Journal of Contextual Behavioral Science 4 (2015) 292–299

Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

journal homepage: www.elsevier.com/locate/jcbs

Empirical Research

The Implicit Relational Assessment procedure (IRAP) and attractiveness bias

Carol Murphy, Tara Hussey, Dermot Barnes-Holmes, Michelle E. Kelly*

The Department of Psychology, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland

ARTICLE INFO

Article history:
Received 18 April 2014
Received in revised form
20 July 2015
Accepted 17 August 2015

ABSTRACT

The Implicit Relational Assessment Procedure (IRAP) was used in the area of attractiveness bias and attributions of successfulness. Alternate IRAP trial-blocks required participants to affirm consistent (attractive–successful) and inconsistent (unattractive–successful) relations; shorter mean response latencies across consistent trial-blocks were interpreted as implicit attractiveness.stereotyping. Participants also completed a rating scale for successfulness of attractive versus unattractive individuals. Both implicit and explicit (rating data) data showed statistically significant attractiveness bias for male and female participants. Directionality of bias was analyzed via the IRAP 4 trial-type methodology to determine if it was pro-attractive or anti-unattractive, or if bias was evident.in both directions, or if no bias was shown. For both gender groups, bias was shown to be proattractive.and not antiunattractive. Findings are discussed with regard to a comprehensive account of attractiveness bias, directionality, and contextual influences.

1. Introduction

The term “attractiveness bias” refers to an apparently commonly held belief that physically attractive people have more socially desirable personality traits compared to physically unattractive people. This has been well documented in research literature using explicit measures (Dion, Berscheid, & Walster, 1972; Eagly, Ashmore, Makhijani, & Longo, 1991; Ryan & Costa-Giomi, 2004; Wapnick, Darrow, & Kovacs, 1997). Attractive people are judged as having more positive characteristics, as being more intelligent, more likely to attempt a college education, more desirable as potential romantic partners, more sensitive, kind, interesting and sociable compared to unattractive individuals, and less likely to have psychological difficulties (Jones, Hansson, & Phillips, 1978; Langlois et al., 2000; Puleo, 2006). The phrase “What is Beautiful is Good” (WBIG) was coined by Dion, Berscheidet, and Walster (1972) in reference to stereotypical judgments about attractiveness; physical beauty has been said to create a “halo” effect of a positive evaluation that leads observers to be positively biased in their judgments of attractive individuals (Lucker, Beane, & Helmreich, 1981).

Despite evidence to support findings of attractiveness bias (Eagly et al., 1991), a small number of studies show that contextual factors might result in conflicting outcomes. A contrary “Beauty is Beastly” effect (Heilman & Stopeck, 1985; Hosoda, Stone-Romero, & Coats, 2003) indicated that attractiveness for women was disadvantageous in certain contexts, for example, when being considered for jobs traditionally viewed as masculine. In organizational research, Agthe, Spörrle, and Maner (2010) reported that attractiveness of an individual produced a positive bias in different-sex evaluators, but a negative bias in same-sex evaluators. It appears therefore that the hypothesis on attractiveness bias may not be straightforward and that contextual influences should be explored prior to conclusions being drawn in either direction.

A shortcoming in attractiveness bias research is that the bulk of the literature examines attractiveness bias in the area of social attributes and comprises self-report or questionnaire data (Griffin & Langlois, 2006). Only a small number of studies have used measures of participant behavior, such as recording reaction times on computer generated tasks including a modified Stroop (Van Leeuwen and Macrae, 2004); and on the Go/No Go Association task (e.g. Buhlmann, Teachman, & Kathmann, 2010); to demonstrate implicit positive bias toward attractive compared to unattractive individuals. The accuracy of data from self-report measures in psychological research is considered somewhat dubious as participants can readily provide fake data for reasons related to self-presentation or impression management (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Nosek, Greenwald, & Banaji, 2007). This may be particularly relevant in research that examines bias in socially sensitive topics (e.g., prejudice toward minority social groups) compared to research on topics such as consumer preferences and

* Correspondence to: 2nd Floor John Hume Building, Maynooth University, Maynooth, Co. Kildare, Ireland. Fax: +353 86 1527917.

http://dx.doi.org/10.1016/j.jcbs.2015.08.001
2212-1447/

clinical phenomena (Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

Furthermore, Feingold (1990) speculated that the research data on attractiveness bias may be contaminated by greater influence of self-presentational concerns in women. Specifically, preferential ratings of attractive versus unattractive individuals may be perceived as unjust or "shallow", and thus participants, particularly women, might be influenced to conceal this type of bias (Feingold, 1990). Whether or not Feingold's perception of gender differences in self-presentational concerns were correct, an enormous disparity in numbers of women and men undergoing cosmetic procedures suggest that gender differences are also present in concerns about physical attractiveness of self. The Beacon Face and Dermatology Clinic reports over 10,000 facelift procedures conducted in Ireland in 2011, the majority on women (www.beaconfaceanddermatology.ie). The gender disparity in uptake of cosmetic procedures is consistent with information from other sources (e.g., statistics from the British Association of Aesthetic Plastic Surgeons show that in the UK, women underwent 90.5% of all cosmetic procedures in 2013; www.baaps.org.uk. Also, the American Society of Plastic Surgery reported women underwent 91% of all cosmetic procedures in 2012; www.plasticsurgery.org). Thus, the issue of participant gender effects would seem important in research on attractiveness bias.

In seeking to address issues of bias with explicit self-report measures, researchers have recently designed "implicit" measures of social bias focused on behavioral responses as an alternative, or complementary, to self-report data. Measures of participant response latencies (speed of responding) are used as behavioral indicators of prejudice, or lack of it, in research examining implicit bias. For example, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is an implicit measure that examines associations between two concepts. In the IAT, faster responding when categorizing a particular concept (e.g. "thin") with positive attributions (e.g. smart, friendly, vibrant) compared to when categorizing a contrasting concept (e.g. "fat") with the same positive attributes, may be deemed evidence of a pro-thin bias. Results from IAT research support existing research on attractiveness bias (Van Leeuwen and Macrae, 2004; Buhlmann et al., 2009). Specifically, response latencies were faster when photographs of attractive individuals were paired with positive words (e.g. kind, intelligent) and photographs of unattractive individuals were paired with negative words (e.g. stupid, cruel) compared to trials when attractive–negative and unattractive–positive words were paired together (Buhlmann et al., 2009).

Despite these advances in the research literature on attractiveness bias, a shortcoming is that the issue of directionality has not yet been adequately addressed. That is, the question of how, or to what extent attractiveness is good or unattractiveness is bad has been subjected to limited scrutiny. Recent findings suggest that it may be more often the case that unattractiveness is "bad" rather than that attractiveness is "good", though stereotype directionality varied depending upon domain of judgment (Griffin & Langlois, 2006). Specifically, anti-unattractive bias only was evident in judgments of altruism and intelligence, whereas both a pro-attractive and anti-unattractive bias was shown in judgments of sociability. This finding highlights the need for further research to explain the directionality of bias. Understanding the extent to which attractiveness is advantageous to the individual and/or unattractiveness is disadvantageous (Eagly et al., 1991; Langlois et al., 2000), and examining the dynamics of attractiveness bias under contextual influences (including gender of evaluators) and in domains other than sociability might allow researchers to provide a more coherent and comprehensive account of this phenomenon.

A recent development in the measurement of implicit bias could facilitate a more thorough examination of attractiveness bias including directionality. The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009) has been adapted from the previously referred to IAT associationist test procedure (Greenwald, McGhee, & Schwartz, 1998). An important conceptual difference, however, is that the IRAP is non-associationist, and instead is proposed as a behavioral measure of pre-experimental verbal relations (software and sample instructions available to researchers upon request). The conceptual foundation for the IRAP program is 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 speed and accuracy of relational responding (e.g. responding to one stimulus in terms of another) is thought to be an index of the verbal histories that participants have been exposed to, as opposed to representations of mental associations, as conceived in the IAT.

The IRAP, like many implicit tests including the IAT, the Extrinsic Affective Simon Test (EAST; De Houwer, 2003), and the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001), uses participants' response latencies to infer bias; for example, participants may be asked to affirm verbal relations such as thin-positive-true and fat-positive-true across alternate trial-blocks. The former relations are thought to be "consistent" with verbal relations prevalent in the wider social community, and the latter relations are thought "inconsistent" with that relational pattern. Faster responding (shorter response latencies) for the former would be interpreted as implicit bias. The term "implicit" as used in IRAP research is not intended as a mentalistic description; the preferred behavioral term for the type of automatic or impulsive responding captured via implicit measures is "brief immediate relational responding" (BIRRS; Barnes-Homes, Barnes-Holmes, Stewart & Boles, 2010). The more intentional or deliberative responding thought to be involved in self-report measures is described behaviorally as "relational elaboration and coherence" or REC. Under time pressure to respond rapidly, a participant's BIRR may affirm *men-smart* more rapidly than *women-smart* suggesting agreement that men are smarter than women; 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 the participant may fail to report agreement that men are smarter than women (for an expanded discussion see Hughes, Barnes-Holmes, & Vahey, 2012).

The IRAP program involves onscreen presentation of one of two sample stimuli that appear 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 affirm thin-positive relations and during inconsistent trials to confirm fat-positive relations. 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 in four trial-types means that the IRAP can provide information regarding directionality of any bias detected (e.g., thin-positive, fat-negative, fat-positive, thin-negative); to continue with the example, whether there is a pro-thin and anti-fat bias, or an anti-fat but no pro-thin bias, or a combination of bias in both directions.

The IRAP has been effective in detecting participants' implicit bias in research that has been replicated across a number of domains, including social stereotyping (Barnes-Holmes, Murphy,

Barnes-Holmes, & Stewart, 2010; Power et al., 2009); age-related attitudes (Cullen, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009); implicit self-esteem (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009); and implicit sexual beliefs (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009). Also, preliminary findings have provided support for the IRAP in terms of reliability (Power et al., 2009) and validity (Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009), and have indicated that the IRAP was not amenable to "fake-ability" when participants were instructed to control responding (Hussey & Barnes-Holmes, 2012; McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007). The IRAP has been shown also to have predictive validity toward behavior (see Dawson et al., 2009; Roddy et al., 2010). Taken together, this evidence suggests that the IRAP might be an appropriate measure to address the shortcomings noted in extant attractiveness bias research literature.

The current study sought to use both the IRAP (implicit measure) and a Likert-type questionnaire (explicit measure) to examine attractiveness bias in the context of attributes of successfulness, in order to determine how successfulness (or professional ability) might be affected by attractiveness bias. We also undertook a gender analysis with participants' IRAP data and explicit data. Findings of consistent gender influence in attractiveness bias could have implications for the design of practical strategies to counter negative effects of attractiveness bias in important real world situations. For example, if one or other gender group showed consistently greater vulnerability to attractiveness bias across domains, then equal numbers of male–female evaluators may not be optimum for interview or jury panels. This study focused on facial attractiveness (photographic facial images of attractive versus unattractive individuals) because facial attractiveness was deemed of primary importance in an individual's overall attractiveness (Chung & Leung, 1988; Dickey-Bryant, Lautenschlager, Mendoza, & Abrahams, 1986; Heilman & Stopeck, 1985). The IRAP presented alternate consistent and inconsistent trial-blocks representing *Attractive–Successful* and *Unattractive–Successful* relations respectively, and automatically recorded response latencies. Shorter mean response latencies across consistent compared to inconsistent trial-blocks were interpreted in terms of implicit attractiveness bias.

The aims of the current study may be summarized as follows: (i) to test the IRAP as an effective behaviorally-based implicit measure in the domain of attractiveness bias. Specifically, response latency data were used to determine if college students ($N=30$; 15 male and 15 female) more rapidly affirmed that attractive versus unattractive facial images were deemed successful; (ii) to use the IRAP methodology to examine directionality of any IRAP bias detected; i.e., was there a positive bias toward attractiveness and a negative (or neutral) bias toward unattractiveness; (iii) to explore contextual influences on attractiveness bias; that is, different-sex vs. same-sex evaluators (e.g. Agthe et al., 2010); (iv) to examine participants' Belief About Appearance Scale (BAAS; Spangler, 1999) scores in rating the importance of their own appearance, and to explore correlations between these explicit scores and implicit IRAP data; and (v) to analyze both implicit and explicit data for influence of participant gender on attractiveness bias.

2. Method

2.1. Participants

Thirty participants (15 men and 15 women) with an age range of 19–26 years ($M=21.45$) took part in the experiment. An additional three participants volunteered for the study but were excluded due to failure to meet the predetermined performance criteria of more than 75% accuracy on at least two of the six practice blocks of the IRAP (see below for more detail). Participants were psychology undergraduate students at the National University of Ireland Maynooth. Although this was a sample of convenience, data from college students may be representative of young middle-class male and female adults who may at some point form part of employment selection panels where beauty bias effects may be important. All participants were Caucasian, fluent English speakers and had normal or corrected-to-normal vision. No inducements of any kind were offered for participation. 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.

2.2. Apparatus and stimuli

2.2.1. Explicit measure

Beliefs About Appearance Scale questionnaire (BAAS; Spangler, 1999) is a 20-item, self-report scale that rates the degree of participant affirmation of beliefs about the impact of their own appearance in four areas (subscales): "interpersonal relationships", "personal achievement", "self-perception", and "emotional well-being". Questionnaire items included: "My life will be more exciting or rewarding if I look good" and "The opinion others have of me is based on my appearance". The degree of agreement with statements about appearance in the four domains was rated on a 5-point scale ranging from 0 (not at all) to 4 (extremely). High scores on the BAAS may range from 23.34 ($SD=14.86$) to 30.57 ($SD=18.02$) (Spangler, 1999). Higher scores indicate greater endorsement of beliefs that positive self-perception, emotional well-being, and interpersonal and work success are dependent upon appearance. The BAAS 20-item scale possesses internal consistency (coefficient alpha.95) and test-retest reliability (r.83) (Spangler & Stice, 2001), and had a Cronbach's alpha of.90 (Clerkin & Teachman, 2009).

2.2.2. Implicit measure

The Implicit Relational Assessment Procedure (IRAP) was used to measure participants' implicit beliefs regarding attractiveness. The computer-based IRAP software (available upon request) was written in Microsoft Visual Basic 6 by Professor Dermot Barnes-Holmes. Participants completed the IRAP program on an Acer laptop with a Pentium 4 processor, standard keyboard and color monitor. Each IRAP trial presented one of two types of sample images; either an image of a person's face deemed attractive, or an image of a person's face deemed unattractive. Below each image a single positive or negative target word was presented (see Table 1), with the response options "True" and "False" presented below the target words (see Fig. 1). Positive and negative target words that reflected successfulness and unsuccessfulness were selected by a panel of six final year psychology students. The

Table 1
Sample, target, and relational stimuli presented in the IRAP.

Sample 1	Sample 2
Attractive image	Unattractive image
Response option 1	Response option 2
True	False
Consistent target (successful) 1	Consistent targets (unsuccessful) 2
Good	Bad
Intelligent	Stupid
Friendly	Unfriendly
Qualified	Useless
Out-going	Shy
Kind	Cruel

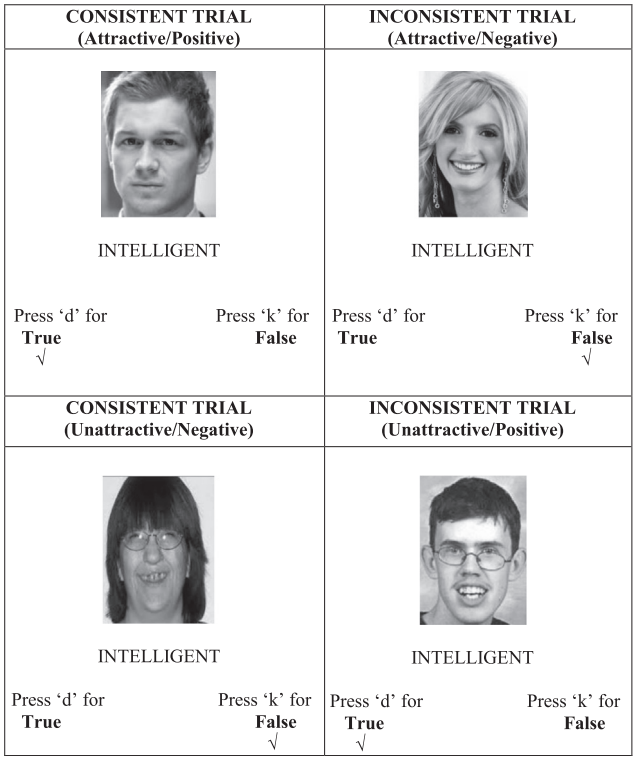

Fig. 1. An example of each of the four IRAP trial-types presented.

students were asked to think about people they deemed successful and to generate adjectives that reflected the characteristics of these individuals. The students agreed on six positive adjectives, which they agreed captured the characteristics of 'successfulness'; and subsequently generated negative adjectives based on opposites of each positive adjective.

The sample stimuli were digital color photographic stimuli 352 pixels tall, 373 pixels wide, and in 256-color format. Six of the pictures presented images of individuals that were deemed attractive (three attractive men and three attractive women) and six presented images of individuals that were deemed unattractive (three unattractive men and three unattractive women). The images of women were downloaded from a publicly accessible website (www.inyourface.ocregister.com), and were judged categorically in terms of "unattractive" or "attractive" by internet voters. The investigator selected the first, third and fifth images that appeared onscreen from both the "unattractive" and "attractive" categories for use in the current research. Images of men were not available from the former website, and were therefore downloaded from a different website (www.hotornot.com). This site also provided internet voter ratings of attractiveness. The investigator selected 20 images of men with high attractiveness ratings and 20 images with low attractiveness ratings. From these, the aforementioned panel of six final year psychology students selected three high-rated or "attractive" and three low-rated or "unattractive" images for use in the current study. Attractiveness was thus designated to stimuli on the basis of voter rating and the term is not intended in any absolute sense, however, for brevity purposes, images used will henceforth be referred to as "attractive/unattractive" without reminder of the designation process used. All images portrayed individuals between the age of 20 and 27 years; all had smiling facial expression, were looking at the camera, and were alone.

2.3. Procedure

The study was conducted in a quiet room at the Department of Psychology, National University of Ireland Maynooth. Due to unavailability of this room at certain periods, two of the 30 participants were required to complete the IRAP and explicit measures in a similar quiet room in the investigators residence. All instructions and stimuli remained constant for all participants regardless of location. Before beginning, each participant was given general information about the study and participation and each was presented with a consent form. Participants were free to discontinue their participation at any stage, and were assured that confidentiality would be upheld and their data would not be identifiable at an individual level.

2.3.1. Explicit measure (BAAS)

After completing the consent form, participants were presented with a BAAS questionnaire sheet (see Materials). Participants were instructed to respond to each item by ticking the response that most accurately represented their response to each one of 20 statements on a 5-point scale from 0 (not at all) to 4 (extremely). The investigator left the room while participants completed the questionnaire.

2.3.2. Implicit measure (IRAP)

After questionnaire completion, participants commenced the IRAP procedure. Participants were given oral instructions by the investigator (instructions may be important in reducing participant attrition rates; generic IRAP experimenter instructions used in this and other IRAP studies are available at http://irapresearch.files.wordpress.com/2011/11/irap-2012-experimenters-script1.pdf). Participants were seated at a desk in front of the Acer laptop. An automated set of detailed instructions were presented onscreen that participants read in their own time. The instructions described the format of the IRAP task and presented illustrations of the four trial-types that would be presented to them throughout the experiment (Fig. 1). Participants were informed that the IRAP would be presented in discrete trial-blocks and that the general rule for responding correctly would alternate across these blocks. They were also made aware that they would be required to respond in a way that was inconsistent with their beliefs. However, they were assured that this was part of the experiment, and that it was important for them to respond quickly and accurately for all trials, regardless of whether they truly agreed with the relations presented or not. Participants were told that the first two trial-blocks were for practice, that accuracy must be greater than 75% correct and that the median response latency across each trial-block must be below 2000 ms. Rapid accurate responding is required because the IRAP may fail to detect implicit bias if responding is too slow. Participants were told that if their performance did not meet these criteria practice-blocks could be repeated by pressing the space-bar.

Participants were asked to rest their left and right index fingers on the "*d*" and "*k*" keys on the keyboard. These keys corresponded to one of two onscreen response options (True/False), and participants had to press either "*d*" or "*k*" to select a response option. For each IRAP task, four stimuli were presented at the same time. The pictorial stimulus (either an attractive or unattractive person), appeared at the top of the screen, the target evaluative word appeared in the center of the screen, and the two response options, "True" and "False" appeared at the bottom left and right corners of the screen (see Fig. 1). The left–right position of the response options alternated quasi-randomly across trials, the same position occurring across no more than two successive trials. All four stimuli remained onscreen until the participant selected a response by pressing either the "*d*" or "*k*" key corresponding to "True" or

"False", respectively. If a participant responded correctly for a particular trial, all four stimuli were removed from the screen for a 400 ms interval before the next trial was displayed. If a participant's response was incorrect, a red X was presented directly under the target word. The X remained on-screen until the participant responded correctly. If a participant failed to emit a correct response within 2000 ms on any trial, the words "Too Slow" appeared on screen; the message cleared when a correct response was made. The IRAP consisted of a minimum two practice blocks and a fixed set of six test blocks, each containing twenty-four trials. Only data from the test blocks were used in subsequent analyzes. During each test block the twelve target words were presented in a quasi-random sequence, with each target presented twice, once with each type of sample picture.

The first block of practice trials required participants to affirm relations consistent with stereotypical verbal relations thought to operate in the wider social community (e.g., "What is Beautiful is Good" or WBIG stereotype). For example, during consistent trials if a sample picture with an attractive individual and a positive word appeared on-screen, a correct response involved pressing the response option "True". If an unattractive sample picture and a negative word appeared on-screen, a correct response involved pressing the response option "True". After 24 trials were complete, participants were presented with very brief feedback indicating the percentage of correct responses (accuracy) and median response time (latency/speed). Before the next trial-block commenced, instructions for the next 24 trials were displayed, stating that during the next practice trials the feedback contingencies would be reversed relative to the previous block. Participants pressed the space-bar to continue to the next block.

The second block of 24 practice trials required participants to affirm relations that were inconsistent with the WBIG stereotype. For example, if a picture with an attractive individual appeared on-screen with a positive word, a correct response involved pressing the response option of "False". If a sample picture with an unattractive individual appeared with a negative word appeared on-screen, a correct response involved pressing "False". Feedback was again displayed at the end of a trial-block that indicated the percentage of correct responses and median response time. If, after this second practice block, the percentage of correct responses was less than 75% or if the median response latency was greater than 2000 ms, an onscreen presentation screen reminded participants of the performance standard and the standard of responding they had achieved. Participants were allowed three attempts (a total of six practice blocks) to achieve the performance criteria. If they failed to meet the criteria across three pairs of practice blocks (messages reminding participants of the performance criteria were presented after each pair of practice blocks), the screen cleared and a message appeared indicating that the experiment was complete. These participants were thanked, debriefed and their data were discarded. This is because the detection of IRAP effects requires a high level of speed and accuracy in responding (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008).

If a participant achieved above 75% correct and a median response latency of less than 2000 ms across a pair of practice blocks, the program proceeded to the six test blocks. The procedure for the test blocks was similar to the sequence of practice blocks except that before each test block the following message appeared on screen; "This is a test—go fast. Making a few errors is okay." In the first, third and fifth blocks, participants were required to affirm relations consistent with a WBIG stereotype, while in the second, fourth and sixth blocks, participants were required to affirm relations that were inconsistent with the WBIG stereotype. No performance criteria were applied during the test blocks in order to proceed. After the sixth block of trials, the screen cleared and a message appeared informing the participant that the experiment was over and to report to the experimenter. Participants were thanked for their co-operation and fully debriefed. All participants completed the experiment in a single session that lasted approximately 20–30 min.

3. Results

3.1. Explicit measure

3.1.1. BAAS questionnaire data

The overall mean scores for male and female participants on the BAAS showed that both gender groups rated their appearance as highly important (male participants, $M=21.40$, $SD=12.24$; female participants, $M=23.73$, $SD=13.78$; see Fig. 2). The BAAS data for male and female participants were subjected to statistical analysis using a 2×4 mixed repeated measures analysis of variance (ANOVA) with gender as the between-participant variable and the four BAAS questionnaire sub-scales as the within-participant variable (i.e., beliefs of the importance of attractiveness to their interpersonal interactions, personal achievement, self-perception and emotional well-being). There was no statistically significant main effect for gender, $F(1, 28)=.394$, $p=.53$; or subscale, $F(3, 24)=2.70$, $p=.06$, but there was a statistically significant interaction effect, $F(3, 24)=6.55$, $p<.001$.

Thus, four one-way between-participants ANOVAs were conducted to identify which of the particular subscales showed a statistically significant difference dependent on participant gender (Fig. 2). These results showed a statistically significant gender difference for the subscale "self-perception" with higher scores for female versus male participants in rating the importance of appearance, $F(1, 28)=4.25$, $p<.05$. There were nonsignificant gender differences ($p>0.05$) shown for the other three subscales: "interpersonal relationships", $F(1, 28)=.08$; "personal achievement", $F(1, 28)=3.00$; "emotional well-being", $F(1, 28)=2.56$.

3.2. Implicit measure (IRAP)

3.2.1. Statistical analyzes

Positive *D*-IRAP (difference) scores indicated a pro-attractive bias and negative *D*-IRAP scores indicated an anti-unattractive bias (see Figs. 3 and 4). The steps involved in calculating *D*-IRAP scores are described in Barnes-Holmes et al. (2010). A 2×4 mixed repeated measures ANOVA was conducted with gender and trial-type as the between and within-participant IVs, respectively, and *D*-IRAP scores as the DV. There was a main effect for trial-type that

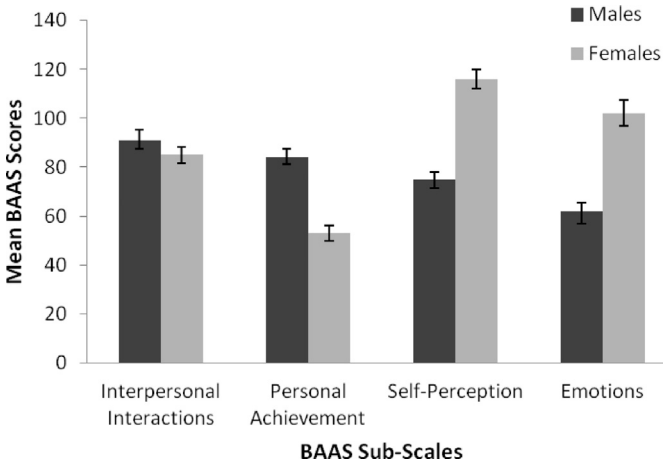

Fig. 2. The mean BAAS scores for males and females on each of the four sub-scales. Error bars denote one standard deviation around the mean.

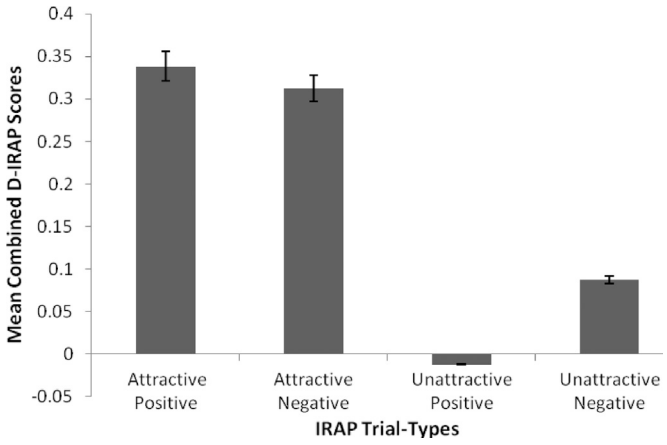

Fig. 3. The mean combined *D*-IRAP scores for the entire group of participants (males and females) for each of the four trial-types. Positive *D*-scores indicate a pro-attractive bias and negative *D*-scores indicate an anti-attractive bias.

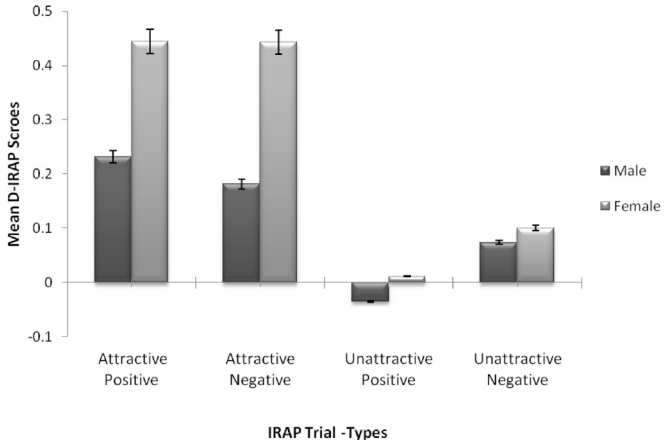

Fig. 4. The mean D-IRAP scores for males and females for each of the four trial-types. Positive D-scores indicate a pro-attractive bias and negative D-scores indicate an anti-attractive bias.

Table 2
Mean overall *D*-IRAP scores. Results of the eight planned one-sample *t*-tests conducted on trial-type.

Trial type		Group					
	Overall *D*-IRAP	Male participants			Female participants		
	M	*M*	*t*	*p*	*M*	*t*	*p*
Attractive–positive	.34	.23	3.15	< .01*	.45	5.37	.01*
Attractive–negative	.31	.18	1.91	.08	.44	4.43	< .01*
Unattractive–positive	−.01	−.04	−.38	.11	.01	.98	.92
Unattractive–negative	.11	.12	1.14	.28	.10	.45	.66

was statistically significant, $F(3, 24)=6.52$, $p=.002$, partial eta squared=.43 (with the direction of the bias detected being pro-attractive, see Fig. 3). There was a nonsignificant main effect for gender, $F(1, 28)=1.93$, $p=.17$ and a nonsignificant interaction effect $F(1, 28)= 3.68$, $p=.22$.

Eight one-sample *t*-tests were conducted with the *D*-IRAP scores for each trial-type in each gender group, to determine which trial-type (/s) differed from zero at the statistically significant level. Bonferroni corrections were applied for multiple comparisons. Results indicated that the *D*-IRAP scores for the *attractive–successful–true* trial-type (proattractive) were different from zero and statistically significant for male participants, $t(14)=3.15$, $p<.01$; data for the other three IRAP trial-types showed nonsignificant differences for male participants (Table 2). The *D*-IRAP scores differed significantly from zero for female participants for both the *attractive–successful–true* trial-type (pro-attractive), $t(14)=5.37$, $p=.01$, and the *attractive–unsuccessful–false* trial-type (pro-attractive), $t(14)=4.43$, $p<.01$ trial-types (Table 2, Fig. 4). The *D*-IRAP scores for the remaining two IRAP trial-types showed nonsignificant differences for female participants (Table 2).

3.3. Implicit/explicit correlations

Two correlation matrices of scores on implicit (IRAP) and explicit (BAAS) measures were calculated; one for male and one for female participants. Each matrix thus involved testing for correlations between participants' *D*-IRAP trial-type scores and scores for males and females on each of the four BAAS subscales. Correlations were nonsignificant for both participant groups.

In summary, both gender groups showed a pro-attractive preference on both explicit and implicit measures. Explicit data showed that both gender groups rated appearance as highly important, and analyzes of the BAAS subscale data indicated that female participants rated "self-perception" regarding appearance to be more important compared to their male counterparts. Implicit data demonstrated via response latencies (speed of responding) that both male and female participant groups more readily affirmed that attractive facial images were successful compared to unattractive facial images. Detailed statistical analyzes facilitated via the IRAP four trial-type methodology provided information regarding the direction of participant bias; for both gender groups prejudice was shown to be pro-attractive and not anti-unattractive.

4. Discussion

The findings in the current study provided preliminary support for the IRAP as a behavior-based implicit measure that is sensitive to attractiveness bias, and that can provide information regarding directionality and gender differences in attractiveness bias. Overall participant data ($N=30$) showed a statistically significant IRAP effect of attractiveness bias favoring successfulness of attractive vs. unattractive facial images. Specifically, overall group responding showed shorter mean response latencies for consistent versus inconsistent trial-blocks, indicating that participants responded faster when the task was to affirm attractive–successful compared to unattractive–successful relations. These results are consistent with the body of research literature using explicit measures showing a favorable bias toward attractive compared to unattractive individuals; and also with other implicit measures showing attractiveness bias such as a modified Stroop task (Van Leeuwen & Maacre, 2004), an IAT (Buhlmann et al., 2009) and a Go/No-Go Association Task (Buhlman et al., 2010).

This IRAP research advanced prior findings by examining the direction and nature of the implicit bias. Specifically, analysis of data from four trial-types was used to examine directionality, and indicated that the pro-attractive implicit bias regarding successfulness was not accompanied by a corresponding implicit anti-unattractive prejudice. The presence of a pro-attractive bias and absence of an anti-unattractive bias was present across all four IRAP trial-types. Interestingly, these findings did not support the conclusions of Griffin and Langlois (2006) who reported that it is generally more often the case that unattractiveness is bad rather

than attractiveness is good, particularly in the domains of intelligence and altruism; although bi-directional stereotyping (attractiveness-is-good and unattractiveness-is-bad) was evident in the domain of sociability (Griffin & Langlois, 2006). The current results in the domain of successfulness are somewhat consistent with those on sociability therefore; to the extent that participant data showed a beauty-is-good attractiveness bias. Taken together, these findings indicate that the directionality of attractiveness bias can be influenced by the domain or attribute being targeted for evaluation. Further research on directionality in attractiveness bias in the domains of intelligence and altruism, and other domains, is needed to clarify conditions that are necessary and sufficient to promote or reduce attractiveness-is-good and unattractiveness-is-bad effects with male and female evaluators. Directionality of attractiveness bias in the domain of "successfulness" may be relevant to any attempts to counter evaluators' beauty bias in selection panels in employment, where competence judgements are at issue. For example, measures designed to reduce anti-unattractive bias effects might be redundant in conditions where attractiveness bias is not comprised of anti-unattractive bias and primarily consists of pro-attractive prejudice. Future research on directionality of attractiveness bias in domains such as competence or intelligence may be informative.

The gender of targets evaluated, and the gender of those providing evaluations are also relevant in attractiveness bias research (see Agthe et al., 2010). The current study used male and female target stimuli (photographic facial images), whereas much of the previous research in facial attractiveness has been conducted with only women targets (as noted by Griffin & Langlois, 2006, who also used women targets only). This creates a problem regarding generaliseability of the bulk of the research on attractiveness bias; it is not scientifically rigorous to *assume* that findings would be similar if the target stimuli were mixed male and female, or indeed were all male, and, moreover, regardless of gender of evaluators. Future research in the area should therefore focus on effects of attractiveness of both women and men, taking account also of gender of evaluators, to determine if bias towards attractive women in areas of intelligence, altruism, and sociability is comparable to bias towards attractive men in similar domains. Such issues may be considered relevant in applied psychology, particularly in guiding policy for important real-world evaluator committees and selection panels such as jury panels.

The gender analysis of data for the four IRAP trial-types in this study indicated a pro-attractive bias on two (*attractive–successful–true* and *attractive–unsuccessful–false*) out of four trial-types for female participants and a pro-attractive bias on one (*attractive–successful–true*) out of four trial-types for male participants. This suggests that for both male and female participants, attractive individuals were more likely to be deemed successful, although responding for female participants suggested that female students also deemed attractive individuals unlikely to be unsuccessful. In addition, there was a significant gender difference shown for the BAAS subscale of "self-perception", with higher scores for female versus male participants. This divergence is somewhat unsurprising given cosmetic surgery statistics; and indicates that females place more value on attractiveness than men, at least in terms of successfulness and self-perception. Further research is required to determine if this is the case in other domains such as intelligence, and how this might impact on an individual in areas such as confidence, belief in one's own abilities, academic/career success, ambition, work ethic, promotion, and so on. This research may also be relevant in clinical practice in developing interventions to promote a greater focus (e.g. of students/employees) on intelligence, ability, determination, and so on in academic and work environments.

Related to this point, and an area for future investigation, is the extent to which attractiveness-bias in terms of successfulness actually impacts interactions and behavior towards others. In prior IRAP research, implicit attitudes towards overweight individuals predicted behavioral intentions toward an overweight target (Roddy et al. 2010). Although this has yet to be replicated in the current context, university students' attractive-is-successful bias may influence not only beliefs about their own abilities, but also interactions with other students and colleagues in the future. Further research is required to determine the extent to which attractiveness-bias in successfulness actually influences the behavior of individuals towards others, or effects interpretations of others abilities or skills. This might have implications for areas such as teacher training, management training, or organizational behavior analysis.

A limitation of the current study is that the small sample size of 30 participants limits the generalizability of the results. Although the results should be interpreted with caution, it should be noted that findings of statistical significance are typically more difficult to detect with smaller samples. A further limitation of this study is that the effects of gender of target stimuli were not analyzed. Recent research has shown however, that such an analysis is possible using the IRAP (Nolan et al., 2013), which could also facilitate analysis of same-sex and different-sex dyads of evaluators and target individuals. Another possible limitation is that the concept of "successfulness" was not clearly defined for participants and the term was used in a general sense; future research might define "successfulness" more specifically as related to particular areas (e.g., employment, interpersonal relationships). In addition, some of the target words (e.g. Friendly/Unfriendly; Out-going/Shy) chosen by our panel could be said to relate to sociability rather than successfulness; in this regard, the general concept of "successfulness" that we tried to encapsulate in the current context was perhaps vulnerable to varied interpretations. Future studies might attempt to investigate attractiveness stereotype effects with "successfulness" that is more specific and encompasses more strict definitions (e.g., career success/political success). Unlike the current research, most of the existent documented research pertains to attractiveness bias related to sociability rather than aptitude or successfulness. Although this limits the direct comparability of current results, future IRAP investigations of attractiveness bias might focus on social attributes or intelligence, as in prior studies, to examine if results previously shown on explicit measures are replicated with an implicit measure. Future research may also include participants explicit ratings of the facial targets used in the IRAP to allow for more direct comparisons between implicit and explicit rating patterns.

Overall, the current study extended the research literature on implicit measures by demonstrating that the IRAP was sensitive to participants' attractiveness bias in the domain of successfulness. This is important because the IRAP is a relatively recent implicit measure, and requires extensive testing in a variety of domains. The IRAP was also shown to be efficacious in elucidating directionality in implicit attractiveness bias, and as such, has potential utility in many domains. Addressing directionality in applied practice is important because in order to address any prejudice one needs to know what precisely the nature of that prejudice is. Future research in the area of attractiveness bias should focus on the extent to which bias in relation to successfulness (and indeed other domains) influences actual patterns of behavior; whether differences emerge in evaluators ratings of female versus male targets; and gender differences of evaluators in relation to other domains such as intelligence. Such findings may have practical implications for a number of settings including industry, education, and society as a whole.

References

Agthe, M., Spörrle, M., & Maner, J. K. (2010). 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.

The British Association of Aesthetic Plastic Surgeons ⟨www.baaps.org.uk⟩. *Annual audit of cosmetic surgery procedures.* ⟨http://baaps.org.uk/about-us/audit/1856-britain-sucks⟩.

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., 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*, 497–516.

Barnes-Holmes, D., Murphy, C., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure: exploring the impact of private versus public context and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record*, *60*, 57–79.

Barnes-Holmes, D., Murtagh, L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure (IRAP) to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *The Psychological Record*, *59*, 389–406.

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.

Buhlmann, U., Teachman, B. A., & 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.

Buhlmann, U., Teachman, B. A., Naumann, E., Fehlinger, T., & Rief, W. (2009). The meaning of beauty: implicit and explicit self-esteem and attractiveness belief in body dysmorphic disorder. *Journal of Anxiety Disorders*, *23*, 694–702.

Cullen, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit Relational Assessment Procedure (IRAP) and the malleability of ageist attitudes. *The Psychological Record*, *59*, 591–620.

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

Clerkin, E. M., & Teachman, B. A. (2009). Automatic and strategic measures as predictors of mirror gazing among individuals with body dysmorphic disorder symptoms. *Journal of Nervous and Mental Disorders*, *197*, 589–598.

Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J. P., & Gore, N. J. (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. (2003). The extrinsic affective Simon test. *Experimental Psychology*, *50*, 77–85.

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

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

Dovidio, J. F., 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, H., Ashmore, D., Makhijani, M., & Longo, C. (1991). What is beautiful is good: a meta-analytic review of research on the physical attractiveness stereotype. *Psychology Bulletin*, *110*, 107–128.

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.

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

Greenwald, A. G., Poehlman, T. A., Uhlmann, E., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. MEta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping. Is beauty good or is ugly bad? *Social Cognition*, *24*, 187–286.

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

Heilman, M. E., & Stopeck, M. H. (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*(2), 202–215.

Hosoda, M., Stone-Romero, E. F., & Coats, G. (2003). The effects of physical attractiveness on job-related outcomes: a meta-analysis of experimental studies. *Personnel Psychology*, *56*, 431–462. http://dx.doi.org/10.1111/j.1744-6570.2003.tb00157.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: a voyage through implicit cognition research. *Implicit Cognition* (pp. 1–81), 1–81.

Hussey, I., & Barnes-Holmes, D. (2012). The IRAP as a measure of implicit depression and the role of psychological flexibility. *Cognitive and Behavioral Practice*, *19*, 573–582.

Jones, W. H., Hansson, R. O., & Phillips, A. L. (1978). Physical attractiveness and judgements of psychopathology. *The Journal of Social Psychology*, *105*, 79–84.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, *126* (390–323).

Lucker, G. W., Beane, W. E., & Helmreich, R. L. (1981). The Strength of the halo effect in physical attractiveness research. *Journal of Psychology: Interdisciplinary Applied*, *107*, 69–75.

McKenna, I. M., 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.

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.

Nosek, B., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–666.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review (pp. 265–292) In: J. A. Bargh (Ed.), Psychology Press.

Power, P. M., 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.

Puleo, R. (2006). Are beautiful people thought to be better? An example of the relationship between physical attractiveness and expectations. *Journal of Undergraduate Psychological Research*, *1*, 35–37.

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.

Ryan, C., & Costa-Giomi, E. (2004). Attractiveness bias in the evaluation of young pianists' performances. *Journal of Research in Music Education*, *52*, 141–154.

Spangler, D. L. (1999). The Beliefs About Appearance Scale. Unpublished Manuscript.

Spangler, D.L, & Stice, E. (2001). Validation of the beliefs about appearance scale. *Cognitive Therapy and Research*, *25*, 813–827.

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.

Van Leeuwen, M. J., & Macrae, C. N. (2004). Is beautiful always good? Implicit benefits of facial attractiveness. *Social Cognition*, *22*, 637–649.

Wapnick, J., Darrow, A. A., & Kovacs, J. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, *45* (470–47).