International Journal of Psychology and Psychological Therapy, 2014, 14, 3, 333-351

Implicit Relational Assessment Procedure and Attractiveness Bias: Directionality of Bias and Influence of Gender of Participants

Carol Murphy*, Saoirse MacCarthaigh, Dermot Barnes-Holmes

National University of Ireland, Maynooth, Ireland

Abstract

The behaviorally-based Implicit Relational Assessment Procedure (IRAP) was used for the first time in the area of attractiveness bias with adult participants (*N*=47). Alternate IRAP trial-blocks required participants to affirm relations consistent (attractive-successful) and inconsistent (unattractive-successful) with attractiveness bias; shorter mean response latencies (faster responding) across consistent trial-blocks were interpreted in terms of implicit stereotype. Participants also completed a Likert-type scale rating of the successfulness of attractive versus unattractive individuals. Statistical analyses were conducted with implicit and explicit data, and both indicated attractiveness bias for male and female participants. Directionality of implicit bias was analyzed via the IRAP 4 trial-type methodology to determine if bias was pro-attractive or anti-unattractive, or if it was evident in both directions; a statistically significant implicit proattractive and antiunattractive bias was shown for male and female participants, and the effect was greater for male participants. Modest correlations were found between implicit and explicit data. Findings are discussed regarding a comprehensive and nuanced account of attractiveness bias, directionality, and contextual influences.

Key words: attractiveness bias, implicit bias, IRAP, RFT.

Novelty and Significance

What is already known about the topic?

- Attractiveness bias is a robust finding in the cognitive/associationist (indirect) research literature using self-report methods in the domain of social interpersonal relations.
- In the area of mate selection, males show greater attractiveness bias.
- The IRAP is a relatively recent program that has been useful as a measure of implicit bias in socially sensitive areas, when participants may not wish to report prejudice.

What this paper adds?

- Tested the IRAP as a new behavioral (direct) measure that avoided relying on participants to honestly report if they had attractiveness bias in the context of evaluations of successfulness.
- Shed light upon directionality of prejudice (i.e., if there is bias, is it proattractive but not antiunattractive? is it antiunattractive but not proattractive? is it a combination of both?).
- Showed greater attractiveness bias for males in evaluations of successfulness - infrequently examined outside of mate selection research; important from a basic research perspective, and regarding practical strategies aimed at reducing negative prejudicial effects.
- Examined participants' implicit attractiveness data as well as their self-reported (rating) data; the IRAP implicit measure showed greater magnitude in attractiveness bias shown for males but the self-report method did not detect this.

Attractiveness is thought to impact on all interpersonal relationships (Feingold, 1992) and research has shown that humans are vulnerable to influence of attractiveness bias in multiple social contexts (Eagly, Ashmore, Makhijani, & Longo, 1991). Research

* Correspondence concerning this article: Carol Murphy, Department of Psychology, John Hume Building, National University of Ireland, Maynooth, Maynooth, Co Kildare, Ireland; E-mail: Carol.A.Murphy@nuim.ie

participants reported more favorable first impressions for attractive versus unattractive individuals (Eagly *et al*. 1991); participants rated attractive v. unattractive individuals as more persuasive (Chaiken, 1979); attractive v. unattractive pianists' performances were rated more positively (Ryan & Costa-Giomi, 2004); and attractive v. unattractive singers were rated more positively (Wapnick, Darrow, & Kovacs, 1997). Attractiveness bias has been demonstrated when participants evaluated individuals in date and mate selection decisions (Adams, 1977), regarding helpful behavior (Benson, Karabenick, & Lerner, 1976) and even in judgments in simulated court trials (Mazella & Feingold 2006). The phenomenon that attractive individuals are frequently perceived by others in more favorable terms than less attractive counterparts was termed the "What is beautiful is good" stereotype by Dion, Berscheid, and Walster (1972). Importantly, employee selection may also be influenced by attractiveness, and research has shown that better-looking candidates were selected over equally qualified but less attractive candidates (Marlowe, Schneider, & Nelson, 1996). Meta-analyses suggest the findings of an attractiveness bias is a robust phenomenon (Eagly *et al*. 1991), however, some relevant qualifications should be noted regarding contextual influences and limitations in the available research literature.

There have been a small number of conflicting findings in which contextual factors appeared to exert a contrary effect, in that attractiveness for women was disadvantageous when considerations were for jobs traditionally viewed as masculine ("Beauty is Beastly effect"; Heilman & Stopeck, 1985). In the context of employment and applicant selection, research participants showed an antiunattractiveness bias toward unattractive female applicants but not toward unattractive male applicants (Hosoda, Stone-Romero, & Coat, 2003). Beauty was found to be beastly also if evaluators were the same sex as individuals evaluated; recent findings in organizational research indicates that attractiveness may produce a positive bias in different-sex evaluators but may not result in a positive bias and may even impact negatively if the evaluator is the same sex as the target evaluatee (Agthe, Spörrle, & Maner, 2011). Further, although the bulk of research studies in the literature on beauty bias is robust in indicating a positive impact for attractiveness, studies are mostly comprised of self-report or questionnaire ratings related to the social domain. This presents a limitation in that vulnerabilities regarding the accuracy of self-report measures in psychological research have been well documented (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Nosek, Greenwald, & Banaji, 2007). Self-report data may be prone to confounding influences of self-presentation or social desirability effects. Particularly when socially sensitive issues are investigated, self-presentational problems may exert influence; consider that in questionnaire investigations into racial prejudice participants might not wish to indicate that they perceive black individuals with guns as more dangerous than white individuals with guns, because of the implication of an undesirable racial bias (Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010); thus they may purposefully and readily provide fake ratings that avoid showing negative bias. Regarding attractiveness bias, Feingold (1990) suggested that questionnaire research on attractiveness in mate selection could be distorted by sex differences in self-presentational concerns. Feingold's meta analysis of the influence of attractiveness

in mate selection consistently showed that men but not women selected primarily based on attractiveness; Feingold noted that if it were more socially desirable for men to select based on attractiveness than for women to do so (why this might be the case was not made clear), women might be disproportionately influenced by a wish to conceal evidence of "shallowness". This may affect data but another factor is that in many cases the self-report research literature on attractiveness bias in the general social domain is based on group data that has not been analyzed gender for influences. If female, or indeed male, participants purposefully conceal beauty bias in explicit measures, implicit behavioral measures may provide alternative or complementary enlightening information showing divergence between explicit and implicit data. Thus, a more complete and coherent account of attractiveness bias might be facilitated by viewing participants' implicit (see below) as well as explicit data, and conducting gender analysis. Recent research findings with participants with Body Dysmorphic Disorder (BDD) used explicit and implicit measures and found that implicit data showed greater emphasis placed on the importance of attractiveness for the BDD group compared to control groups, but the effect was not shown in a comparison of explicit data for the BDD and control participants (Buhlmann, Teachman, & Kathmann, 2010).

An additional problem is that even if participants wish to respond honestly, explicit self-report measures can only evaluate in terms that the participant is consciously aware of and can introspect accurately. Introspection in psychological research has been found to be problematic, however, and many participants may not be aware of their own bias toward social groups (De Houwer, 2002, 2008). To overcome problems with explicit self-report data, researchers have recently begun to use measures of behavioral responses, or "implicit" measures of bias or prejudice (Greenwald, McGhee, & Schwartz, 1998). For example, measures of participants speed of responding via response latencies are considered useful behavioral indicators of prejudice; faster categorizations of "thin" with positive attributions (smart, attractive, energetic) compared to slower categorizations of "fat" with similar positive attributes may be deemed to be evidence of a pro-thin bias. It has been found when research uses behavioral measures in addition to self-report data in socially sensitive research that participant behaviors may fail to accord with explicitly reported data (Van Lange, 2006). Furthermore, self-report measures may be poor predictors of racial discrimination behavior (Crosby, Bromley, & Saxe, 1980). White research participants may respond positively toward black people in a questionnaire test for racial bias, but their responding on an implicit measure may show racial bias. Research has shown that participants who claimed not to hold racist attitudes more readily categorized names typical of White persons with positive words and names typical of Black persons with negative words, but responding was less speedy and accurate when participants were asked to categorize White with negative and Black with positive (Greenwald *et al*., 2002). Self-report and explicit rating measures may be considered an important research tool for psychologists in that they are efficient for use with large participant samples; it should be noted also they have well-established predictive validity in areas such as consumer preferences, political preferences, and clinical phenomena (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Nonetheless, considering their inherent limitations, complementing explicit reports with results from

implicit behavioral measures may facilitate a more nuanced and comprehensive approach to examining attractiveness bias than would be possible with either measure used alone. Also the data obtained from explicit and implicit measures, whether convergent or divergent, may be informative about different types of responding. It is thought that explicit measures may involve intentional, controlled, and conscious deliberation, whereas implicit measures involve rapid responding under time pressure, and may detect heuristic "automatic" or impulsive responding (e.g., Fazio & Olson, 2003; Spalding & Hardin, 1999). In the context of the current study the term "implicit" should not be understood as a mentalistic description, and is merely used to refer to bias that is not reported by participants but instead is inferred by the researcher on the basis of behavioral measures (i.e., "implicit" in the pattern of participant's recorded response latencies). The behavioral term used to refer to the type of automatic or impulsive responding captured via implicit measures is "brief, immediate relational responding" (BIRR), as distinct from more intentional or deliberative responding found in self-report measures which appears to involve "relational elaboration and coherence or REC" (for an expanded discussion see Barnes-Homes, Barnes-Holmes, Stewart, & Boles, 2010). To illustrate, under time pressure to respond rapidly, participants' BIRRs may affirm men-smart more rapidly than women-smart, indicating prejudice based on sexism; however, given time to reflect and report on the matter in a questionnaire, more complex relations may come into play, such as sexism-bad, egalitarianism-good, evidence-more-reliable-than-supposition. Thus participants may report on the explicit questionnaire that men and women are equally smart, and the implicit and explicit measures may diverge.

The most popular of the current implicit measures is the Implicit Association Test (IAT; Greenwald *et al*., 1998), which is frequently used to measure bias in socially sensitive areas. The IAT is a computerized program that was designed based on a relatively simple associative assumption, which is that it should be easier to categorize concepts that are strongly associated in participant memory compared to concepts that are weakly associated. The IAT task requires participants to respond quickly when trials present paired stimuli in alternate trial-blocks that are considered consistent or inconsistent with preexperimental memory associations. If participants respond more rapidly across trial-blocks presenting pairings such as 'thin-positive' (consistent), versus trial-blocks presenting 'fat-positive' (inconsistent), the differential in the response latencies is interpreted in terms of an implicit pro-thin bias. Speed of responding is calculated by averaging recorded response latencies across consistent and inconsistent trial-blocks. The IAT automatically records response latencies and subsequently the mean response latencies for consistent trial-blocks is subtracted from the mean latencies for inconsistent trial-blocks, providing a score for the differential that is subjected to statistical analysis for significant bias effect. The IAT has been useful as a measure of implicit bias in various socially sensitive domains, such as homophobia (e.g., Banse, Seise, & Zerbes, 2001), gender stereotypes (e.g., Rudman & Glick, 2001), racism (e.g., Greenwald *et al*., 1998), religious stereotyping (e.g., Rudman, Greenwald, Mellott, & Schwartz, 1999), and even self-esteem (e.g., Bosson, Swann, & Pennebaker, 2000). A relatively recent development is a behavioral adaptation of the IAT known as the Implicit Relational Assessment Procedures (IRAP; Barnes-Holmes *et al*., 2006; Power, Barnes-Holmes,

Barnes-Holmes, & Stewart, 2009). The IRAP differs conceptually from the IAT in that it is nonassociationist, and is proposed as a behavioral measure of preexperimentally learned verbal relations rather than prelearned memory associations. The IRAP methodology may be considered more direct and parsimonious in that the focus is on relational behavioral responding avoiding reference to indirect (hypothetical) underlying mental phenomena. Further, an important aspect of the IRAP as an implicit measure is that four relational trial-types facilitate subsequent specification of directionality of any bias detected, whereas other implicit measures including the IAT may fail to provide information on directionality of stereotype.

The IRAP is a computerized program that is freely available to researchers via the internet (software and sample instructions are available from http://irapresearch.org/downloads-and-training/). The program is derived from a comprehensive behavioral approach to human language and cognition known as Relational Frame Theory (for a complete account see RFT; Hayes, Barnes-Holmes, & Roche, 2001). The RFT approach posits that the core components of advanced cognition are relational processes, rather than associations. The IRAP may be considered a behaviorally-based adaptation of the IAT that similarly uses a computer-based task to present 'consistent' and 'inconsistent' trials, and automatically records response latency data to determine speed of participant responding across alternate trial-blocks. A differential score for response latencies averaged across consistent and inconsistent IRAP trial-block is calculated similarly as with the IAT. The IRAP program involves onscreen presentation of one of two sample stimuli appearing on separate trials, with one of a range of target positive or negative attributes, and two relational response options. For example, a photographic image of a thin or an overweight individual might be presented with either the attribute 'intelligent' or 'stupid', and participants must select a relational response option such as 'similar' or 'opposite'. During consistent trials, participants may be required to confirm thin-intelligent-similar relations and during inconsistent trials to confirm fat-intelligent-similar relations, and faster responding during consistent trials would be interpreted in terms of an implicit pro-thin bias (e.g., see Nolan, Murphy, & Barnes-Holmes, 2013; Roddy, Stewart, & Barnes-Holmes, 2010). The procedural use of relational terms (e.g., Same/Different; True/False; Similar/Opposite) and four trial-types means that the IRAP may detect that participants respond faster to thin-positive v. fat-positive relations, but can also provide information as to whether the difference comprises a pro-thin bias, an anti-fat bias or a combination of both (Roddy *et al*., 2010; Nolan *et al*., 2013). Preliminary findings have provided support for IRAP reliability (Power *et al*., 2009) and validity (Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009), and indicated that the IRAP was not amenable to "fake-ability" when participants were instructed to control responding (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007). The IRAP has been shown also to have predictive validity toward behavior (see Roddy *et al*., 2010; Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009).

In the area of attractiveness bias and relevant to the current research, the IRAP may be a useful research tool in that it may function as an implicit measure that can provide information regarding directionality of any beauty bias evident with research participants. The issue of directionality may be considered important to a comprehensive

understanding of this phenomenon, because lack of clarity on this point has been noted; the question of whether beauty is good or ugly is bad has been subjected to limited scrutiny in the research literature (Griffin & Langlois, 2006). Specifically, the extent to which attractiveness is advantageous to the individual, or unattractiveness is disadvantageous may need to be elucidated (Eagly *et al*., 1991; Langlois *et al*., 2000). Recent findings suggest that it may be more often the case that unattractiveness is "bad" rather than that beauty is "good"; on the other hand, stereotype directionality was shown to vary depending upon domain of judgment (Griffin & Langlois). For example, anti-unattractive bias was evident in judgments of altruism and intelligence, whereas both a pro-attractive and an anti-unattractive bias was shown related to the domain of sociability.

The current research sought to use both implicit (IRAP) and explicit measures (Likert-type questionnaire) to examine attractiveness bias with adult participants in the context of successfulness attributions. The study used facial photographs and focused on facial attractiveness because in evaluating an individual's overall-attractiveness, facial attractiveness is thought to be of primary importance (Dickey-Bryant, Lautenschlager, Mendoza, & Abrahams, 1986; Chung & Leung, 1988; Heilman & Stopeck, 1985). The IRAP program presented an attribute stimulus (e.g., Successful/Unsuccessful) with an attractive or unattractive photo facial image above two response option (Similar/Opposite) and participants were required to affirm Attractive-Successful and Unattractive-Successful relations across alternated consistent and inconsistent trial-blocks. Latency data were recorded automatically by the program and were used to determine if participants more rapidly affirmed consistent rather than inconsistent relations. Directionality of bias detected was also examined via the data from four relational trial-types in the IRAP methodology. The explicit measure used was a Likert type rating scale that required participants to rate "Successfulness" of attractive v. unattractive photographic facial images identical to those presented during IRAP trials. Specific research aims were (a) to determine if the IRAP had potential as a behavioral measure sensitive to implicit attractiveness bias with college students (N= 47); (b) to elucidate directionality of any implicit IRAP bias effect detected regarding a proattractive bias, an antiunattractiveness bias, some combination of both, or indeed no stereotype bias whatever; d) to examine participants' explicit and implicit data for effects of gender of participants; (c) to analyze participants' explicit and implicit data for correlations.

METHOD

Participants

Participants were undergraduate students at the National University of Ireland, Maynooth. Only college students enrolled in a course other than psychology were included as participants in the experiment, therefore participants had no prior experience with the IRAP procedure. Fifty participants (25 male, 25 female) with an age range of 18-28 years of age (M= 23, SD= 2) took part in the experiment at the Department of Psychology at NUIM. All participant volunteers were of Caucasian ethnicity, English-speaking with

normal or corrected to normal vision. Data from 3 participants were excluded because they failed to achieve the predetermined performance criterion of 75% accuracy on the IRAP. Prior to the commencement of the experiment, participants were briefed as to the general nature of the study. They were informed that the experiment would include a computer-based task and two brief questionnaires, and that the data would be analyzed at a group level. No financial or other incentive, other than the knowledge that they were assisting in scientific research, was offered for participation in the experiment. Participants signed an informed consent form, and all procedures were conducted in accordance with current ethical standards in psychology. The research project was approved by the Ethical Committee at NUI, Maynooth, and procedures were undertaken in accordance with current ethical standards in psychology and behavior analysis taking due regard to participant voluntariness, informed consent and data confidentiality and protection.

Materials

The Implicit Relational Assessment Procedure (IRAP) was administered using a portable Intel Pentium 4 laptop with a 15 inch monitor operating with the Microsoft Windows 7 system. The IRAP software was used to present the experimental trials and record participants' responses. Each IRAP trial consisted of a presentation of two category labels "Successful" or "Unsuccessful," one of twelve target stimuli (facial images), and two response options, "Similar" and "Opposite." The twelve target stimuli were digital color photographs that had been culled from a number of copyright free internet sources (http://www.uni-regensburg.de, http://www.thaimedicalnews.com, http://www.flickr.com). Of the twelve photographic stimuli, eight images were of adult "Attractive" faces (four male and four female) and four images were of adult "Unattractive" faces (two male and two female). Digital images available from the websites depict attractive and unattractive males and female faces that do not exist in reality, and images are in fact a composite of about 30 faces morphed together using computer software to create "attractive" and "unattractive" faces. Digital composite faces are frequently used in this type of research to readily manipulate attractiveness and avoid ethical problems related to use of images of real people. Research has shown that youthfulness and symmetry are preferred by both sexes universally, and computerized "averaging" or blending of multiple images makes people more attractive (an effect also seen for birds, fish and cars; see Halberstadt & Rhodes, 2003). The number of attractive images in the current study exceeded that of the unattractive ones in order to adjust for a potentially greater salience of negative compared to positive stimuli that was assumed to parallel greater salience attributed to losses over gains. This adjustment was modeled from research conducted by Yamamoto, Ariely, Chi, Langleben, & Elman (2009) examining gender differences and motivational processing of facial attractiveness. Since all participants were of Caucasian ethnicity, all facial images used in the experiment were of Caucasian individuals in order to avoid racial in-group prejudices confounding the results. To ensure image consistency and symmetry, the stimuli were standardized for size and equalization of distances between standard facial landmarks (for instance there was a pupil to pupil distance of 2.3cm and temple to temple distance of 3.2cm). Adobe Photoshop 12.0 was used to ensure

image consistency. The size of each image was 336 x 339 dpi with RGB color. The same photographic images were used for implicit and explicit measures; in addition the latter also involved two Likert-type measures for rating attractiveness and successfulness.

Procedure

Participants (the majority) undertook experimental procedures in the Department of Psychology at the National University of Maynooth, Ireland, or else in their own homes. All experiments were conducted on an individual basis in a quiet room free from distractions with the participant seated comfortably at a table in front of a PC. Participants first completed explicit rating procedures and then completed the IRAP procedure.

Explicit Measures

Attractiveness Rating Measure. The composite photographic images were pre-designated as attractive or unattractive when drawn from relevant websites (on the basis of website user votes). To ensure that the current research participants concurred with pre-designated categories, an Attractiveness 7-point Likert-type questionnaire was used to measure participant ratings of images. The Attractiveness scales ranged from -3 (Unattractive) to +3 (Very Attractive) and participants were instructed to rate attractiveness by circling the number they felt was appropriate. Overall mean attractiveness ratings of the attractive/unattractive images were calculated for both male and female participants, to determine if there was a gender difference in attractiveness ratings.

Successfulness Rating Measure. Participants also rated the photographic images in terms of successfulness with a second Likert-type scale designed for the purpose. The 'Successfulness' scale was a 7-point scale ranging from -3 (Unsuccessful) to +3 (Very Successful). This was used as an explicit measure of perceptions of successfulness for attractive v. unattractive facial images in order to compare results with implicit measures. Participants were instructed to indicate how successful they estimated the individual in the photo by circling the appropriate number on the scale. The Successfulness questionnaire data were calculated by averaging the total numerical scores for attractive/unattractive facial images for male and female participants.

I*mplicit Relational Assessment Procedure*. Before commencing, participants were provided with oral instructions by the researcher (appropriate instruction may be important to participants' successful completion thus reducing attrition rates, so researchers may wish to view exemplar instructions available online at http://irapresearch.files.wordpress.com/2011/11/irap-2012-experimenters-script1.pdf; the IRAP computer software program itself can be downloaded for free at: http://irapresearch.org/downloads-and-training/. The instructions informed the participants that on each trial one of two labels, "Successful" or "Unsuccessful," would appear at the top of the screen along with a photographic facial image presented in the centre of the screen. Participants were also told that the response options "Similar" and "Opposite" would appear at the bottom of the screen, and they were required to choose one of these options on each trial by pressing either the "D" or "K" Key. They were also told that the left-right positions of these response options would switch randomly from trial to trial. The instructions explained that the IRAP consisted of four different trial types (Attractive-Successful, Attractive-Unsuccessful, Unattractive-Successful, Unattractive-Unsuccessful) and illustrated examples of these were provided (see Figure 1).

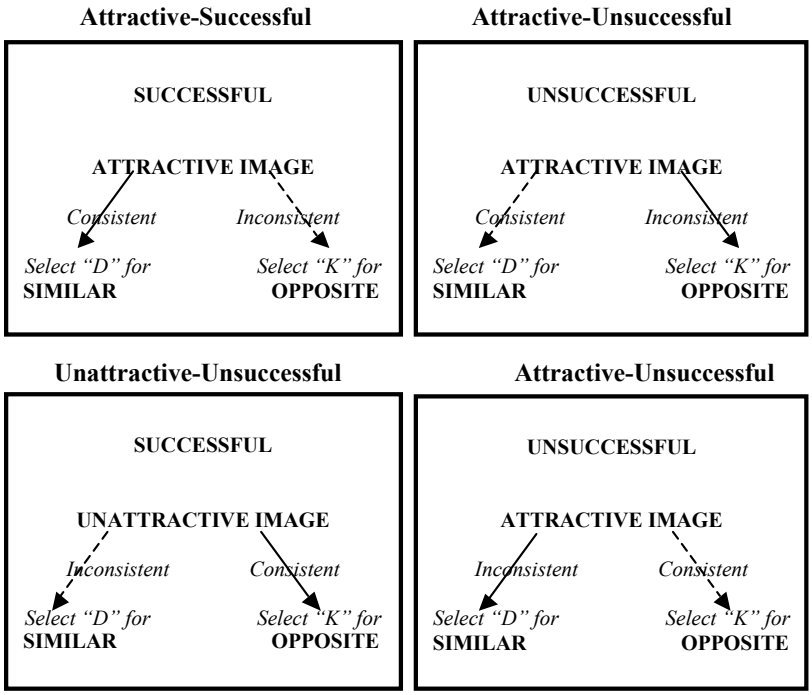

Figure 1. Representations of four IRAP trial-types. The attribute label stimulus ("Successful" or "Unsuccessful") appeared at the top of the screen while the target stimulus (a photo of either an attractive or an unattractive face) appeared in the centre of the screen. Response options (Similar/Opposite) appeared simultaneously on each trial at the bottom of the screen. The superimposed arrows and labels indicate what would be considered an attractiveness preference (Consistent) or an unattractiveness preference (Inconsistent) response for each trial type. (The boxes and arrows are for illustration purposes and did not appear on the screen).

Next, participants were informed that sometimes they would be required to respond in a way that was consistent with their beliefs and at other times they would be required to respond in a way that was inconsistent with their beliefs. Participants were assured that this was part of the experiment, and that it was important for them to respond as quickly and accurately as possible on all trials. Participants were also informed that correct responses (designated according to consistent or inconsistent trials) would allow them to progress to the next trial, but incorrect responses would produce a red X in the centre of the screen, which could only be removed by pressing the correct response key. The programmed feedback contingencies alternated from block to block between consistent and inconsistent trials. Prior to the commencement of each new block, each participant was informed via instructions onscreen that the previously correct and incorrect answers would be reversed. In addition to oral instructions, the IRAP program began with a set of instructions that described the task by illustrating the layout of the screen and explaining the response options.

The first block of the IRAP presented relations thought to be consistent with pro-attractiveness stereotyping (e.g., attractive-successful; unattractive-unsuccessful) potentially established preexperimentally for participants within the wider verbal community. The IRAP consisted

of a minimum of two practice blocks (one consistent and one inconsistent trial-block) prior to six test blocks (three consistent and three inconsistent trial-block); all trial-blocks comprised 24 trials. For the first two practice blocks, participants were informed that it was a practice block and that errors were expected. In order to progress to the test blocks participants were required to reach a standard of ≥ 75% accuracy in responding and a mean response time of ≤ 3000 milliseconds (ms) across the trial-block. The criteria set may be influential in the detection an IRAP effect, which may not be evident with longer response times. If participant responding failed to meet criteria during either of the two practice blocks, the standard they achieved and the required standard was presented as on-screen feedback. Each participant was allowed a maximum of 4 practice blocks to achieve the required practice criteria; if they failed to do so they were thanked for their participation, debriefed, and their data were subsequently discarded. Each target picture stimulus was presented twice in the presence of the each of the attributes "Successful" and "Unsuccessful" and the relational response options Similar and Opposite. The trials were presented quasi-randomly with the constraint that none of the four trial-types could be presented twice in succession. The left-right positioning of the two onscreen response options was alternated quasi-randomly, and they could not appear in the same position three times in succession.

RESULTS

Mean data were calculated for the 7-point Attractiveness Likert scale completed by male and female participants, and the overall ratings were subjected to a 2x2 mixed repeated measures analysis of variance (ANOVA) with gender (male versus female) as the between participant variable, and picture-type ("attractive" versus "unattractive") as the within-participants variable. The main effect for gender was not significant; $F(1, 45)= 3.78$, $p= 0.06$, partial Eta squared= .077. The effect for picture-type was significant; Wilks Lambda= .047, $F(1, 45)= 922.01$, $p= .0005$, partial Eta squared= .953. There was no significant interaction effect. In summary, both gender groups clearly discriminated between the pictures of "attractive" and "unattractive" facial images, with both male and female participants showing similar bias levels (see graphic data representation in Figure 2). Male and female participant ratings of attractiveness for photographic images were found to accord with "attractive" and "unattractive" categories predesignated by internet voters at the source (websites) from which the composite facial images were drawn.

Mean scores were calculated for participants' explicit ratings of successfulness for attractive versus unattractive facial images portrayed with the Likert questionnaire, and these data were subjected to a 2×2 mixed repeated measures ANOVA with gender as the between-participant variable and picture type (attractive/unattractive) as the repeated measure. The main effect for gender was significant; $F(1, 45)= 6.56$, $p= 0.04$, partial Eta squared= .172. The effect for picture-type was also significant, Wilks Lambda= .34, $F= (1, 45)= 252.75$, $p= .0005$, partial Eta squared= .172. There was no significant interaction effect; Wilks Lambda= .981, $F(1, 45)= 4.00$, $p= 0.51$, partial Eta squared= .082. The overall mean ratings showed that both male and female participants rated attractive facial images as more successful than unattractive facial images and findings were statistically significant (Figure 3).

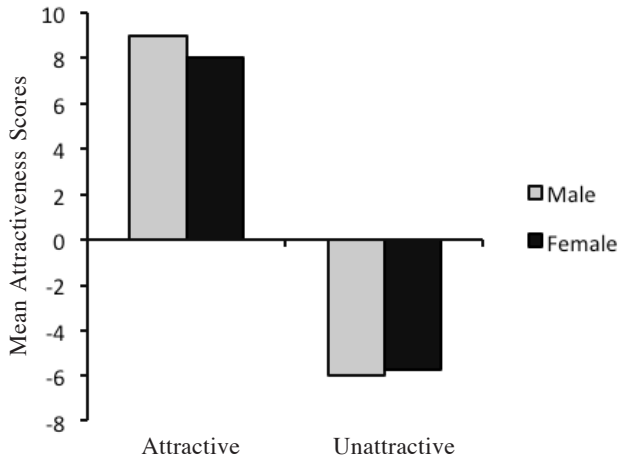

Figure 2. Male and female participants' mean ratings of attractiveness of target images.

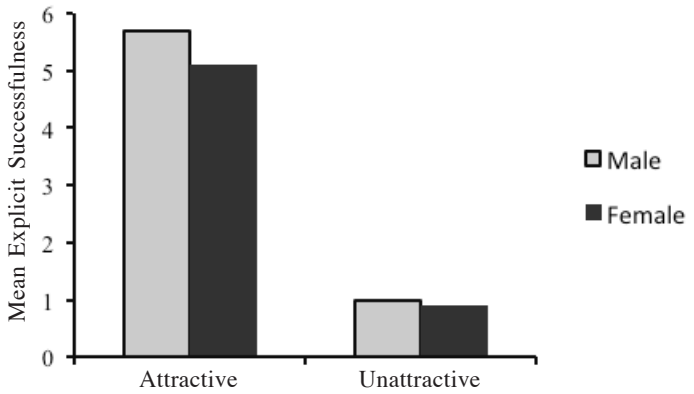

Figure 3. Male and female participants' mean ratings of successfulness of attractive versus unattractive target images.

The primary datum examined using the IRAP was response latency, defined as the time in milliseconds (ms) that elapsed between the onset of the onscreen stimulus presentation on each trial and the first correct response emitted by the participant. Due to failure to reach performance criteria, data from 3 participants were emitted from the data set, therefore response latency data from 47 participants were analyzed. The response latency data for each participant were transformed into difference or *D* scores (see *D*-algorithm; Greenwald, Nosek, & Banaji, 2003). An adaptation of the *D*-algorithm was utilized (detailed description of the D-algorithm adaptation is available in Barnes-Holmes *et al*., 2009; Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010). Transforming the data to *D*-scores may counter effects of individual variability in participant responding related to nonrelevant factors such as age or motor and cognitive skills (see Barnes-Holmes, *et al*., 2010). The steps involved in calculating the *D*-IRAP

scores were as follows: (1) Only response-latency data from the six test blocks were used; (2) latencies above 3000 ms were removed from the dataset; (3) all data for a participant were removed if he or she produced more than 10% of test-block trials with mean response latencies greater than 3000ms; (4) twelve standard deviations for the four trial types were computed: four for the four response latencies from test-blocks 1 and 2, four from the latencies from test blocks 3 and 4, and a further four from test blocks 5 and 6; (5) 24 mean latencies were then calculated for the four trial-types in each of the test-blocks; (6) difference scores were calculated for each of the four-trial types for each pair of test blocks by subtracting the mean latency of the consistent (pro-attractive) test-blocks from the mean latency of the corresponding inconsistent (pro-unattractive) test blocks; (7) each difference score was then divided by its corresponding standard deviation from step 4, yielding 12 *D*-IRAP scores; one score for each of four trial-types for each pair of test blocks; (8) four overall trial-type *D*-IRAP scores were calculated by averaging the three scores for each trial-type across the three pairs of test blocks; (9) two *D*-IRAP scores, one for Attractive and one for Unattractive were then calculated by averaging the Attractive and Unattractive trial type scores; (10) an overall *D*-IRAP score was calculated by averaging all 12 trial-type *D*-IRAP scores from step 7.

The *D*-IRAP data were then subjected to statistical analysis via a 2×4 mixed repeated measures ANOVA. The four IRAP trial-types were the within-participant independent variable (IV), and gender was the between participant IV; *D*-IRAP scores were the dependent variable (DV). The analysis revealed a significant main effect for trial-type; Wilks Lambda= .35, $F(3, 43)= 26.32$, $p <.0005$, partial Eta squared= .65, and a significant main effect for gender (sex): $F(1, 45)= 76.63$, $p <.0005$, partial Eta squared= .63. There was also a significant interaction effect; Wilks Lambda= .80, $F(3, 43)= 3.62$, $p <.02$, partial Eta squared= .13. Eight one-sample t-tests were conducted to determine if the *D*-IRAP scores for each of the four trial-types differed significantly from 0 for male and female participants. The resultant data are presented in Tables 1 and 2, respectively. For both male and female participants, the difference from 0 was found to be significant for all four trial types: Attractive-Successful-Similar (male participants: $M=1.6$, $SD= .18$, $t(23)= 42.9$, $p < .0005$; female participants $M=1.2$, $SD=.42$, $t(22)= 14.0$, $p <.0005$; Unattractive-Successful-Opposite (male participants: $M=1.3$, $SD= .29$, $t(23)= 22.0$, $p <.0005$; female participants $M=.67$, $SD=.55$, $t(22)= 5.8$, $p <.0005$; Unattractive-Unsuccessful-Opposite (male participants: $M=.1.3$, $SD=.36$, $t(23)= 17.9$, $p <.0005$; female participants: $M=.94$, $SD= .35$, $t(22)= 12.9$, $p <.0005$; Attractive-Unsuccessful-Similar (male participants: $M=1.5$, $SD= .15$, $t(23)= 48.0$, $p <.0005$; female participants $M=.80$, $SD=.42$, $t(22)= 9.2$, $p <.0005$. In summary, both male and female participants' responding showed a pro-attractive and anti-unattractive bias that were statistically significant (attractiveness bias interpreted via latency data indicating speed of affirmations of successfulness of attractive v. unattractive facial images); the magnitude of attractiveness bias for male participants was shown to be stronger across all four IRAP trial-types, and these effects were statistically significant (see graphic representation, Figure 4).

In order to assess the internal consistency of the IRAP, split-half reliability scores were calculated, one for consistent and inconsistent trial-types. Thus, two scores were calculated, one for odd trials and the second for even trials. These two scores

Table 1. Results of planned comparison *t*-tests for 4 IRAP trial-types for male participants.

IRAP Trial-type	Mean	*SD*	*t*	*p* value
Attractive-Successful-Similar	1.6	.18	*t*(23)= 42.9	.0005
Unattractive-Successful-Opposite	1.3	.29	*t*(23)= 22.0	.0005
Unattractive-Unsuccessful-Opposite	1.3	.36	*t*(23)= 17.9	.0005
Attractive-Unsuccessful-Similar	1.5	.15	*t*(23)= 48	.0005

Table 2. Results of planned comparison *t*-tests with 4 IRAP trial-types for female participants.

IRAP Trial-type	Mean	*SD*	*t*	*p* value
Attractive-Successful-Similar	1.2	.42	*t*(22)= 14.0	.0005
Unattractive-Successful-Opposite	.67	.55	*t*(22)= 5.8	.0005
Unattractive-Unsuccessful-Opposite	.94	.35	*t*(22)= 12.9	.0005
Attractive-Unsuccessful-Similar	.80	.42	*t*(22)= 9.2	.0005

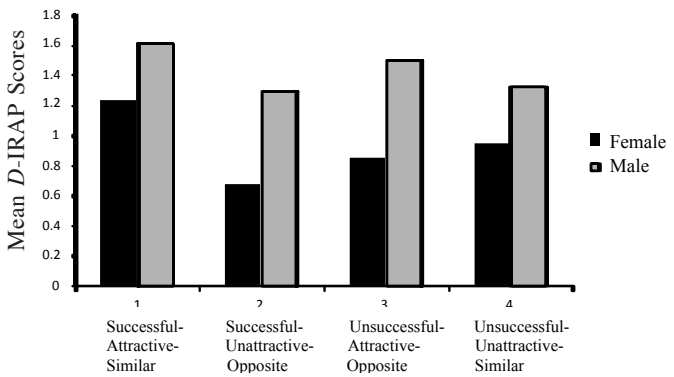

Figure 4. Mean D-IRAP Scores for four trial-types for male and female participant.

were calculated in the same way as described for the *D*-IRAP scores, except that the algorithm was applied separately to all odd trials and to all even trials. The split-half correlation between odd and even D-IRAP scores calculated across all participants indicated significant internal consistency (r= .616, n= 47, p <.0004).

Modest positive correlations were detected between the successfulness ratings for attractive facial images and the *D*-IRAP scores for three out of four trial-types and these were statistically significant (p <.01; see Table 3). There was no significant correlation found between the fourth D-IRAP trial-type data (Unattractive-Unsuccessful-Similar) and explicit data, and no significant correlations between IRAP data and explicit successfulness ratings for unattractive facial images.

Table 3. Correlations between explicit and implicit data.

IRAP Trial-type	Successfulness Scale	
	Attractive Image	Unattractive Image
Attractive-Successful-Similar	.405**	-.187
Unattractive-Successful-Opposite	.418**	-.161
Attractive-Unsuccessful-Opposite	.508**	-.22
Unattractive-Unsuccessful-Similar	.136	-.85

Note: ** $p < .01$

DISCUSSION

The current research was the first study to use the IRAP behavioral measure in the area of implicit attractiveness bias with adult participants (N= 47). Findings provided preliminary support for the IRAP as a sensitive measure of implicit beauty bias in both male and female participant groups, and the split-half reliability test indicated support for the internal consistency of the IRAP in the current context. Importantly, the IRAP program design facilitated specification of directionality of the attractiveness bias detected; statistical analyses of *D*-scores for male and female participants showed implicit pro-attractive bias and implicit anti-unattractive bias regarding successfulness of attractive v. unattractive photographic facial images that was statistically significant in both cases and this applied across all four IRAP trial-types. Interestingly, gender analysis of IRAP participant data showed a difference in that the magnitude of attractiveness bias was greater for male compared to female participants; the gender difference was statistically significant and again applied across all four IRAP trial-types. Statistical analysis of participants' explicit data rating the successfulness of attractive v. unattractive faces showed a beauty bias for both male and female participants that was statistically significant. Unlike the IRAP data, however, statistical analysis of the explicit data failed to indicate any gender difference. Correlational analysis indicated moderate positive correlations between ratings of successfulness of attractive faces and *D* (difference) scores on three of four IRAP trial-types.

The current study has extended the IRAP research literature by providing support for the successful use of the program in a novel area. Furthermore, the research in attractiveness bias has also been extended by the specification of directionality of the implicit attractiveness bias detected, given that a dearth of research on directionality of attractiveness bias has been noted (Langlois *et al*., 2006; Agthe *et al*., 2011). The IRAP could prove to be a very useful tool in the area of attractiveness bias; the program has gained support as a measure of implicit bias that can provide nuanced information about the directionality of stereotyping in many socially sensitive areas such as racial stereotyping (Barnes-Holmes, Murphy, *et al*., 2010; Power *et al*., 2009), implicit ageism (Cullen, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), implicit self-esteem (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), implicit deviant attitudes in child sex offenders (Dawson *et al*., 2009) body-weight bias (Roddy *et al*. 2010; Nolan *et al*. 2013) depression (Hussey & Barnes-Holmes, in press) and other areas (for a detailed review see Hughes & Barnes-Holmes, 2011). The capacity to provide directionality

information may be considered an advantage over other implicit measures in the area of attractiveness bias, however, it should be acknowledged that the IRAP is a relatively recent development and thus more extensive studies are required to support the utility of the IRAP in this area. Currently, however, it can be safely stated that the IRAP may be considered an important behavioral measurement tool that has begun to successfully penetrate areas previously considered to be the realm of associative or cognitive researchers rather than behavioral researchers.

The IRAP findings of a more pronounced attractiveness bias for male participants are broadly consistent with research findings of a stronger attractiveness bias in male participants in the context of partner selection for sex and marriage; for example, men show favorable bias toward women who are physically attractive, whereas women show bias favorable toward men with higher socioeconomic status (Berscheid & Walster, 1974; Buss & Barnes, 1986; Feingold, 1990; Bereczkei, 1997). Attractiveness bias research literature in the more general social domain has not reported a greater propensity toward beauty bias in male participants, however, more nuanced future investigations of contextual influences in a variety of domains may be illuminative (Langlois *et al*., 2006).

The current findings of gender disparity shown via participants' implicit data but not explicit data, indicating a stronger attractiveness bias effect for male participants, appears contrary to Feingold's (1990) caveat that women may be more inclined than men to conceal attractiveness bias but of course further research is needed to clarify the issue. To the extent, that the explicit data did not indicate a gender disparity, it appears that the IRAP was the more sensitive measure, but again more research would be required to support the apparent advantage. As stated previously, discovery of consistent unidirectional gender disparity in the extent of attractiveness bias in important domains could be relevant to practical purposes aimed at promoting fairness and egalitarianism in human situations. Findings of gender differences in attractiveness bias could be used to guide real world counter measures; for instance if one or other gender was shown to be consistently more prone to stronger attractiveness bias in particular domains, an equal gender balance may not be optimal in important evaluator panels (e.g., jury panels, job interview panels). The issue is perhaps further complicated by the influences exerted when evaluators and evaluatees are same-sex or different sex (Agthe *et al*., 2011). Expanded knowledge about gender differences in prejudicial tendencies may ultimately provide useful information and caveats regarding evaluator situations, but is in any event worthy of interest at the level of basic research.

The current findings of modest positive correlations between participants' explicit ratings of successfulness of attractive faces and *D*-scores for three of the four IRAP trial-types (it is currently unclear as to why correlations should occur with three and not four IRAP trial-types, but this may be a spurious finding specific to the current study) might be interpreted tentatively as follows: The correlations are perhaps unsurprising given the robust attractiveness bias shown in previous explicit research literature which suggests that participants do not typically conceal this bias; that the correlations were modest rather than strong could be related to the explicit/implicit types of behavioral responding, or to the different measurement systems used. The debate on whether and when researchers should expect to find explicit-implicit correlations is ongoing within

the field of implicit cognition. For instance, Greenwald *et al*. (2003) purported that the greater the concordance between explicit and implicit measures, the more valid the measurement of the topic is. Conversely, it has been argued that the degree of correlation may depend upon participants' motivation and opportunity to consciously deliberate (Dovidio & Fazio, 1992; Barnes-Holmes *et al*., 2010). It appears from the data that participants in the current research were not motivated to entirely conceal their attractiveness bias in the explicit rating scale, but the gender disparity shown in implicit but not explicit data might suggest that male participants failed to reveal the full extent of their attractiveness bias in explicit ratings. Such possibilities would require much more extensive investigation, however, before any interpretation could be made with anything like science-based conviction.

A limitation in the current study was that an analysis of influence of gender of target facial images on attractiveness bias was not undertaken. Future research with the IRAP program, however, could readily adapt the IRAP procedure to assess influence of gender of target individuals (see Nolan *et al*., 2013), and this would also facilitate analysis of same-sex and different-sex dyads of evaluators and target individuals. Another possible limitation was that the concept of "successfulness" was not clearly defined in the current research. The aim was to capture or encompass perceptions of general successfulness including career successfulness; however, in keeping the concept broad in scope it is possible that participants may have interpreted it differently. The small sample size limits the generaliseability of the findings, but adds credence regarding the findings of statistical significance; it is more difficult to detect significant differences with small groups whereas very large samples are unlikely not to indicate a difference that is statistically significant. Notwithstanding possible limitations, the current findings suggest that the IRAP may be a sensitive tool with satisfactory internal consistency suitable for use in further explorations of implicit attractiveness bias, and especially for elucidating directionality of participants' attractiveness bias in domains other than the social domain. The sample attributes utilized in the current IRAP procedure ("Successful" v. "Unsuccessful") could be readily modified to "Conscientious" v. "Lazy", "Intelligent" v. "Stupid", or a host of other representative bipolar, or less polarized, attitudes (BIRRS) in the area of attractiveness bias. To facilitate a more nuanced investigation of attractiveness bias, future IRAP investigations might consider reducing the response latency criterion from 3000ms in the current research to 2000ms, because research has shown that IRAP effects are more readily detected with a shorter speed criterion. For example, increased racial stereotyping effects were evident in an IRAP test with a response latency of 2000ms compared to 3000ms (Barnes-Holmes *et al*., 2010). Such findings are perhaps not surprising and provide support for the influence of heuristics in biased automatic responding, in that when participants were required to respond more rapidly, prejudice was increased.

In conclusion, it has been noted that the research into stereotyping and prejudice contributes to a greater understanding of these complex human issues, and the research in implicit attitudes has been particularly enlightening regarding the insidiousness of the stereotypical judgments we make when responding with speed under time pressure (Hardin & Banaji, 2013). The IRAP is an implicit measure that may facilitate further

and more nuanced investigation into contextual influences on attractiveness bias, and perhaps of more practical importance, into the malleability of attractiveness bias via contextual manipulations (Hughes & Barnes-Holmes, 2011).

References

Adams GR (1977). Physical attractiveness research: Toward a developmental social psychology of beauty. *Human Development, 20,* 217-239.

Agthe M, Spörrle M, & Maner JK (2011). Don't hate me because I'm beautiful: Anti-attractiveness bias in organizational evaluation and decision making. *Journal of Experimental Social Psychology*, *46*, 1151-1154.

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

Barnes-Holmes D, Barnes-Holmes Y, Power P, Hayden E, Milne R, & Stewart I (2006). Do You Really Know What You Believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a Direct Measure of Implicit Beliefs. *The Irish Psychologist, 32,* 169-177.

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, Murtagh L, Barnes-Holmes Y, & Stewart I (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *The Psychological Record, 60,* 287-306.

Barnes-Holmes D, Murphy A, Barnes-Holmes Y, & Stewart I (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables among vegetarians and meat-eaters. *The Psychological Record, 60,* 207-306.

Barnes-Holmes D, Waldron D, Barnes-Holmes Y, & Stewart I (2009). Testing the validity of the Implicit Relational Assessment Procedure (IRAP) and the Implicit Association Test (IAT): Measuring attitudes towards Dublin and country life in Ireland. *The Psychological Record, 59,* 389-406.

Benson PL, Karabenick SA, & Lerner RM (1976). Pretty please: The effects of physical attractiveness, race, and sex on receiving help. *Journal of Experimental Social Psychology, 12,* 409-415.

Berscheid E & Walster E (1974). Physical Attractiveness. *Advances in Experimental Social Psychology, 7,* 157-152.

Bosson JK, Swann WB Jr., & Pennebaker JW (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79,* 631-643.

Buhlmann U, Teachman BA, & Kathmann N (2010). Evaluating implicit attractiveness beliefs in body dysmorphic disorder using the Go/No-go Association Task. *Journal of Behavior Therapy and Experimental Psychiatry, 1,* 1-28.

Buss DM & Barnes ML (1986). Preferences in Human Mate Selection. *Journal of Personality and Social Psychology, 50,* 559-570

Chaiken S (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology, 37,* 1387-1397.

Cheung PP & Leung K (1988). Effects of performance information and physical attractiveness on managerial decisions about promotion. *The Journal of Social Psychology, 128,* 791-801.

Crosby F, Bromley S, & Saxe L (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin, 87,* 546-563.

Dawson DL, Barnes-Holmes D, Gresswell DM, Hart AJP, & Gore NJ (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. Sexual Abuse: *A Journal of Research and Treatment, 21*, 57-75.

De Houwer J (2002). The implicit association test as a tool for studying dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behavior Therapy & Experimental Psychiatry, 33,* 115-133.

De Houwer J (2008). Comparing measures of attitudes at the functional and procedural level: Analysis and implications. In R Petty, RH Fazio & P Brionol (Eds.), *Attitudes: Insights from the new implicit measures* (pp 361-390). Erlbaum.

Dickey-Bryant L, Lautenschlager GJ, Mendoza JL, & Abrahams N (1986). Facial attractiveness and relation to occupational success. *Journal of Applied Psychology, 71,* 16-19.

Dion KK, Berscheid E, & Walster E (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24,* 285-290.

Dovidio JF, & Fazio RH (1992). New technologies for the direct and indirect assessment of attitudes. *Questions about Questions: Inquiries into the Cognitive Bases of Surveys, 44,* 204-237.

Dovidio JF, Kawakami K, Johnson C, Johnson B, & Howard A (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33*, 510-540.

Eagly AH, Ashmore, RD, Makhijani MG, & Longo LC (1991). What is beautiful is good, but…: A Meta analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109-128.

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

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

Feingold A (1990). Gender differences in effects of physical attractiveness on romantic attraction: A comparison across five research paradigms. *Journal of Personality and Social Psychology, 59,* 981-993.

Feingold A (1992). Good-looking people are not what we think. *Psychological Bulletin, 111,* 304-341.

Greenwald AG, McGhee DE, & Schwartz JLK (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1146-1480.

Greenwald AG, Banaji MR, Rudman LA, Farnham SD, Nosek BA, & Mellott DS (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3-25.

Greenwald AG, Nosek BA, & Banaji MR (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216.

Greenwald AG, Poehlman TA, Uhlmann E, & Banaji MR (2009). Understanding and using the Implicit Association Test: III. MEta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17-41.

Griffin AM & Langlois JH (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad? *Social Cognition, 24,* 212-246.

Halberstadt J & Rhodes G (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin and Review*, *10,* 149-156.

Hardin CD & Banaji MR (2013). The nature of implicit prejudice: Implications for personal and public policy. In E Shafir (Ed.), *The behavioral foundations of public policy* (pp. 13-32). New Jersey: Princeton University Press.

Hayes SC, Barnes-Holmes D, & Roche B (2001). *Relational Frame Theory: A post-Skinnerian account*

of language and cognition. New York: Plenum.
Heilman ME & Stopeck MH (1985). Being attractive, advantage or disadvantage? Performance-based evaluations and recommended personnel actions as a function of appearance, sex, and job type. *Organizational Behavior and Human Decision Processes, 35,* 202-215.
Hosoda M, Stone-Romero EF, & Coat G (2003). The effects of physical attractiveness on job- related outcomes: A meta-analysis of experimental studies. *Personnel Psychology, 56,* 431-462.
Hussey I, & Barnes-Holmes D (in press). The IRAP as a measure of implicit depression and the role experiential avoidance. *Cognitive and Behavioral Practice.*
Langlois H, Kalakanis L, Rubenstein AJ, Larson A, Hallam M, & Smoot M (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*, 390-423.
Mazzella R & Feingold A (2006). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology, 24,* 1315-1338.
Marlowe CM, Schneider SL, & Nelson CE (1996). Gender and attractiveness biases in hiring decisions: are more experienced managers less biased. *Journal of Applied Psychology, 81,* 11-21.
McKenna IM, 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,* 123-138.
Nolan J, Murphy C, & Barnes-Holmes D (2013). Implicit perceptions of intelligence in slim and overweight individuals: Effects of gender of participants and target stimuli. *The Psychological Record*, *63,* 467-488.
Power P, Barnes-Holmes D, Barnes-Holmes Y, & Stewart I (2009). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record, 59,* 621-640.
Roddy S, Stewart I & Barnes-Holmes D (2010). Anti-fat, pro-slim, or both? Using two reaction time based measures to assess implicit attitudes to the slim and overweight. *Journal of Health Psychology, 15,* 416-425.
Rudman LA & Glick P (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57,* 743-762.
Rudman LA, Greenwald AG, Mellott DS, & Schwartz JLK (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition*, *17*, 437-446.
Ryan C & Costa-Giomi E (2004). Attractiveness bias in the evaluation of young pianists' performances. *Journal of Research in Music Education, 52,* 141-154.
Spalding LR & Hardin CD (1999). Unconscious unease and self-handicapping: Behavioral consequences of individual differences in implicit and explicit self-esteem. *Psychological Science, 10,* 535-539.
Vahey NA, Barnes-Holmes D, Barnes-Holmes Y, & Stewart I (2009). A first test of Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record, 59*, 371-388.
Van Lange PAM (2006). *Bridging social psychology: Benefits of transdisciplinary approaches*. New Jersey, New York: Routledge.
Wapnick J, Darrow AA, & Kovacs J (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education, 45,* 470-47
Wilson TD, Lindsey S, & Schooler TY (2000). A model of dual attitudes. *Psychological Review, 107,* 101-126.
Yamamoto R, Ariely D, Chi W, Langleben DD, & Elman I (2009). *PLoS ONE, 4,* 6, e202 doi:10.1371/journal.pone.0006042

Received, July 30, 2013
Final Acceptance, June 20, 2014