, for example, while English was the first language of all of the white participants, this may not have been the case for all of the black participants, which may have impacted in some unexpected way on the IRAP performances. On balance, all participants who provided data for the final set of analyses were required to achieve the same performance criteria on the practice blocks of the IRAP, and maintain them during the test blocks. Furthermore, the *D*-algorithm (Greenwald, Nosek, & Banaji, 2003) that was used to transform the IRAP latency data controls, to some extent, for potentially confounding variables. As such, language and other demographic variables would not be expected to produce the observed between-group differences, and in particular the quite dramatic difference observed on the *Black-Negative* trial-type. On a related note, some demographic variables could prove potentially interesting for conducting future work. For example, as mentioned previously, there is some difficulty in directly comparing the current research with the previous Drake et al. studies because of the different demographics of participants used (i.e., recent immigrants versus participants who may not have been or for whom this information was unknown). Taken together, perhaps future research could manipulate this variable directly by investigating potential differences in response biases between recent immigrants, and those who have resided in the country of interest for multiple generations.

Another limitation to the current study is that there was no attempt to determine if the differences found on the IRAP or self-report measures actually predicted racially-biased behavior, using for example some form of behavioral approach task (Amodio & Devine, 2006). In addition, no attempt was made in the current study to explore methods for reversing or at least reducing the negative response bias obtained on the IRAP for the white participants. Future research could certainly pursue these and related issues. For example, research conducted by Lillis and Hayes (2007) compared two approaches to reducing racial and ethnic prejudice: one protocol based on Acceptance and Commitment Therapy and an education-based protocol drawn from a well-known textbook on the psychology of racial differences. Perhaps future studies could examine the impact of protocols such as these on IRAP performances and other measures of racial bias to determine if the various measures are equally or differentially affected by the protocols.

Compliance with Ethical Standards

Declaration of Interest: This article was prepared with the support of an Odysseus Group 1 grant awarded to the third author by the Flanders Science Foundation (FWO). The data collected for this research was attained with the support of funding awarded to the first author by the Irish Research Council (IRC). The authors declare no other conflicts of interest.

Ethical Approval: All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent: Informed consent was obtained from all individual participants included in the study.

References

Amodio, D.M. & Devine, P.G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology, 91*(4), 652-661. doi: 10.1037/0022-3514.91.4.652

Barnes, D., Lawlor, H., Smeets, P.M., & Roche, B. (1996). Stimulus equivalence and academic self-concept among mildly mentally handicapped and nonhandicapped children. *The Psychological Record, 46*, 87-107.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*, 527-542.

Barnes-Holmes, D. Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record, 58*, 497-515.

Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record, 60*, 57-66.

Central Statistics Office Ireland (2006). Retrieved December 10, 2009, from http://www.cso.ie/census/default.htm.

Cohen, J. (1988).*Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Conley, T.D. (2012). Beautiful, self-absorbed, and shallow: People of color perceive white women as an ethnically marked category. *Journal of Applied Social Psychology, 43*(1), 45-56. doi: 10.1111/j.1559-1816.2012.00980.x

Conley, T.D., Rabinowitz, J.L., & Rabow, J. (2010). Gordon Gekkos, frat boys and nice guys: The content, dimensions, and structural determinants of multiple ethnic minority groups' stereotypes about White men. *Analyses of Social Issues and Public Policy, 10*, 69-96. doi: 10.1111/j.1530-2415.2010.01209.x

Conley, T.D., & Ramsey, L R. (2011). Killing us softly? Investigating portrayals of women and men in contemporary magazine advertisements. *Psychology of Women Quarterly, 35*(3), 469-478. doi: 10.1177/0361684311413383

Dixon, M., Rehfeldt, R.A., Zlomke, K.M., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record, 56,* 83-103.

Drake, C.E., Kellum, K.K., Wilson, K.G., Luoma, J.B., Weinstein, J.H., & Adams, C.H. (2010). Examining the Implicit Relational Assessment Procedure: Four preliminary studies. *The Psychological Record, 60,* 81-86.

Drake, C.E., Kramer, S., Sain, T., Swiatek, R., Kohn, K., & Murphy, M. (2015). Exploring the reliability and convergent validity of implicit racial evaluations. *Behavior and Social Issues, 24,* 68-87. doi: 10.5210/bsi.v.24i0.5496

Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioral dynamics of the Implicit Relational Assessment Procedure: The impact of three types of introductory rules. *The Psychological Record, 66*, 309-321. doi: 10.1007/s40732-016-0173-4

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. doi: 10.1037/0022-3514.85.2.197

Hayes, S.C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A post Skinnerian account of human language and cognition.* New York, NY: Plenum.

Leslie, J.C., Tierney, K.J., Robinson, C.P., Keenan, M., Watt., A., & Barnes, D. (1993). Differences between clinically anxious and non-anxious subjects in a stimulus equivalence training task involving threat words. *The Psychological Record, 43,* 153-161.

Lillis, J., & Hayes, S.C. (2007). Applying acceptance, mindfulness, and values to the reduction of prejudice: A pilot study. *Behavior Modification, 31*(4), 389-411. doi: 10.1177/0145445506298413

Maloney, E. & Barnes-Holmes, D. (2016). Exploring the behavioral dynamics of the Implicit Relational Assessment Procedure: The role of relational contextual cues versus relational coherence indicators as response options. *The Psychological Record, 66*, 395-403. doi:10.1007/s40732-016-0180-5

Merwin, I.M., & Wilson, K.G. (2005). Preliminary findings on the effects of self-referring and evaluative stimuli on stimulus equivalence class formation. *The Psychological Record, 55*, 561-575.

Nicholson, E. & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure (IRAP) as a measure of spider fear. *The Psychological Record, 62,* 263-278.

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamic Theory, Research, and Practice, 6*(1), 101-115.

McKenna, I. Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology and Psychological Therapy, 7*, 253-268.

Olson, M. A.**,** Crawford, M. T., & Devlin, W. (2009). Evidence for the underestimation of implicit in-group favoritism among low status groups. *Journal of Experimental Social Psychology*, *45*, 1111-1116.

Power, P.M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of relative preferences: A first study. *The Psychological Record*, *59*, 621-640.

Vahey, N.A., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, *59*, 371-388.

Vahey, N.A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry, 48,* 59-65. doi: 10.1016/j.jbtep.2015.01.004

Watt, A.W., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record*, *41,* 371-388.

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

Table 1

Summary of Six Significant (and Two Approaching Significant) Correlations between the Implicit and Explicit Measures Calculated Across Black and White Participants (*N* = 34).

IRAP Trial-type	**Self-Report Measure**	***r***	***p***
White-Positive	Black feeling thermometer	-.35	.04*
Black-Positive	Black semantic differential	.39	.02*
Black-Positive	Black feeling thermometer	.41	.01*
Black-Negative	Black semantic differential	.34	.05*
Black-Negative	Black feeling thermometer	.34	.04*
Overall *D*-IRAP score	Black feeling thermometer	-.38	.02*
White-Negative	Diversity scale	-.33	.06
Overall *D*-IRAP score	Black semantic differential	-.32	.07

**p* < .05

Table 2

Summary of Hierarchical Logistical Regression Analysis for the Variables Predicting Race of Participants (*N* = 34).

Step 1				Step 2				
Self-Report Measure				Self-Report Measure + IRAP				
Predictor Variables	B	R^2	*p*	Predictor Variables	B	R^2	*p*	R^2 Change
Discrimination Scale	1.82	.13	.03*	Discrimination Scale +				
				White-Pos D-IRAP	1.35	.15	.32	.02
				White-Neg D-IRAP	0.07	.13	.94	0
				Black-Pos D-IRAP	3.09	.27	.03*	.14
				Black-Neg D-IRAP	7.53	.54	.02*	.41
				Overall D-IRAP	6.25	.34	.03*	.21
Diversity Scale	0.49	.02	.30	Diversity Scale +				
				White-Pos D-IRAP	1.48	.06	.25	.04
				White-Neg D-IRAP	0.78	.04	.39	.02
				Black-Pos D-IRAP	2.67	.15	.04*	.13
				Black-Neg D-IRAP	6.50	.45	.02*	.43
				Overall D-IRAP	4.57	.19	.02*	.17
Semantic Differential (S	0.97	.14	.02*	SD Black +				
Black				White-Pos D-IRAP	1.18	.15	.38	.01
				White-Neg D-IRAP	0.64	.14	.52	0
				Black-Pos D-IRAP	2.04	.20	.13	.06
				Black-Neg D-IRAP	7.42	.50	.02*	.36
				Overall D-IRAP	4.35	.26	.04*	.12
Semantic Differential (S	0.47	.03	.26	SD White +				
White				White-Pos D-IRAP	1.37	.05	.30	.02
				White-Neg D-IRAP	0.59	.04	.53	.01
				Black-Pos D-IRAP	2.65	.15	.05*	.12
				Black-Neg D-IRAP	6.56	.45	.02*	.42
				Overall D-IRAP	4.97	.21	.02*	.18
Feeling Thermometer	0.04	.08	.08	FT Black +				
(FT) Black				White-Pos D-IRAP	1.91	.09	.50	.01
				White-Neg D-IRAP	0.69	.09	.49	.01
				Black-Pos D-IRAP	2.22	.15	.10	.07
				Black-Neg D-IRAP	6.69	.45	.02*	.37
				Overall D-IRAP	4.16	.21	.04*	.13
Feeling Thermometer	0.00	.00	.99	FT White +				
(FT) White				White-Pos D-IRAP	1.65	.04	.22	.04
				White-Neg D-IRAP	0.68	.01	.44	.01
				Black-Pos D-IRAP	2.72	.13	.04*	.13
				Black-Neg D-IRAP	6.63	.44	.01*	.44
				Overall D-IRAP	4.39	.17	.02*	.17

**p* < .05

Table 3

Summary of Discriminant Analyses for the Variables Predicting Race of Participants (*N* = 34).

	χ2	df	*p*	Group	Predicted Percentage Group Membership		Overall Prediction
					Black	White	
White-Pos *D*-IRAP	1.41	1, 32	.23	Black	62.5	37.5	67.6
				White	27.8	72.2	
White-Neg *D*-IRAP	.56	1, 32	.46	Black	37.5	62.5	58.8
				White	22.2	77.8	
Black-Pos *D*-IRAP	5.31	1, 32	.02*	Black	68.8	31.3	67.6
				White	33.3	66.7	
Black-Neg *D*-IRAP	16.38	1, 32	.00*	Black	93.8	6.3	82.4
				White	27.8	72.2	
Overall *D*-IRAP	7.23	1, 32	.01*	Black	68.8	31.3	64.7
				White	38.9	61.1	
Discrimination Scale	5.52	1, 32	.02*	Black	87.5	12.5	73.5
				White	38.9	61.1	
Diversity Scale	1.06	1, 32	.30	Black	43.8	56.3	47.1
				White	50.0	50.0	
Semantic Differential Black	6.04	1, 32	.01*	Black	68.8	31.3	67.6
				White	33.3	66.7	
Semantic Differential White	1.26	1, 32	.26	Black	75.0	25.0	64.7
				White	44.4	55.6	
Feeling Thermometer Black	3.38	1, 32	.07	Black	56.3	43.8	61.8
				White	33.3	66.7	
Feeling Thermometer White	.00	1, 32	.99	Black	50.0	50.0	41.2
				White	66.7	33.3	

*p < .05

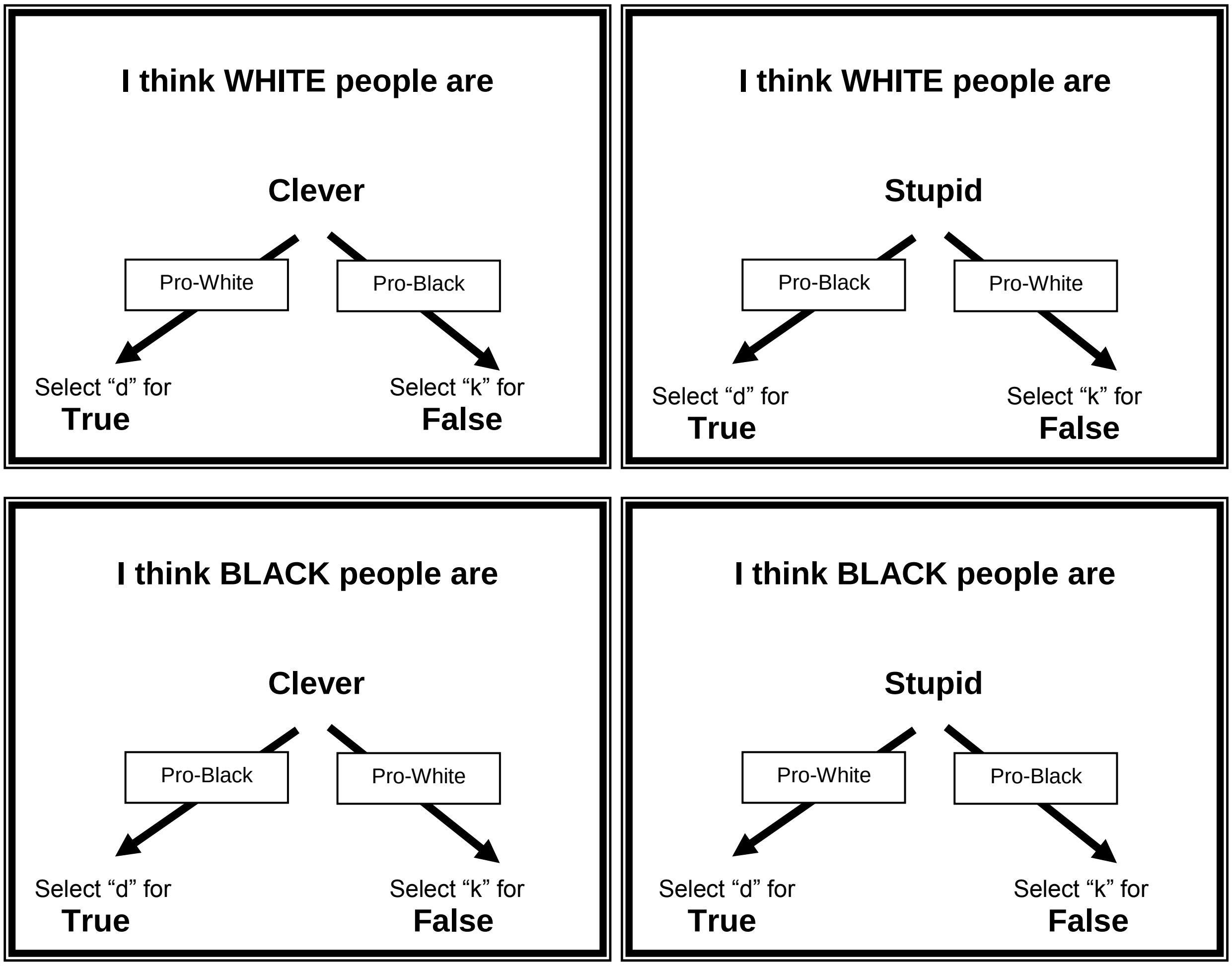

Fig. 1 Diagrammatic representation of the four IRAP trial-types. Arrows and boxes containing the words *Pro-White* and *Pro-Black* did not appear on-screen. The four IRAP trial-types were denoted as: *White People-Positive; Black People-Negative; Black People-Positive;* and *White People-Negative*

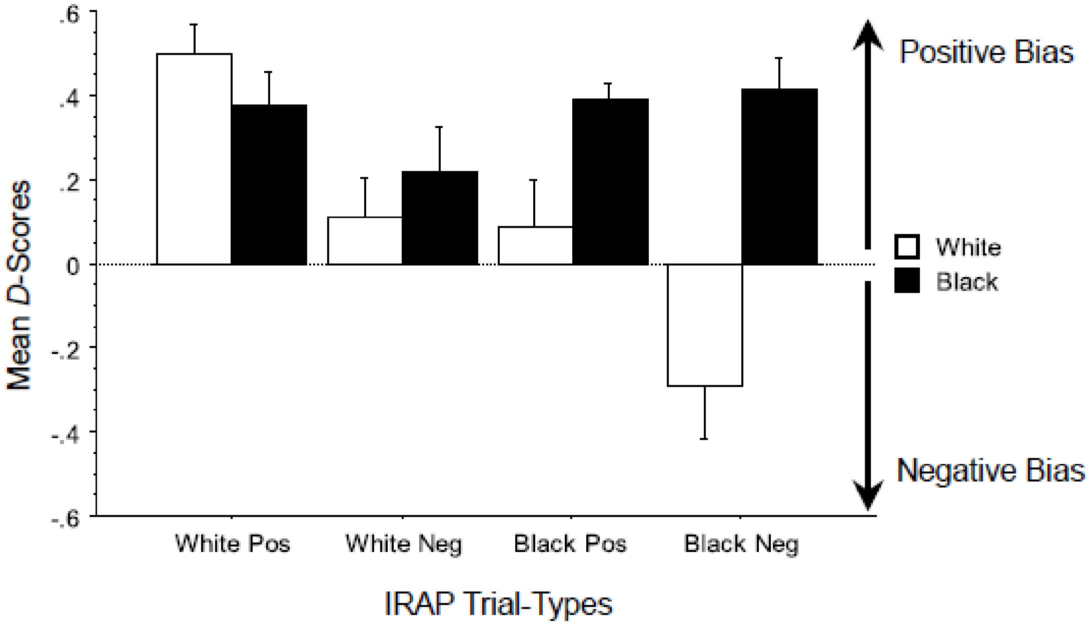

Fig. 2 The mean *D*-IRAP scores, with standard error bars, for the IRAP four trial-types.