International Journal of Psychology and Psychological Therapy, 2020, 20, 3, 273-285
20th ANNIVERSARY EDITION // EDICIÓN XX ANIVERSARIO

Measuring Implicit Attractiveness Bias in the Context of Innocence and Guilt Evaluations

Hannah Rice, Carol Murphy
National University of Ireland, Maynooth, Ireland
Conor Nolan, Michelle Kelly*
National College of Ireland, Dublin, Ireland

Abstract

Previous research on mock-jury trials has shown an explicit attractiveness bias in participant attributes of innocence. This study used the Implicit Relational Assessment Procedure (IRAP) to measure attractiveness-bias in implicit evaluations of innocence or guilt with a sample of 46 college students. Alternate IRAP trial-blocks required participants to affirm relations consistent and inconsistent with attractiveness bias (attractive-innocent/unattractive guilty versus unattractive-innocent/ attractive-guilty). Faster responding across consistent trial-blocks was interpreted in terms of implicit stereotype. Participants' beliefs about the importance of their own appearances were examined using the Beliefs about Appearances Scale (BAAS) and explicit attractiveness ratings for the IRAP photographic stimuli were measured using Likert scales; analysis examined relationships between these beliefs and IRAP scores. Results revealed statistically significant attractiveness bias for both male and female participants; specifically, both a pro-attractiveness and anti-unattractive bias. Findings are discussed regarding research in implicit evaluations of innocence or guilt and effects of attractiveness bias.
Keywords: IRAP, attractiveness bias, beauty bias, guilt or innocence.

How to cite this paper: Rice H, Murphy C, Nolan C, & Kelly M (2020). Measuring Implicit Attractiveness Bias in the Context of Innocence and Guilt Evaluations. *International Journal of Psychology & Psychological Therapy, 20, 3,* 273-275.

Novelty and Significance

What is already known about the topic?

- It is still unclear what factors contribute to positive or negative attractiveness biases when considering guilt or innocence.
- Interest in this area is ever-increasing; with the rising popularity of True Crime documentaries, the expanding portrayal of criminals has come under ethical and moral debate.

What this paper adds?

- This study incorporates objective and subjective measures of attractiveness bias, and data are analysed for the influence of gender on attractiveness bias in accord with Sex and Gender Equity in Research guidelines..
- Results support a "what is beautiful is good effect", and there were no significant gender differences reported.

The term 'attractiveness bias' refers to the influence of attractiveness on individuals' evaluations of others. Specifically, physically attractive people are perceived as having more socially desirable personality traits compared to physically unattractive individuals (Dion, Berscheid & Walster, 1972). Many studies have examined the role that attractiveness plays in social judgment and found in almost every context, that attractive people are perceived more favourably than unattractive people (Eagly, Ashmore, Makhijani, & Longo, 1991; Langlois, Kalakanis, Rubenstein, Larson, Hallam, & Smoot, 2000). For instance, attractive individuals are typically regarded as more humorous, amiable, intelligent, helpful, and socially skilled than less attractive individuals (Feingold, 1992; Benson, Karabenick, & Lerner, 1976). Furthermore, employee selection research has shown that more attractive candidates were chosen over equally qualified but less attractive individuals

* *Correspondence*: Michelle Kelly, National College of Ireland, Mayor Street IFSC, Dublin 1, Ireland. Email: michelle.kelly@ncirl.ie. *Data Sharing*: Anonymised data files are available from the first author upon request.

(Hosoda, Stone-Romero, & Coats, 2013). Dion *et alia* (1972) termed this phenomenon the 'what is beautiful is good stereotype'. As such, attractiveness bias may have many implications for social and political decisions. Of concern is whether attractiveness-bias might influence decisions with serious implications, such as in the legal system.

One of the most widely studied extra-legal variables is the defendant's physical attractiveness (Willis & Todorov, 2006), and research literature suggests that physically unattractive defendants are generally at a disadvantage, in both the likelihood of being convicted guilty, and the severity of the recommended sentence. It is suggested that this is due to 'dangerous decisions theory' (DDT). That is, a defendant's untrustworthiness or dangerousness is assessed almost immediately upon first seeing a defendant's face (Willis & Todorov, 2006; Porter, ten Brinke, & Gustaw, 2010). Those perceived as untrustworthy or dangerous in initial judgments are more likely to be found guilty by a judge or jury, and to be given longer sentences (Porter *et alia*, 2010).

Related to these issues, "baby-faced" individuals, or those with small noses, large eyes, a small chin, and a round face are perceived to be weaker and more affectionate than mature-faced individuals (Zebrowitz & Montepare, 2008). These positive social characteristics associated with having a baby-face reduce a defendant's likelihood of being found guilty, and reduce sentence length compared to those with more "mature" faces (Berry & Zebrowitz-McArthur, 1988). There is even research suggesting that benefits accruing to the physically attractive (i.e., advantages in the employment arena) may be somewhat protective against criminal engagement. An association between unattractiveness and increased criminality was found for men and women, although this effect was more pronounced for women (Cavior, Hayes, & Cavior, 1974; Cavior & Howard, 1973). This could in turn supplement the stereotype 'what is beautiful is good'.

It is important to note however, that there have been a small number of conflicting findings in which contextual factors appear to produce contrary results. Termed the 'beauty is beastly' effect (Heilman & Stopeck, 1985), attractiveness was conversely found to disadvantage women in particular employment contexts (e.g., with same-sex evaluations for competitive positions). In courtroom situations, although attractive defendants generally seem to have an advantage, research suggests that this might only be the case for certain crimes, such as rape and robbery (Mazzella & Feingold, 1994). For other crimes, including swindle (Sigall & Ostrove, 1975; Smith & Hed, 1979) and negligent homicide (Mazzella & Feingold, 1994), physically attractive defendants tend to be treated more harshly, as they are perceived to have used their appearance to their advantage; they are also perceived as being capable of better judgment and thus more responsible ('reverse halo effect'; Mazzella & Feingold, 1994).

Stereotypes related to the reverse halo effect are particularly prevalent in the portrayal of crime in the media. Historically, attractive TV criminals were portrayed either as psychopaths that prey on weak and vulnerable victims or as professionals that are shrewd, ruthless, and violent (Surrette, 1989); both types of crimes outlined in the research above. However, the rising popularity of True Crime documentaries and the expanding portrayal of criminals have come under ethical and moral debate (Bonn, 2014). Of particular note, the film Extremely Wicked, Shockingly Evil and Vile (2019) that follows the case of Ted Bundy has come under criticism for the physically attractive portrayal of the serial killer, and the effects of physical attractiveness on public perception of the case (Crenshaw & Stroud, 2019). To date, there is little scientific research quantifying the effects of attractive actors portraying criminals and the effects this may have on public perception retrospectively. However, the presence

of an attractiveness-bias could provide psychological insight into the effects of these portrayals on public opinion.

In addition to the issue of anomalous findings, a considerable limitation in the literature on attractiveness bias is that the bulk of the research has examined attractiveness bias in the social domain, and it is largely comprised of self-report or questionnaire data (Griffin & Langlois, 2006). Questionnaires are an efficient means of collecting data, but have well-documented vulnerabilities related to introspection and presentation bias (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Nosek, Greenwald, & Banaji, 2007). This is 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 clinical phenomena (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Furthermore, extant data in attractiveness bias typically did not disaggregate data for male and female participants or examine potential gender effects. A more nuanced approach in such investigations may help to provide a more comprehensive and informative account of factors involved in attractiveness bias.

Only a small number of studies have attempted to use more objective measures of participant behaviour to demonstrate implicit positive bias toward attractive versus unattractive individuals. These include studies recording reactions times on computer generated tasks such as the modified Stroop (Van Leeuwen & Macrae, 2004), the Go/No Go Association task (e.g. Buhlmann, Teachmann, & Kathmann, 2011) and the Implicit Relational Assessment Procedure (IRAP; Murphy, McCarthaigh, & Barnes-Holmes, 2014; Murphy, Hussey, Barnes-Holmes, & Kelly, 2015). The IRAP (Barnes-Holmes *et alia*, 2006) was developed from a modern behaviour-analytic account of language and cognition called relational frame theory (RFT; see Hayes, Barnes-Holmes, & Roche, 2001), and has been used to examine attractiveness bias in other domains (Murphy *et alia*, 2015) as well as in areas such as implicit self-esteem (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009; Ritzert *et alia*, 2016) and sexual beliefs (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009). Support has been provided for the IRAP in terms of reliability (Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009) and validity (Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009). The IRAP, in common with other implicit measures, has also been shown to have predictive validity toward behaviour (Dawson *et alia*, 2009; Vahey, Nicholson, & Barnes-Holmes, 2015).

The current study aimed to determine if participant responding on the IRAP would show pro-attractive bias, anti-unattractive bias, or both, or no bias, in the context of evaluations of guilt or innocence toward attractive versus unattractive facial photo images. Participants were required to respond under time pressure to relations presented via a computer programme, alternately affirming or denying across trial-blocks relations that were consistent or inconsistent with beauty-positive stereotyping. For example, shorter mean response latencies (i.e., more rapid responding) for affirming consistent relations were deemed an implicit attractiveness bias. Directionality of bias (pro-attractive/ anti-unattractive) was investigated using the IRAP four trial-type methodology; these were *attractive-innocent/ attractive-guilty*; *unattractive-innocent/ unattractive-guilty.* Trial-types one and four presented relations consistent with attractiveness-bias, while trial-types two and three presented relations inconsistent with attractiveness-bias. The current study focused on facial attractiveness (photographic facial images of attractive v. unattractive individuals) because facial attractiveness is deemed of primary importance in an individual's overall attractiveness (Heilman & Stopeck, 1985; Willis & Todorov,

2006; Porter *et alia*, 2009). An explicit attractiveness rating scale was used to determine whether participants deemed the images as attractive or unattractive as intended. Participant's beliefs about their own appearance were measured through the Beliefs about Appearances Scale (BAAS; Spangler, 2001) and data were examined for correlations. All data were analysed for influence of participant gender on attractiveness bias in accord with Sex and Gender Equity in Research (SAGER) guidelines (Heidari, Babor, De Castro, Tort, & Curno, 2016).

METHOD

Participants

Fifty undergraduate students (25 female) participated in the experiment. Participants were recruited using an ad-hoc sampling method of convenience and included both psychology and non-psychology students. All participant volunteers were of Caucasian ethnicity, English-speaking with normal or corrected to normal vision. 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 analysed 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. Data from 4 (2 female) participants were excluded because they failed to achieve the predetermined performance criterion of 80% accuracy on the IRAP. The study adheres to the ethical guidelines as stipulated by the American Psychological Association (APA) Code of Conduct and by the Psychological Society of Ireland (PSI) Code of Professional Ethics. The study received ethical approval from the ethics committee of the first authors' University.

Materials

The Implicit Relational Assessment Procedure (IRAP). The IRAP programme was administered using a portable ASUS X553M Series laptop with a 15 inch monitor operating with the Microsoft Windows 10 system. The IRAP software (Version 2016) presented the experimental trials and recorded participants' responses. Each IRAP trial consisted of a presentation of one of twelve category labels "Guilty" or "Innocent," one of twelve target stimuli (photographic facial images), and two response options, "True" and "False" (see Table 1). The twelve target stimuli were digital colour photographs of averaged faces (e.g., DeBruine, Jones, Unger, Little & Feinberg, 2007) from the Face Research Lab London (http://faceresearch.org/demos/average: consent for inclusion of facial images in scientific research is indicated on the website). The

Table 1. Stimuli and response options of the attractiveness innocent IRAP.

Label stimuli	
Sample 1 (e.g. Attractive)	Sample 2 (e.g. Unattractive)
Positive target words	*Negative target words*
Innocent	Guilty
Good	Bad
Right	Wrong
Sensible	Senseless
Lawful	Criminal
Honest	Dishonest
Response Options	
True	False

twelve photographic stimuli consisted of six images of adult "Attractive" faces and six images of adult "Unattractive" faces. Both 'Attractive' and 'Unattractive' face images consisted equally of three men and three women. Since all participants were of Caucasian ethnicity, all facial images used in the experiment were of Caucasian images in order to avoid racial in-group prejudices confounding the results.

The Attractiveness Rating Measure. The Attractiveness Rating Measure was a questionnaire constructed expressly for this study. Participants filled out an attractiveness 7-point Likert-type questionnaire. Participants were given statements 'this person is attractive' and 'this person is unattractive' alongside each photographic image, and were asked to what extent they agreed with the statement and to circle the corresponding number. The attractiveness scale ranged from 1 (entirely disagree) to 7 (entirely agree). The accompanying photographic images were pre-designated as attractive or unattractive when drawn from the website. Therefore, the attractiveness Likert-scale was used to ensure that the current research participants concurred with pre-designated categories. 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.

The Beliefs about Appearance Scale (BAAS; Spangler & Stice, 2001). The BAAS is a 20- item self-report scale that assesses the degree of endorsement of beliefs about the consequences of appearance for relationships, achievement, self-view, and feelings. Higher scores indicate stronger beliefs, that positive feelings, self-worth, and interpersonal and work success are dependent upon appearance. Participants were asked to mark their degree of agreement with statements about appearance on a 5-point scale ranging from 0 (not at all) to 4 (extremely). This scale possesses acceptable internal consistency (coefficient α= .95), test-retest reliability (r= .83), Cronbach's α (r= 90) as well as good convergent, discriminant, and predictive validity (Spangler & Stice, 2001).

Procedure

Participants completed the research in a testing cubicle or a quiet classroom. Verbal instructions informed participants about the IRAP procedure in accordance with the guidelines of Hussey et alia (2016). Upon commencing the IRAP programme, participants read onscreen instructions asking them to respond to specific rules in each block of trials. The IRAP programme comprised two rules for responding. One rule was consistent with likely existing verbal relations ("attractive innocent and unattractive guilty") while the other rule was inconsistent with these ("attractive guilty and unattractive innocent"). The rule was switched from block to block and the order in which the two types of blocks were presented was counterbalanced across participants. The IRAP commenced with a minimum of one pair of practice blocks, consisting of one consistent and then inconsistent block. When participants selected the response option that was deemed correct within that block of trials the label, target, and response option stimuli were removed immediately from the screen for an inter-trial interval of 400ms, after which the next trial was presented (see Figure 1). When participants selected the response option that was deemed incorrect for that block of trials, the stimuli remained on screen and a red X appeared beneath the target stimulus. The participants were required to select the correct response option, and only then did the program proceed directly to the 400 ms inter-trial interval (followed immediately by the next trial). Participants were informed that trial-types may be consistent or inconsistent with their personal beliefs; however participants should try to respond to the rule provided as fast and as accurately as possible regardless.

If participants failed to achieve both accuracy (>80%) and latency (<2000ms) criteria across a pair of blocks, they received automated feedback, and practice blocks continued to a maximum of four pairs of blocks. Failing to meet the criteria after four pairs of practice blocks, participation was terminated and these data were discarded.

When the criteria were reached on a pair of practice blocks, participants proceeded automatically to three pairs of test blocks. No performance criteria were employed for participants to progress across the three pairs of test blocks, but performance feedback was presented at the end of each block to encourage participants to maintain the criteria. 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 20-30min. The program automatically recorded response accuracy (based on the first response emitted on each trial) and response latency (time in ms between trial onset and emission of a correct response) on each trial. Upon completion of the IRAP programme, participants were asked to complete the attractiveness rating measure and the BAAS using paper and pencil. The completion of the explicit measures took approximately ten minutes, after which participants were debriefed and thanked for their time.

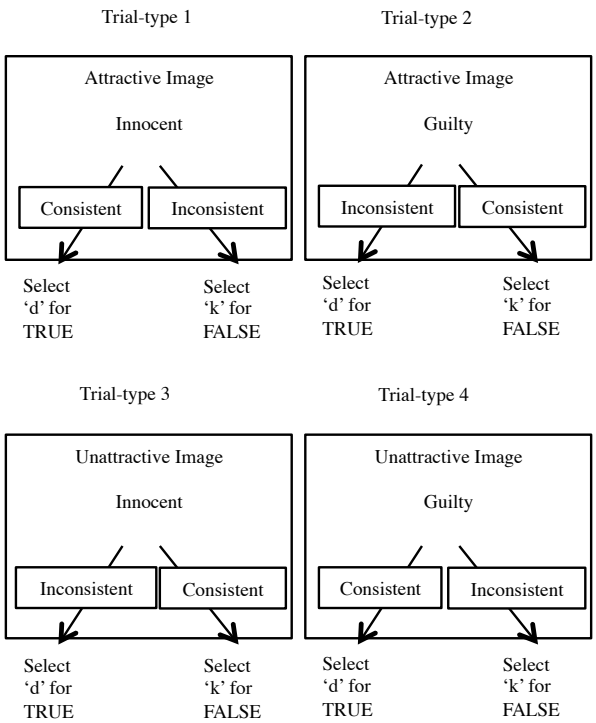

Figure 1. Representations of the four IRAP trial-types. The target stimulus (a photo of either an attractive or unattractive face) appeared at the top of the screen, while the attribute label stimulus (e.g. "Innocent" or "Guilty") appeared in the centre of the screen. Response options (True/False) appeared 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.

Results

The IRAP response latencies, defined as the time in milliseconds (ms) that elapsed between the onset of stimulus presentation on each trial and the first correct response emitted by the participant, were transformed into difference or *D*-scores. *D*-scores were calculated using an adaptation of the Greenwald *et alia* (2003) *D* algorithm (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010). The steps involved in calculating the D_{IRAP} scores have been outlined previously (see Hussey, Thompson, McEnteggart, Barnes-Holmes & Barnes-Holmes, 2015). The data calculations resulted in four overall trial-type D_{IRAP} scores. Initial analysis confirmed that data adhered to the assumption of homogenity of variance and equality of error in variances ($p \geq .18$). Tests of normality for the four IRAP trial-types were not significant. However, as tests of normality were significant for the explicit measure (BAAS) $p \leq .01$, Spearmans rho was used for correlational analyses.

Figure 2 shows the mean D_{IRAP} scores for males and females across the four trial types: attractive-innocent, attractive-guilty, unattractive-innocent, and unattractive-guilty. Positive scores indicate responding consistent with attractiveness bias, while negative scores indicate the reverse. When analysing the relationship between trial-types, descriptive statisitics were initally run for each of the four trial-type *D*-scores (see mean *D*-scores in Figure 2). The D_{IRAP} data were then subjected to statistical analysis via a 2×4 mixed repeated measures ANOVA, with IRAP trial-type as the within-participant factor, gender as the between participant factor, and D_{IRAP} scores as the dependent variable (DV). The analysis revealed a significant main effect for IRAP trial-type; Wilks $\lambda = .37$, $F(3, 42) = 26.32$, $p > .01$, partial $\eta^2 = .63$. However, there was no significant main effect for gender ($p = .59$) and no significant interaction effect.

The strength of the IRAP effects for each trial-type were analysed by conducting one-sample *t*-tests. *D*-scores were statistically significant from 0 on the consistent trial types Attractive-Innocent, for both male and female participants (male participants:

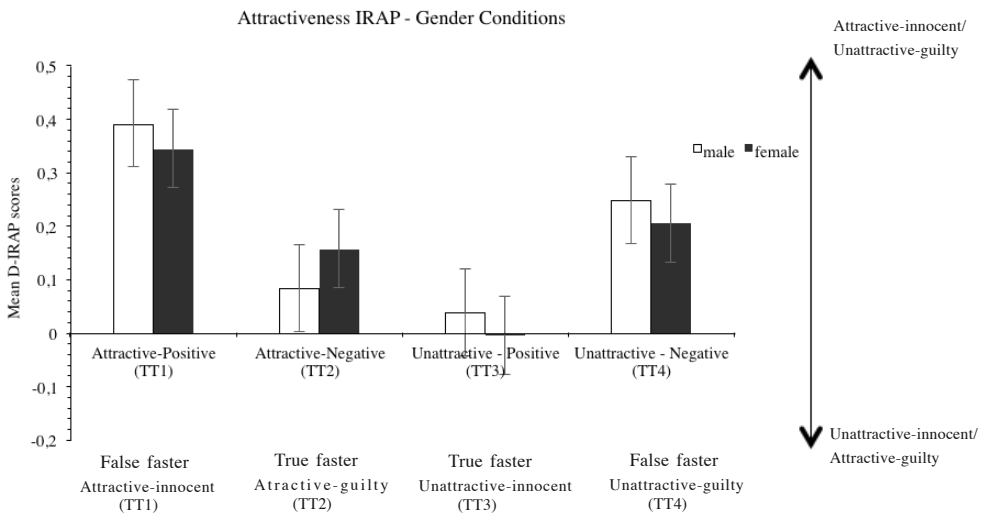

Figure 2. Mean D_{IRAP} scores on attractiveness IRAP for each of the four trial-types across male and female participants. Upper figure: Positive *D*-scores indicate an attractive-innocent/unattractive-guilty bias and negative *D*-scores indicate an unattractive-innocent/attractive-guilty bias.

M= .390, *SD*= .339, *t*(22)= 5.553, *p* <.001; female participants: *M*= .345, *SD*= .324, *t*(22)= 5.098, *p* <.001) and Unattractive-Guilty (male participants: *M*= .249, *SD*= .354, *t*(22)= 3.367 *p*= .003; female participants: *M*= .206, *SD*= .310, *t*(22)= 3.188, *p*= .004). Responses on the Attractive-Guilty and Unattractive-Innocent trial types were not statistically significant from zero (see Table 2).

Table 2. Results of one-sample t-tests with 4 IRAP trial-types for males (M) and females (F). Table shows Mean (*M*), Standard Deviation (*SD*), *t*-value, and *p*-value.

Trial-Type	*M*	*SD*	*t*	*p*-value
Attractive-Innocent-True	M= .390	M= .339	M= 5.553	M= .000
	F= .345	F= .324	F= 5.098	F= .000
Attractive-Guilty-False	M= .083	M= .455	M= .870	M= .393
	F= .157	F= .434	F= 1.741	F= .096
Unattractive-Innocent-False	M= .039	M= .391	M= .471	M= .642
	F= -.005	F= .445	F= -.054	F= .958
Unattractive-Guilty-True	M= .249	M= .354	M= 3.367	M= .003
	F= .206	F= .310	F= 3.188	F= .004

In summary, there was statistically significant attractiveness bias shown on two IRAP trial-types (attractive-innocent/ unattractive guilty) indicating both a pro-attractive and anti-unattractive direction. The tendency was slightly more pronounced for male compared to female participants, but there was no significant gender difference.

Mean data were calculated for the 7-point Attractiveness and Unattractiveness Likert-scales completed by male ("attractive" pictures: *M*= 69.48, *SD*= 1.45; "unattractive" pictures: *M*= 33.88, *SD*= 1.58) and female participants ("attractive" pictures: *M*= 72.76, *SD*= 1.22; "unattractive pictures": *M*= 35.12, *SD*= 1.71). The overall ratings were subjected to a 2x2 mixed repeated measures 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 significant, $F(1, 48)= 4.04$, $p= .05$, partial $\eta^2= .08$. The effect for picture-type was significant, Wilks $\lambda= .10$, $F(1, 45)= 404.54$, $p <.01$, partial $\eta^2= .89$. There was no significant interaction effect. In summary, scores differed significantly between the gender groups as females rated "attractive" pictures as more attractive than males, and rated "unattractive" pictures as more unattractive than males. In addition, both gender groups clearly discriminated between the pictures of "attractive" and "unattractive" facial images; ratings of attractiveness for images were found to accord with "attractive" and "unattractive" categories that had been predesignated.

The overall mean scores for male and female participants on the BAAS showed that both gender groups rated their appearance as important (male participants, *M*= 33.76, *SD*= 12.81; female participants, *M*= 48.64, *SD*= 15.18). The results of an independent *t*-test showed a statistically significant difference between the mean scores for males and females on the BAAS: *t*= -3.75, *p* <.01 (two-tailed).

For males, there was a significant positive correlation between the attractive-guilty trial-type and ratings for attractive faces ($p \leq .01$), and a significant negative correlation between the attractive-guilty trial-type and ratings for unattractive faces ($p \leq .05$). For females, there was a significant negative correlation between unattractive innocent trial-type and ratings for unattractive faces ($p \leq .01$). This suggests that for male

participants, the higher the pictures were scored on attractiveness, the stronger the IRAP effect showing a pro-attractive bias. For females, the more unattractive facial images were judged as unattractive, the greater their implicit unattractive-is-not-innocent bias. There were no significant correlations between BAAS scores and the four trial-types for either male or female participants.

Discussion

The findings in the current study support the IRAP as behavioural measure of 'in-flight' relational responding, and attractiveness bias was evident in participants' faster affirmation of relational presentations that were consistent with this stereotype. Overall, participant data (N= 46) showed a statistically significant bias, favouring implicit evaluations of 'innocence' for attractive v. unattractive facial images. These results are consistent with the robust and well-documented beauty bias shown in explicit measures used in various social domains, and add to the implicit research in attractiveness bias in the context of evaluations of guilt or innocence. The results are consistent also with previous IRAP studies showing implicit attractiveness bias in participant evaluations of successfulness and intelligence (Murphy *et alia*, 2014; Murphy *et alia*, 2015; Ritzert *et alia*, 2016). A pro-attractive bias was shown also in participant explicit ratings, and attractive v. unattractive facial images were rated higher for 'innocence'.

An interpretation of the IRAP data taking account of recent research and theoretical discussion may be warranted. The Differential Arbitrarily Applicable Relational Responding Effects (DAARRE; Finn, Barnes-Holmes, & McEnteggart, 2018) model of IRAP performance predicts the Single Trial-Type Dominance Effects (STTDE; Finn *et alia,* 2018), which is influenced by increased coherence between IRAP stimuli (e.g., label, attribute, response option) on more dominant compared to less dominant trial-types. In the current research trial-types 1 and 4 were both dominant, however trial-type 1 was the most dominant. In this regard, orienting functions of stimuli may be relevant also, contributing to a type of positivity bias (O'Shea, Watson, & Brown, 2015), thus for example, the response option "True" when presented with *Attractive-Innocent* may induce a stronger orienting function, rather than presentation of the response option "False" with *Unattractive-Guilty* (see also Maloney & Barnes-Holmes, 2016 regarding True/False as natural language relational coherence indicators (RCIs) versus relational response options such as Similar/Different; the latter has been found to enhance IRAP effects). These recent research findings suggest that pre-experimental verbal learning history may not be the sole influence in results observed in the current IRAP research. Additionally, the provision of rules to participants (e.g. respond in accordance with consistent or inconsistent relations) can be found to enhance IRAP effects (Finn, Barnes-Holmes, Hussey & Graddy, 2016). On balance, however, across many studies, a substantial percentage of the participant samples fail to complete an IRAP (e.g., more than 20%, see Remue, De Houwer, Barnes-Holmes, Vanderhasselt, & De Raedt, 2013), and the provision of instructional rules has been shown to reduce participant attrition rates. It was toward this end that the current study provided rules regarding responding on alternate trial-blocks.

Analysis of directionality of the bias shown on the IRAP trial-type data indicated both a pro-attractive and anti-unattractive bias. Previous research has suggested that the bi-directionality of the 'beauty is good' stereotype was specific to the domain of sociability. That is, attractiveness is good and unattractiveness is bad was found in

the context of sociable attributes, but not in relation to attributes such as intelligence (Griffin & Langlois, 2006). Murphy *et alia* (2015) found a pro-attractive bias but not an anti-unattractive bias in the context of successfulness evaluations. These differences in directionality may suggest the importance of context in stereotype. However, the issue is currently unclear and may require further elucidation via research manipulating various contexts to determine influence of different domains on directionality of stereotype bias.

In terms of gender influencing attractiveness bias, no significant difference was observed between data for male and female participants for the implicit (IRAP) measure. This finding is inconsistent with previous research, which found male participants showed a stronger implicit attractiveness-bias compared to females (Murphy *et alia*, 2014). However, the study by Murphy *et alia* (2014) had a higher ratio of female to male facial images compared to the even number of male and female facial images in the current study. Previous research measuring explicit attractiveness-innocence bias has reported a stronger attractive leniency effect for participants judging the other gender (Wuensch, Castellow & Moore, 1991). Furthermore, this effect was stronger for male participants judging female defendants than female participants judging male defendants (Efran, 1974). In the current study, attractiveness bias shown for male participants in IRAP trial-type *D*-scores was marginally stronger, but the difference was not statistically significant.

It may be worth further investigation to determine if a larger sample of participants might show more pronounced gender differences. Although small sample size may limit the generalisability of the current results, Vahey *et alia* (2015) found that a sample size of 29 participants is sufficient to provide a study with a statistical power of .80 when examining the statistical significance of first-order Pearson's *r* correlations between clinically-focused IRAP effects and corresponding criterion variables. A limitation was that the current research did not examine the effects of gender of stimuli (i.e., facial images) for any potential interaction with gender of participants, thus the differences between same-sex and different-sex pairings could not be analysed. Ongoing development of the IRAP program, however, suggests it may be readily adapted to assess influence of gender of target individuals (see IRAP software available at www.go-rft.com).

The analysis of the explicit data showed significant gender differences in scores on the BAAS and in ratings on the attractiveness/unattractiveness of the photographic stimuli. This shows that females in our sample had stronger beliefs that positive feelings, self-worth, and interpersonal and work success are dependent upon their appearance (BAAS) compared to males; and also had stronger opinions about the attractiveness of those depicted in the photographs. Gender differences are relatively common in studies examining explicit opinions about body image or satisfaction (Frost & McKelvie, 2004; Grossbard, Lee, Neighbors & Larimer, 2009) or importance of physical appearance (Gentile, Grabe, Dolan-Pascoe, Wells & Matino, 2004) but our results, and those of Murphy *et alia* (2015) did not find similar group differences when implicit attractiveness bias was measured. Gender differences in opinions about ones' own image or attractiveness might therefore be more likely than implicit opinions about others' character/behaviour based on their attractiveness. This is supported by the fact that we found no correlations between the BAAS and the IRAP; the constructs may be too divergent to be comparable. Although the analysis of the data on the photographic stimuli confirmed that participants discriminated between pictures of "attractive" and "unattractive" facial images in accordance with the predesignated attractive and unattractive categories, it was unclear if the images functioned as expected for each participant. Future studies might consider the feasibility

of selecting stimuli for the IRAP individually, based on each participant's ratings, as this may improve the extent to which the function of the stimuli could be predicted.

The issue of correlations between explicit and implicit data is at times complicated, particularly with research in socially sensitive domains when presentation or introspection effects may confound explicit findings. In such cases, no stereotype may be evident in explicit data in contrast with implicit results, and no correlations may be evident between explicit and implicit data (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). If explicit data fails to show a stereotype bias but implicit measures do show such bias, presentation effects may be suspected related to the former. At times, it might be intuitively expected that results would show correlations but researchers should consider whether their implicit and explicit variables should logically be associated. As yet, it is arguable whether such correlations are necessary; this should be determined on a case-by-case-basis, as opposed to 'by default'.

In conclusion, the current research has contributed to findings of implicit attractiveness bias related to participant evaluations guilt or innocence, albeit that in the current context the latter concepts are presented in the abstract rather than in the context of criminality or court judgements. Notwithstanding, the results are consistent with previous research findings in the area of judgments of guilt or innocence, which have shown bias favouring attractive individuals. We hope that further research in implicit attractiveness bias may attempt some form of replication of the current study, to determine if the results are replicable.

References

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, 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*, 287-305. Doi: 10.1007/BF03395708

Barnes-Holmes D, Waldron D, Barnes-Holmes Y, & Stewart I (2009). Testing the validity of the Implicit Relational Assessment Procedure and the Implicit Association Test: Measuring attitudes toward Dublin and country life in Ireland. *The Psychological Record*, *59*, 389-406. Doi: 10.1007/BF03395671

Benson PL, Karabenick SA, & Lerner RM (1976). Pretty pleases: The effects of physical attractiveness, race, and sex on receiving help. *Journal of Experimental Social Psychology*, *12*, 409-415. Doi: 10.1016/0022-1031(76)90073-1

Berry DS & Zebrowitz-McArthur L (1988). What's in a face? Facial maturity and the attribution of legal responsibility. *Personality and Social Psychology Bulletin*, *14*, 23-33. Doi: 10.1177/0146167288141003

Buhlmann U, Teachman BA, & Kathmann N (2011). Evaluating implicit attractiveness beliefs in body dysmorphic disorder using the Go/No-go Association Task. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*, 192-197. Doi: 10.1016/j.jbtep.2010.10.003

Cavior HE, Hayes SC, & Cavior N (1974). Physical attractiveness of female offenders. Effects on institutional performance. *Correctional Psychologist*, *1*, 321-331. Doi: 10.1177/009385487400100403

Cavior N & Howard LR (1973). Facial attractiveness and juvenile delinquency among black and white offenders. *Journal of Abnormal Child Psychology*, *1*, 202-213.

Dawson DL, Barnes-Holmes D, Gresswell DM, Hart AJ, & Gore NJ (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure. *Sexual Abuse*, *21*, 57–75. Doi: 10.1177/1079063208326928

DeBruine LM, Jones BC, Unger L, Little AC, Feinberg DR (2007) Dissociating averageness and attractiveness: Attractive faces are not always average. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1420-1430. Doi: 10.1037/0096-1523.33.6.1420.

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

24, 285-290. Doi: 10.1037/h0033731

Dovidio JF & Fazio RH (1992). New technologies for the direct and indirect assessment of attitudes. In JM Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (p. 204–237). New York: Russell Sage Foundation.

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. Doi: 10.1006/jesp.1997.1331

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. Doi: 10.1037/0033-2909.110.1.109

Efran MG (1974). The effect of physical appearance on the judgment of guilt, interpersonal attraction, and severity of recommended punishment in a simulated jury task. *Journal of Research in Personality*, *8*, 45-54.

Fazio RH, Jackson JR, Dunton BC, & Williams CJ (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychologyocial*, *69*, 1013-1027. Doi: 10.1037/0022-3514.69.6.1013

Fazio RH & Olson MA (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327. Doi: 10.1146/annurev.psych.54.101601.145225

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. Doi: 10.1037/0022-3514.59.5.981

Feingold A (1992). Good-looking people are not what we think. *Psychological Bulletin*, *111*, 304-341. Doi: 10.1037/0033-2909.111.2.304

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

Finn M, Barnes-Holmes D, & McEnteggart C (2018). Exploring the single-trial-type-dominance-effect in the IRAP: Developing a differential arbitrarily applicable relational responding effects (DAARRE) model. *The Psychological Record, 68,* 11-25. Doi: 10.1007/s40732-017-0262-z.

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*, 1741. Doi: 10.1037/a0015575

Griffin AM & Langlois JH (2006). Stereotype directionality and attractiveness Stereotyping: Is beauty good or is ugly bad? *Social Cognition*, *24*, 187-206. Doi: 10.1521/soco.2006.24.2.187

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

Heidari S, Babor TF, De Castro P, Tort S, Curno M (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review volume,* 1, Article number 2. Doi: 10.1186/s41073-016-0007-6x

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*, 202-215. Doi: 10.1016/0749-5978(85)90035-4

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. Doi: 10.1111/j.1744-6570.2003.tb00157.x

Hussey I, Thompson M, McEnteggart C, Barnes-Holmes D, & Barnes-Holmes Y (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*, *4*, 157-162. Doi: 10.1016/j.jcbs.2015.05.001

Langlois JH, 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*, 390-423.

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

Mazzella R & Feingold A (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgment of mock jurors: A meta-analysis. *Journal of Applied Social Psychology*,

24, 1315-1344. Doi: 10.1111/j.1559-1816.1994.tb01552.x

Murphy C, Hussey T, Barnes-Holmes D, & Kelly ME (2015). The Implicit Relational Assessment procedure (IRAP) and attractiveness bias. *Journal of Contextual Behavioral Science*, *4*, 292-299. Doi: 10.1016/j.jcbs.2015.08.001

Murphy C, MacCarthaigh S, & Barnes-Holmes D (2014). Implicit relational assessment procedure and attractiveness bias: Directionality of bias and influence of gender of participants. *International Journal of Psychology and Psychological Therapy*, *14*, 333-351.

Nosek BA, Greenwald AG, & Banaji MR (2007). The Implicit Association Test at age 7: A methodological and conceptual review. *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*, 265-292.

O'Shea B, Watson DG, & Brown GD (2015). Measuring implicit attitudes: A positive framing bias flaw in the Implicit Relational Assessment Procedure (IRAP). *Psychological Assessment, 28, 158-170*. Doi: 10.1037/pas0000172

Porter S, ten Brinke L, & Gustaw C (2010). Dangerous decisions: the impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. *Psychology, Crime & Law*, *16*, 477-491. Doi: 10.1080/10683160902926141

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. Doi: doi.org/10.1007/BF03395684

Remue J, De Houwer J, Barnes-Holmes D, Vanderhasselt MA, & De Raedt R (2013). Self-esteem revisited: Performance on the implicit relational assessment procedure as a measure of self-versus ideal self-related cognitions in dysphoria. *Cognition & Emotion*, *27*, 1441-1449. Doi: 10.1080/02699931.2013.786681

Ritzert TR, Anderson LM, Reilly EE, Gorrell S, Forsyth JP, & Anderson DA (2016). Assessment of weight/shape implicit bias related to attractiveness, fear, and disgust. *The Psychological Record*, *66*, 405-417. Doi: 10.1007/s40732-016-0181-4

Sigall H & Ostrove N (1975). Beautiful but dangerous: Effects of offender attractiveness and nature of the crime on juridic judgment. *Journal of Personality and Social Psychology*, *31*, 410-414. Doi: 10.1037/h0076472

Smith ED & Hed A (1979). Effects of offenders' age and attractiveness on sentencing by mock juries. *Psychological Reports*, *44*, 691-694. Doi: 10.2466/pr0.1979.44.3.691

Spangler DL & Stice E (2001). Validation of the beliefs about appearance scale. *Cognitive Therapy and Research*, *25*, 813-827. Doi: 10.1023/A:1012931709434

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

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

Van Leeuwen ML & Neil Macrae C (2004). Is beautiful always good? Implicit benefits of facial attractiveness. *Social Cognition*, *22*, 637-649. Doi: 10.1521/soco.22.6.637.54819

Willis J & Todorov A (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592-598. Doi: j.1467-9280.2006.01750.x

Wuensch KL, Castellow WA, & Moore CH (1991). Effects of defendant attractiveness and type of crime on juridic judgment. *Journal of Social Behavior and Personality, 6,* 1-12.

Zebrowitz LA & Montepare JM (2008). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*, 1497-1517. Doi: 10.1111/j.1751-9004.2008.00109.x

Received, January 24, 2020
Final Acceptance, June 2, 2020