



The implicit measurement of psychopathy

Jennifer Pink^{a,*}, Robert J. Snowden^b, Nicola S. Gray^{a,c}

^a Swansea University, Singleton Park Campus, Swansea SA2 8PP, Wales, UK

^b Cardiff University, Tower Building, 70 Park Place, Cardiff CF10 3AT, Wales, UK

^c Swansea Bay University Health Board, Caswell Clinic, Glanrhyd Hospital, Tondur Road, Bridgend CF31 4LN, Wales, UK

ARTICLE INFO

Keywords:

Implicit Association Test
Psychopathy
Personality
Aggression
TriPM

ABSTRACT

This study explored whether psychopathic traits might be measured using indirect or “implicit” measures. Three Implicit Association Tests, reflecting each domain of the Triarchic model of psychopathy: Boldness, Meanness and Disinhibition, were completed by mixed-gender, community-based samples ($n_s = 334, 376, 358$). Each IAT demonstrated good internal reliability and showed concurrent validity. Associations with self-report scores on a validated explicit measure of Triarchic psychopathy exceeded those generally reported in the personality IAT literature. The IATs displayed discriminant validity, only correlating with their counterpart domain scores, predictive validity to some explicit measures of known correlates of psychopathy (e.g., aggression), and were robust to socially desirable responding. With further development, the IATs may be a useful research tool alongside self-report psychopathy measures.

1. Introduction

The psychopathic personality is one of the most important concepts in the field of forensic psychology with its strong associations to criminal, antisocial, and violent behaviours (Hare & Neumann, 2008). The core features of the psychopathic personality include a callous disregard for others and a bold and impulsive manner that allows them to violate social norms without guilt or regret (Hare, 1991). A number of instruments have been devised to try to quantify these traits. The most notable is the Psychopathy Checklist-Revised (PCL-R; Hare, 1991) which is clinically administered and requires a trained clinician, a lengthy interview with the person under investigation, as well as comprehensive collateral information relating to criminal history and offending behaviour. Given these demands, attempts have been made to produce psychopathy measures based upon self-report (e.g., Self-Report Psychopathy Scale; SRP-4 (Paulhus et al., 2014), PPI-R (Lilienfeld & Widows, 2005), and TriPM (Patrick, 2010)), and these instruments appear to be producing a wealth of new information relating to the nature of psychopathy and its underpinning psychological (e.g., Segarra et al., 2022; Snowden et al., 2022; Snowden et al., 2017) and neuro-physiological (e.g., Esteller et al. 2016; Koenig et al., 2020; Miglin et al., 2021; Vanova et al., 2022) processes. In this paper, we ask whether an alternate method of measuring psychopathic personality traits, based on indirect or “implicit” measures, may also be possible.

Explicit measures attempt to get the person to report their attitudes, behaviours, likes and dislikes, etc. Such measures require the person to be able to accurately recall this information, evaluate it in relation to contextual information and to others’ behaviour, and be honest about these deliberations (De Cuyper et al., 2017; Puroil et al., 2022). As such, they are seen as the outcome of careful, controlled processes which the person could manipulate if they so wished. Implicit measures aim to bypass these controlled processes and look at early automatic evaluations of attitudes and preferences, etc. For example, implicit and explicit attitudes towards black people have been shown to predict quite different behaviours towards such individuals. Green et al. (2007) noted that physicians’ recommendations for treatment of black individuals differed based on their implicit attitudes but not on their explicit attitudes. Hence, while implicit and explicit attitudes may have some overlap, they also appear to be somewhat distinct, and a wealth of evidence shows that each can predict behaviour under different circumstances (see review by Perugini et al. 2010 and also Buttrick et al. 2020). Could the same techniques also be applied to the field of personality research, with the idea that some aspects of our personality may depend upon controlled and deliberate processes (which are well assessed via self-report measures) while other aspects of personality may be more automatic and be better assayed via implicit measures?

There are now numerous studies that have examined if implicit methods might successfully measure personality or self-concept traits

* Corresponding author at: Department of Psychology, Swansea University, Singleton Park Campus, Swansea SA2 8PP, UK.

E-mail address: 938525@swansea.ac.uk (J. Pink).

<https://doi.org/10.1016/j.jrp.2023.104339>

Received 29 September 2022; Received in revised form 17 December 2022; Accepted 2 January 2023

Available online 5 January 2023

0092-6566/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(Asendorpf et al., 2002; Back et al., 2009; Dentale et al., 2019; Grumm & von Collani, 2007; Kolnes et al., 2021; Puroil et al., 2022; Steffens and König, 2006; Vianello et al., 2010; Vianello et al., 2013; Vecchione et al., 2017); for a review see De Cuyper et al. (2017) – though see Lazarević et al. (2021) for contrary evidence. For example, Vianello et al. (2010: Study 1) developed an implicit measure of “conscientiousness” and found that this was associated with academic performance (defined as the number of exams passed in the last year), whereas an explicit measure of conscientiousness was not related to performance. In study 2, Vianello et al. (2010) showed that both the implicit and explicit measures of conscientiousness predicted future academic performance and that the implicit measure had incremental validity over explicit measures. These results confirm that implicit measures of personality can have value above and beyond what is extracted by explicit, self-report, measures of personality and may become a valuable method as an adjunct to understanding and even predicting behaviour. Given the importance of psychopathic traits in forensic psychology and psychiatry, it is therefore of interest to see if implicit measures of these traits could be developed.

Despite the potential of implicit measures to assay psychopathic traits, there appear to be few published studies of this kind. Florez et al. (2017) constructed two implicit association tests (IAT - see below) that they suggested indexed aspects of psychopathy. One was an “emotional-IAT” with participants categorising pictures as *sadness* or *joy*, and an attribute dimension of *positive* and *negative*. The other IAT was a “morality-IAT” where pictures were categorised as *immoral acts* or *neutral acts*, and an attribute dimension of *positive* and *negative*. However, no difference in IAT scores emerged for either IAT as a function of psychopathy, as indexed by the PCL-R.

The lack of association between IAT scores and psychopathy in Florez et al. (2017)’s study may have originated from the nature of the IAT(s) design. The emotional-IAT assesses whether people associate the concepts of sadness (and joy) as positive or negative. It is not apparent in any model of psychopathic personality that psychopathic people would regard sadness (or joy) as more negative (or positive) than other people. Hence, we do not believe that this IAT has good face validity as a measure of psychopathic personality. The morality-IAT assesses whether people associate an immoral act (in relation to a neutral act) as positive or negative. This IAT does appear to assess an important aspect of psychopathy (Glenn et al., 2009) and should have potential to index antisocial behaviour. However, details of the study and results are not given in the Florez et al. (2017) paper, which does not allow for a critical evaluation of the methodological underpinnings of this study.

Given the potential of implicit methods to aid in understanding psychopathic personality traits, the present study aims to develop these tests. In doing so, we had to make two important decisions: (1) What methodology could be used to measure psychopathic traits? And (2) What are the psychopathic traits that should be measured?

The Implicit Association Test (IAT: Greenwald et al., 1998) involves a dual-classification procedure. For instance, Greenwald et al. (1998) had participants classify images of faces as either “black” or “white” (attribute dimension) as they appeared one at a time on the screen. Interspersed with the face trials were word trials where the person had to classify words as either “pleasant” or “unpleasant” (valence dimension). In one block of trials, responses to the white faces and pleasant words were on the same button (while black and unpleasant were on the other button). On another block of trials, the contingencies were reversed so that white faces and unpleasant words were on the same button (while black and pleasant were on the other button). The rationale behind the IAT is that if people have associations between the attribute and valence dimensions (let us, for example, suggest that someone holds racist beliefs and therefore holds a negative association with black faces), then there will be response interference when “black and pleasant” responses share the same button, but no such interference when “black and unpleasant” share the same button.

The IAT technique has proved popular (as of July 2022, a PubMed

search for the keyword “implicit association test” returns 2413 hits). It has strong psychometric properties (for a review, see Greenwald et al., 2009) and has been used to examine a large variety of attitudes and behaviours. While there is debate, and even controversy, surrounding this test (for instance, see Schimmack, 2021; Brownstein et al., 2019; Jost, 2019), the test appears to be able to capture important aspects of a person’s behaviour and attitudes (Greenwald & Lai, 2020).

Given these results, it is perhaps unsurprising that most work on the implicit measurement of personality traits has adopted the IAT. In this research, the attribute dimension is normally that of a particular personality dimension (e.g., extraversion), and words are classified as either at the high end of this dimension (extravert) or low end (introvert). The valence-dimension of the classic IAT is replaced by a self-reference dimension where words are classified as either belonging to the self (e.g., “me”) or to others. The results from these studies of “personality-IATs” appear highly similar to those from other (non-personality) IATs. A meta-analysis (Hofmann et al., 2005) reported a mean correlation of 0.24 between implicit-explicit measures of personality, while De Cuyper et al. (2017) reported an implicit-explicit correlation of 0.20 from a meta-analysis of 125 personality trait IATs. The mean internal consistency was also comparable between the two meta-analyses, with Hofmann et al. (2005) and De Cuyper et al. (2017) reporting α ’s of 0.79 and 0.82, respectively. The mean predictive effect for observed behaviour for the personality IATs was 0.25, and 0.13 for self-reported behaviour (De Cuyper et al., 2017), which compares with 0.27 overall reported in another meta-analysis by Greenwald et al. (2009). Therefore, the personality IATs appear to perform in a similar manner to IATs that measure other concepts (typically attitudes).

The present research chose to use the IAT as the implicit measure of psychopathic personality traits and constructed three separate IATs relating to the three domains of an established model of psychopathy: the “triarchic” model (Patrick et al., 2009). This model was selected as it attempts to reconcile previous models of psychopathy, along with integrating findings from neurobiology. Additionally, as this research was reliant on using self-report measures of psychopathy, the current study had to select a model of psychopathy that utilised self-report measures, rather than a clinician-administered measure such as the PCL-R (Hare, 1991). Each domain of the triarchic psychopathy model comprises a distinct phenotypic construct: Boldness, which contains the propensity to social dominance, resiliency, and willingness to take risks; Meanness which is defined as aggressive resource seeking without regard for others; and Disinhibition, which reflects a general propensity toward problems of impulse control. The three IATs each aimed to index one of these triarchic domains (see below and Methods for details). The IATs used idiographic stimuli (e.g., “Sarah”) rather than generic pronouns (e.g., “me”) as these are thought to enhance the IAT effect (Bluemke & Friese, 2012) and have proved effective in other research settings (Gray et al., 2021).

The present study (which was not preregistered) constructed three psychopathy-trait IATs reflecting each of the domains of Boldness, Meanness, and Disinhibition. As each dimension could be represented by dichotomies (*bold-wary*, *cruel-kind*, *reckless-sensible*), it was suited to the bipolar structure of the classic IAT (Greenwald et al., 1998). The IAT stimuli were carefully distilled from the items of the self-report measure of the triarchic model, the Triarchic Psychopathy Measure (TriPM; Patrick, 2010) to generate a set of words which reflected the content of the explicit measure as closely as possible. Further detail on this process is given in the Method section.

The study aimed to develop implicit measures of the personality concepts of Boldness, Meanness, and Disinhibition using the IAT. These IATs were then given to a large sample ($N > 1000$) of participants in order to investigate the psychometric properties of the IATs. The explicit measures included the TriPM questionnaire itself, to examine concurrent validity of the IATs. We also constructed simple “personality thermometers” (similar to feeling thermometers used in other research on the IAT, e.g., Gray et al., 2021) that mirrored the bipolar nature of the

IAT measures and used the same labels as the IAT, which may allow for a fairer test of the “explicit vs implicit” relationships (Payne et al., 2008; Snowden et al., 2021).

To examine the external validity of the IATs, we took measures of aggression and prosocial behaviour. The Reactive-Proactive Aggression Questionnaire (RPQ; Raine et al., 2006) is a self-report measure of past aggressive behaviours that divides these into proactive aggression (aggression that is done for a purpose or some instrumental gain) and reactive aggression (aggression caused by anger or frustration). Prior research has shown strong associations between both forms of aggression and the Meanness and Disinhibition scales of the TriPM (Donnellan & Burt, 2016; Pink et al., 2022). Relationships have also emerged between TriPM Boldness and premeditated aggression in some (Gray et al., 2019; Paiva et al., 2020), but not all (Fernandez et al., 2019) studies. Prosocial behaviour was measured via the Prosocialness Scale for Adults (PSA; Caprara et al., 2005). Gatner et al. (2016) have shown a strong negative relationship between TriPM Meanness and PSA scores. Furthermore, prosocial behaviours have been shown to be positively associated with TriPM Boldness (Gatner et al., 2016).

Given that features of psychopathy include manipulation and deceit, there have been concerns about self-report measures of psychopathy (Lilienfeld & Fowler, 2006), with data showing issues of social desirability are present in most self-report measures of psychopathy (Verschuere et al., 2014; Kelsey et al., 2015; Knack et al., 2021). To examine if the psychopathy IATs were resistant or immune to such issues, we used a well-validated measure of social desirability, the Social Desirability Scale (Stöber, 2001) and examined the extent to which the implicit and explicit measures were related to social desirability.

Several hypotheses were made. First, we hypothesised that the three IATs would show high values of internal consistency. Second, that scores on each of the three IATs would correlate with the respective explicit TriPM domain scores, but not with the other scales (or only to the extent that these explicit measures correlate). Third, the IAT scores would positively correlate with ‘personality thermometer’ scores on their respective personality dimension (e.g., Meanness-IAT with Meanness personality thermometer). Fourth, the Meanness-IAT and Disinhibition-IAT scores would be associated with scores on both the Reactive and Proactive scales of the RPQ. Fifth, the Boldness-IAT scores would be associated only with scores on the Proactive scale of the RPQ. Sixth, the Prosocialness Scale (PSA) score would be positively associated with the Boldness-IAT score and negatively associated with the Meanness-IAT score. Lastly, the explicit measures would show associations with the measure of social desirability, whereas the implicit measures would not.

2. Methods and materials

2.1. Participants

2.1.1. Power analysis

This was a between-subjects study to minimise participant demand, and participants undertook one IAT each. The study was powered to detect a correlation of $r = 0.20$ (see De Cuyper et al., 2017), and with standard conditions ($\alpha = 0.05$, 80 % power), this requires a sample of $n = 150$ participants per condition. However, this was increased to $n = 160$ to allow for missing responses and discounted datasets. Thus, with 160 participants required for each of the three conditions (Bold, Mean, Disinhibition), and to allow a comparison of any gender differences should they emerge, the total sample size needed was $N = 960$ ($160 \times 3 \times 2$).

2.1.2. Participant sample

The participant sample was an online, community-based sample recruited through social media channels. The study was advertised as suitable for those aged 18–55 due to possible changes in reaction time with age. It was advertised as suitable only for those fluent in English

due to the complexity of some of the language and concepts described in English. Community participants were eligible for entry in a prize draw to win a £25 shopping voucher, whilst psychology students from the host university received credits for their participation.

A total of 1371 participants commenced the study. After removals (see Results), the final sample for the study was 1068. Of these, 548 were female, 511 were male, 5 were other gender, and 4 preferred not to say. Using the UK Government’s agreed list of ethnic groups (Office for National Statistics, 2021), and consolidating this into their 5-category grouping, 59.4 % of participants reported their ethnicity as White, 4.7 % as Mixed, 12.7 % as Asian, 13.7 % as Black, 8.2 % as other, and 1.3 % preferred not to say. Of the sample, 42.9 % held a first degree, 21.9 % had postgraduate-level qualifications, and 24.1 % had A-Level/A-Level equivalent qualifications. A total of 5.8 % had GCSE/O-Level qualifications, 4.5 % had non-UK or other qualifications, 1.6 % had completed an apprenticeship and 1 % reported no qualifications.

Ethical permission was granted by the host university (Ref. 2021–5034-4155). Informed consent was obtained for experimentation with human subjects. Testing took place between April 2021 and August 2021.

2.2. Materials

2.2.1. Explicit psychopathy

2.2.1.1. Triarchic psychopathy Measure. The TriPM (Patrick, 2010) comprises 58 items across three psychopathy domains: Boldness, Meanness, and Disinhibition. Participants respond to each item on a Likert scale (0 = true, 1 = mostly true, 2 = mostly false and 3 = false). Higher scores indicate more traits of psychopathy.

2.2.1.2. Personality Thermometers. Participants completed two personality thermometers related to their psychopathy domain. Those in the Boldness condition indicated how *bold* they were on a scale of 0 – 100, and on a second thermometer, how *wary* they were. The Meanness participants completed identical thermometers for how *cruel* and how *kind* they were, while the Disinhibition participants completed *reckless* and *sensible* thermometers. We administered two thermometers rather than one where the two ends of a single scale represent the two opposing concepts. This was to account for participants who may consider themselves, for example, to be both partly *bold* and partly *wary*. This evaluation is difficult to achieve on a single dimension thermometer. Descriptions for each personality trait were provided (see supplementary materials). Each thermometer incorporated the IAT stimuli relevant to its domain and the domain descriptions of the TriPM set out by Patrick et al. (2009) upon which the TriPM measure is based. Similar thermometers have been used widely in attitudinal research and demonstrated good reliability in online studies (Liu & Wang, 2015). Explicit ratings of how much various personality traits apply to an individual have been used alongside implicit measures in other studies (e.g., Olson et al., 2007) and have been shown to strongly correlate with other explicit measures (Snowden & Gray, 2013).

2.2.2. Explicit external correlates

2.2.2.1. Reactive and Proactive Aggression. The RPQ (Raine et al., 2006) indexed levels of reactive and proactive aggression. Across 11 items measuring reactive aggression and 12 measuring proactive aggression, participants indicate on a Likert scale (0 = never, 1 = sometimes, 2 = often) the frequency they engage in particular forms of aggressive behaviour.

2.2.2.2. Prosocialness scale for Adults. The PSA (Caprara et al., 2005) was selected to measure levels of prosocial behaviour. Higher scores indicate greater levels of prosocialness. Across 16 items, respondents

indicate on a 5-point Likert scale how true each statement describing a form of prosocial is of them (1 = *never/almost never true*, 2 = *occasionally true*, 3 = *sometimes true*, 4 = *often true*, 5 = *almost always/always true*).

2.2.2.3. Social desirability Scale. The Social Desirability Scale (SDS; Stöber, 2001) assessed levels of socially desirable reporting. Across its 16 items (item 4 was deleted from the original 17-item scale, see Stöber, 2001), participants respond to each with either *true* or *false*. Each item is scored as either 0 or 1, generating a total score of 16 for the measure. Higher scores denote higher levels of socially desirable responding.

2.2.2.4. Attention Checks. To identify careless responders, attention check questions were included in each measure. These appeared midway through each measure and were labelled “This is an attention check question”. Participants were then asked to select a specified point within the Likert scale for that measure. Such attention checks are commonplace in current research and do not affect scale validity (Kung et al., 2018). Data was removed for participants who failed attention checks (see Results section).

2.2.3. Implicit psychopathy

Three implicit association tests (IATs) were developed, one to reflect each of the three TriPM psychopathy domains. In each, participants were required to categorise a series of words as either relating to either end of the psychopathy trait (e.g., *cruel* vs *kind*) or the self (*me* vs *not me*). The IAT category labels and word stimuli were distilled directly from items of the TriPM questionnaire measure to ensure close association to the Triarchic model. Where single trait words were given in the TriPM (e.g., Item 1: *I'm optimistic more often than not*; Item 13: *I am a born leader*), these were included in the stimuli list (*optimistic, leader*). For items which represented traits that were described by more than one word (e.g., Item 3: *I often act on immediate needs*; Item 50: *I don't stack up well against most others*), equivalent single trait words were carefully selected to reflect these (*impulsive, insecure*), being also in line with the descriptions of the domains given by Patrick et al., (2009). As illustrated with these examples, the TriPM includes reverse-scored items which were incorporated into the words used for the non-psychopathic trait stimuli. Not every stimulus word (e.g., *follower, irresponsible*) originated directly from the TriPM. However, these were chosen as bipolar examples of those which were (e.g., *leader, responsible*). Thus, every item chosen either reflected a trait represented by TriPM item(s) or one in diametric opposition. Two of the authors completed this selection process, and agreement from both was required to include each of the stimuli used within the experiment. Each category label was the same as those used in the personality thermometers (see [supplementary materials: Table S1](#)) and, as mentioned previously, each thermometer included the same trait words as those used in the IATs. The participants provided the idiographic *me* and *not me* words. They selected an alternative for each *not me*, which they did not identify with from a selection of prescribed options (see [supplementary materials: Table S2](#)).

Each IAT comprised six blocks (see [supplementary materials: Table S3](#)). All *psychopathy* and *non-psychopathy* stimuli and prompts were displayed in blue text, and idiographic *me* and *not me* stimuli and prompts were shown in green text. This aimed to make the task more straightforward for participants. A fixation cross preceded each trial for 500 ms.

Block 1 was the “attribute discrimination” block (Greenwald et al., 1998), where participants classified the 6 psychopathy and 6 non-psychopathy words as either, e.g., *cruel* or *kind* for the Mean-IAT. On-screen prompts indicated which key to press to categorise, with the L-key for psychopathy and the A-key for non-psychopathy. Participants were given onscreen feedback and could not progress to the subsequent trial if they gave an incorrect response. Trials were presented in a randomised order.

Block 2 was a short practice for Block 3, the “initial combined task”

(Greenwald et al., 1998). This allowed participants to practice sorting traits and personal information as either, e.g., *cruel* or *kind* and *me* or *not me*. Here, a sample of 6 of the psychopathy and non-psychopathy words were combined with 6 of the *me* and *not me* words. In addition to categorising non-psychopathy words with the A-key, participants used the same key to indicate *me* words. Similarly, the psychopathy and *not me* words shared the L-key. For people without psychopathy traits, this key combination (not-psychopathy and *me*) should be congruent and, therefore, fast and easy. Again, participants received onscreen feedback and could not proceed without responding correctly.

Block 3 comprised the data collection “initial combined block”, whereas in Block 2, the concept of *me* was paired with *non-psychopathy*. Each of the 24 stimuli was presented twice (48 trials). Participants were instructed to respond as quickly and accurately as possible. Prompts remained onscreen throughout. However, to minimise the intertrial interval period (Greenwald et al., 2021), and hence, the overall length of the experiment, no feedback was given; participants could move through the trials if they gave an incorrect response.

Block 4 was a reversed “attribute discrimination” block. It was identical to Block 1, with participants classifying the 6 psychopathy and 6 non-psychopathy words as either, e.g., *cruel* or *kind*. However, here the key responses for *psychopathy* and *non-psychopathy* were switched around. Therefore, this time, participants needed to press the A-key for psychopathy words and the L-key for non-psychopathy words. This block prepared the participants for Blocks 5 and 6, where they would be required to sort both traits and personal information as either, e.g., *cruel* or *kind* and *me* or *not me*, but with the response key combinations changed.

Block 5 was a short practice for Block 6, the “reversed combined task” (Greenwald et al., 1998). This was a repeat of Block 2, maintaining the positions for *me* and *not me*, but now pairing *psychopathy* with *me* on the A-key and *not me* with *non-psychopathy* on the L-key. For people without psychopathy traits, this key combination (psychopathy and *me*) should be incongruent and, therefore, slower and more difficult.

Block 6 was the data collection “reversed combined task”. This mirrored Block 3, with each stimulus presented twice (48 trials). However, as in Block 5, *me* shared a response key with *psychopathy*.

2.3. Procedure

The research was completed online on laptops/PCs, using the Gorilla Experiment Builder, <https://www.gorilla.sc> (Anwyl-Irvine et al., 2019) to create and host the study. Participants clicked a link included on social media adverts which took them to the study. They were provided with Participant Information and informed of their right to leave the study without penalty at any point during the research. If they consented, they were asked to provide demographic data of age, gender, ethnicity, and highest educational achievement in line with standardised measures (ONS, 2019). Any participants who indicated they were not fluent in English at this point were automatically thanked and exited from the experiment.

Participants provided their idiographic data for later use in the implicit tasks and selected appropriate equivalents which had no personal relevance to them. Following this, they were randomised into one of the three conditions of Boldness, Meanness, or Disinhibition and completed the relevant personality thermometers for that psychopathy domain. Next, they completed the IAT, and a second implicit task (not reported here) presented in random order. Lastly, participants completed the TriPM, RPQ, SDS and PSA in that order. They were debriefed online, thanked for their time, and offered an opportunity to enter the prize draw or claim participation credits.

2.4. Data analysis

The IATs were scored using standardised procedures (Greenwald et al., 2003) to transform participant scores into single D-scores. A 600

ms penalty was added to the reaction time for each incorrect IAT trial. Trials out of range (less than 300 ms or > 3000 ms) were excluded. The difference in mean RTs between the congruent and incongruent trials was divided by the pooled standard deviation to provide the *D*-score. This process generated positive scores for higher levels of association between psychopathy traits and the self and negative scores for lower levels of association.

Across the questionnaires (Tri-PM, RPQ, SDS, PSA), where items were missing, scores were prorated using complete subscale item scores if there were less than 25 % missing responses. Any questionnaire with 25 % or more missing items was omitted from the analysis.

To assess the reliability and consistency of the IATs, split-half correlations were completed. Tests of association were planned to explore the hypotheses. Analysis was completed using SPSS V26.0. Data from the study is available at [10.17632/gyjw7x6ft2.1](https://doi.org/10.17632/gyjw7x6ft2.1).

3. Results

3.1. Data cleansing

In total, 1371 participants took part in the study. Those under 18 or over 55 were removed from the analysis, along with those not fluent in English, leaving 1296 participants. One hundred forty-eight people failed one or more attention checks, and 80 participants failed to achieve 70 % or greater accuracy on the IAT. The final sample remaining was *N* = 1068 (Boldness *n* = 334, Meanness *n* = 376, Disinhibition *n* = 358).¹

A visual inspection of data histograms for normality identified that distributions for most variables were acceptable for parametric analysis (Tabachnick et al., 2007). The exception to this was the Proactive scale of the Reactive Proactive Questionnaire (RPQ) which showed a strong negative skew. Non-parametric analyses were therefore used for this variable.

3.2. Explicit measures

Table 1 shows the descriptive statistics for the sample for the explicit measures of psychopathy and external correlates associated with psychopathy. Mean scores on the TriPM were similar to those given in previous reports with non-forensic samples, and these scales all showed high internal reliability in line with previous studies (e.g., Sellbom & Phillips, 2013; Kimonis et al., 2020; Pink et al., 2022). Scores on the RPQ were in-line with previous reports in non-forensic samples and showed the same high reliability (e.g., Snowden et al., 2021). The Social Desirability Scale scores were comparable to those reported by Stöber (2001) and, again, demonstrated good reliability. Similarly, the Prosocialness Scale for Adults (PSA) demonstrated excellent reliability and the mean item level score was comparable with those previously reported (Caprara et al., 2005; Gatner et al., 2016).

Associations between the explicit psychopathy measures are given in Table 3. All psychopathic personality thermometers (PT) were strongly and positively associated with their TriPM counterparts. Correspondingly, the non-psychopathic PTs (Wary, Kind, Sensible) were negatively associated with their relevant TriPM domain with moderate to strong effect size.

The pattern of correlations between TriPM domains was similar to those reported elsewhere (e.g., Collison et al., 2021). Comparative correlations between the psychopathy PTs were not available as each participant only completed their respective IAT and thermometer.

¹ There is a possibility that those participants with greater levels of Disinhibition may fail to complete all items on the questionnaires accurately. Thus, the analysis was also run with participants who failed an attention check on one or more of the measures included, which was included as a proxy measure of disinhibition. This modelling generated a very similar pattern of results with similar magnitudes to those reported here.

Table 1

Descriptive statistics for all participants, and per gender (men and women).

	Group	Reported range	Mean (SD)	(α)
Age	All	18–55	27.56 (8.72)	
	Men	18–55	28.36 (8.71)	
	Women	18–55	26.83 (8.68)	
Tri-PM Boldness	All	0 – 52.78	30.30 (8.84)	0.85 [0.84, 0.86]
	Men	6.33 – 52.78	32.97 (8.15)	0.83 [0.81, 0.85]
	Women	0 – 50.00	27.86 (8.75)	0.85 [0.83, 0.87]
Tri-PM Meanness	All	0 – 51	12.34 (8.29)	0.88 [0.87, 0.89]
	Men	0–51	14.97 (8.37)	0.87 [0.85, 0.88]
	Women	0–48	9.88 (7.47)	0.88 [0.86, 0.89]
Tri-PM Disinhibition	All	0 – 57	14.75 (8.00)	0.85 [0.83, 0.86]
	Men	0 – 57	15.90 (8.52)	0.85 [0.83, 0.87]
	Women	0–50	13.61 (7.25)	0.83 [0.81, 0.85]
PT Bold	All	0–100	61.08 (22.86)	
	Men	10–100	67.02 (20.42)	
	Women	0–100	55.23 (23.76)	
PT Wary	All	0–100	44.49 (26.30)	
	Men	0–95	40.30 (25.49)	
	Women	0–100	47.84 (26.37)	
PT Cruel	All	0–100	15.14 (21.88)	
	Men	0–100	18.01 (22.70)	
	Women	0–100	12.60 (21.06)	
PT Kind	All	0–100	77.10 (19.06)	
	Men	1–100	74.66 (19.71)	
	Women	0–100	79.37 (18.34)	
PT Reckless	All	0–100	26.80 (23.39)	
	Men	0 – 100	30.73 (24.84)	
	Women	0–100	22.89 (21.15)	
PT Sensible	All	0–100	72.67 (18.99)	
	Men	0–100	72.20 (19.29)	
	Women	0–100	73.25 (18.68)	
RPQ Reactive	All	0 – 22	7.53 (3.90)	0.82 [0.81, 0.84]
	Men	0–22	8.06 (4.21)	0.84 [0.82, 0.86]
	Women	0–20	7.03 (3.52)	0.80 [0.78, 0.83]
RPQ Proactive ²	All	0 – 22	1.00 (0.00, 2.00)	0.81
	Men	0–22	1.00 (0.00, 3.00)	0.83
	Women	0–16	1.00 (0.00, 2.00)	0.74
SDS	All	0 – 16	10.45 (3.13)	0.72 [0.70, 0.75]
	Men	1–16	10.51 (3.12)	0.72 [0.68, 0.75]
	Women	0–16	10.40 (3.15)	0.73 [0.70, 0.76]
PSA*	All	0 – 5	3.86 (0.66)*	0.91 [0.90, 0.92]
	Men	0 – 5	3.71 (0.68)*	0.91 [0.89, 0.92]
	Women	0–5	3.99 (0.62)*	0.90 [0.89, 0.91]

* For consistency with other papers, range, mean, and SD reported here is at item-level (out of 5), not total score which ranges from 0 to 80. Tri-PM = Tri-archic Psychopathy Measure, PT = Personality Thermometer, RPQ = Reactive Proactive Questionnaire, PSA = Prosocialness Scale for Adults, SDS = Social Desirability Scale.

² As the distribution of scores on the Proactive scale of the RPQ was skewed, median and interquartile range (IQR) and McDonald’s Omega (Hayes & Coutts, 2020) are reported.

3.3. Boldness-IAT

The mean Boldness-IAT score was 0.60 (SEM = 0.03). Thus, most participants associated themselves with being bold rather than wary at this implicit level. Men were significantly bolder than women on this measure. The reliability of the IAT (split-half correlations) was high and comparable to that of TriPM Boldness (see Table 2).

Table 3 displays the correlations of the Boldness-IAT to the explicit measures of psychopathy and the external validity measures. Crucially, the IAT was significantly positively correlated with the Boldness scale of the TriPM and the Bold-PT. Furthermore, it was negatively correlated with the Wary-PT with moderate effect size. As expected, the Boldness-IAT was not significantly correlated with the other TriPM scales and thus showed discriminant validity.

To further examine the relationship between the TriPM scales and the Bold-IAT, a multiple regression was completed, with the Bold-IAT score as the criterion variable, and the three scales of the TriPM entered simultaneously as predictor variables. The model was significant ($R^2 = 0.19, p < .001$) with only the Boldness scale of the TriPM being positively predictive of Bold-IAT score ($\beta = 0.46, p < .001$). Similar results were found when age was added to the model, and when separate models were run for each gender.

The relationship of the Bold-IAT to predict the external indicators of validity are shown in Table 4. The Bold-IAT was not significantly associated with either the RPQ aggression scale or self-reported prosocial behaviour as defined by the PSA. Separate analysis for each gender produced similar results.

3.4. Meanness-IAT

The mean score on the Meanness-IAT was -0.77 (SEM = 0.03). Therefore, participants associated themselves with being kind rather than cruel at this implicit level. Men appeared a little less kind (more cruel) than women on this measure. The reliability of the IAT was high (split-half correlations) and comparable to the TriPM Meanness (see Table 2).

The correlations of the Meanness-IAT to the explicit measures of psychopathy are shown in Table 3. The Meanness-IAT was significantly and positively correlated with the Meanness scale of the TriPM and the Cruel-PT, and negatively correlated with the Kind PT. As expected, it was not significantly correlated with the other TriPM scales and thus showed discriminant validity.

To further examine the relationship between the TriPM scales and the Meanness-IAT, a multiple regression was completed, with the Meanness-IAT score as the criterion variable, and the three scales of the TriPM entered simultaneously as predictor variables. The model was significant ($R^2 = 0.06, p < .001$) with only the Meanness scale of the TriPM being predictive of Mean-IAT score ($\beta = 0.25, p < .001$). Similar results were found when age was added to the model, and when separate

Table 2

Descriptive statistics, reliability, and gender comparison (mean) for the psychopathy Implicit Association Tests (IATs).

	N	(α)	Mean (SD) ¹	Men: Women	p	Effect Size [95 %CI]
Boldness-IAT	334	0.85	0.60 (0.51)	0.66: 0.54	0.029	0.24 [0.02, 0.46]
Meanness-IAT	376	0.86	-0.77 (0.52)	-0.71: -0.83	0.019	0.24 [0.04, 0.45]
Disinhibition-IAT	358	0.86	-0.58 (0.53)	-0.56: -0.61	0.314	0.10 [-0.10, 0.31]

¹ The negative mean IAT scores indicate that overall, most participants associated themselves with being kind rather than cruel, or sensible rather than reckless. The positive mean IAT score for the Bold-IAT indicates that most participants associated themselves with being bold rather than wary.

models were run for each gender.

The Meanness-IAT was positively correlated with the Proactive Aggression scale of the RPQ ($r = 0.16$) while the correlation with Reactive Aggression ($r = 0.08$) showed a positive trend ($p = .05$). The Meanness-IAT was negatively associated with self-reported prosocial behaviour ($r = -0.17$). No significant gender differences emerged between these correlations.

3.5. Disinhibition-IAT

The mean score on the Disinhibition-IAT was -0.58 (SEM = 0.03). Thus, most of the sample associated themselves as being sensible rather than reckless at this implicit level. There was no difference in scores between men and women. The reliability of the IAT was high and comparable to TriPM Disinhibition (see Table 2).

The correlations of the Disinhibition-IAT to the explicit measures of psychopathy and to the external validity measures are shown in Table 3. The IAT was positively correlated with the Disinhibition scale of the TriPM and the Reckless-PT. Moreover, it was negatively correlated with the Sensible-PT. As expected, it was not significantly correlated with the other TriPM scales and thus showed discriminant validity.

To further examine the relationship between the TriPM scales and the Disinhibition-IAT, a multiple regression was completed, with the Disinhibition-IAT score as the criterion variable, and the three scales of the TriPM entered simultaneously as predictor variables. The model was significant ($R^2 = 0.03, p < .001$) with only the Disinhibition scale of the TriPM being predictive of Disinhibition-IAT score ($\beta = 0.16, p = .02$). However, adding age to the model reduced the predictive validity of the Disinhibition TriPM score ($\beta = 0.12, p = .09$). Separate analyses for each gender showed that Disinhibition TriPM score was no longer significant for either gender alone (men $\beta = 0.09$, women $\beta = 0.16, ps > 0.1$).

The Disinhibition-IAT was positively correlated with both the Proactive Aggression and Reactive Aggression scales of the RPQ though with small effect sizes. It was not related to self-reported prosocial behaviours.

3.6. Social desirability and psychopathy

One of the reasons given to consider developing implicit measures of psychopathy was that they might be less affected by issues of desirable responding. Table 4 presents the correlations between social desirability as measured by the SDS and the measures of psychopathy. As hypothesised, none of the three IAT measures was associated with the SDS, while all three TriPM scales showed significant associations. However, it should be noted that while the Meanness and Disinhibition sales showed the expected negative relationship to social desirability (the more that they endorse socially desirable items, the less they endorsed items of Meanness and Disinhibition), the relationship of the SDS to the Boldness scale was positive. These issues will be covered in the Discussion.

Lastly, we examined if desirable responding might influence the relationships of the IAT to the explicit measures and external correlates. The partial correlations between these measures were calculated after controlling for SDS scores (Tables 3 and 4). As expected, for the IAT measure, there were no large changes from the zero-order correlations, as might be hypothesised given that implicit measures should be relatively immune to positive impression management and the lack of relationship between the IAT scores and the SDS.

4. Discussion

In this study, we developed three Implicit Association Tests (IAT; Greenwald et al., 1998). Each represented one of the dimensions of psychopathy defined by the Triarchic model of psychopathy (TriPM; Patrick, 2010): Boldness, Meanness and Disinhibition. Each IAT had good internal reliability, and there was evidence of discriminant validity across several analyses. Furthermore, associations emerged between

Table 3

Zero-order coefficients (r) and partial correlations (partial r) after controlling for social desirability scores on the SDS, between implicit and explicit measures of psychopathy.

		Tri-PM			Personality Thermometers					
		Boldness	Meanness	Disinhibition	Bold	Wary	Cruel	Kind	Reckless	Sensible
IAT	Boldness	0.42* / 0.41*	-0.01 / 0.03	-0.05 / -0.02	0.43* / 0.43*	-0.38* / -0.38*	-	-	-	-
	Meanness	0.01 / 0.02	0.23* / 0.21*	0.12 / 0.08	-	-	0.19* / 0.19*	-0.18* / -0.16	-	-
	Disinhibition	0.03 / 0.04	0.12 / 0.10	0.16 / 0.15	-	-	-	-	0.18* / 0.17*	-0.26* / -0.25*
TriPM	Boldness	-	0.24* / 0.32*	-0.03 / 0.05	0.74* / 0.73*	-0.59* / -0.58*	0.02 / 0.03	0.11 / 0.09	0.07 / 0.14	-0.03 / -0.08
	Meanness	-	-	0.57* / 0.50*	0.12 / 0.19*	-0.07 / -0.13	0.43* / 0.44*	-0.35* / -0.31*	0.46* / 0.40*	-0.32* / -0.26*
	Disinhibition	-	-	-	-0.03 / 0.05	0.12 / 0.07	0.26* / 0.26*	-0.23* / -0.17*	0.58* / 0.52*	-0.39* / -0.31*

Figures in bold $p < .01$; * $p < .001$. IAT = Implicit Association Test, TriPM = Triarchic Psychopathy Measure. Correlations not available between all personality thermometers and IATs as participants only completed their respective thermometer and IAT.

Table 4

Zero-order coefficients (r or rho) and partial correlations (partial r or rho) after controlling for social desirability scores on the SDS, between implicit and explicit measures of psychopathy, and external correlates of psychopathy.

		RPQ		PSA	SDS
		Proactive ¹	Reactive		
IAT	Boldness	0.03 / 0.06	-0.03 / 0.01	0.08 / 0.06	0.09
	Meanness	0.16 / 0.13	0.08 / 0.04	-0.17 / -0.15	-0.09
	Disinhibition	0.13 / 0.12	0.13 / 0.11	-0.06 / -0.04	-0.07
TriPM	Boldness	0.17* / 0.26*	0.00 / 0.08	0.07 / 0.01	0.15*
	Meanness	0.51* / 0.44*	0.47* / 0.37*	-0.52* / -0.45*	-0.35*
	Disinhibition	0.54* / 0.46*	0.48* / 0.34*	-0.32* / -0.19*	-0.43*

Figures in bold $p < .01$; * $p < .001$. IAT = Implicit Association Test, TriPM = Triarchic Psychopathy Measure, RPQ = Reactive Proactive Questionnaire, PSA = Prosocialness Scale for Adults, SDS = Social Desirability Scale.

¹ As the distribution of scores on the Proactive scale of the RPQ was skewed, Spearman's correlations and partial correlations are reported.

measures of aggression and prosocial/antisocial behaviour and the IATs, which replicate previously identified associations with TriPM psychopathy domains. The IATs developed in this study are a promising starting point from which to explore the use of implicit measures, well-utilised in related personality and attitudinal research, in personality disorder research.

4.1. Reliability

As predicted, each of the IATs demonstrated good internal consistency (>0.80), equivalent to that achieved by the TriPM questionnaire. While good internal consistency is typical of the IAT and is a key reason it is selected over some other indirect measurement tasks, the consistency of each psychopathy IAT exceeds the median reported by De Cuyper et al. (2017) in their meta-analysis of 51 personality self-concept IATs, and that reported by Hofmann et al. (2005) in a meta-analysis of a range of IATs. It is encouraging for this level of consistency to emerge for our personality IATs, given that this level is not always achieved (e.g., Vianello et al., 2013).

4.2. Concurrent and discriminant validity

Each of the IATs showed significant correlations to its explicit counterpart scale for both the TriPM and for the Personality

Thermometers – thus, the IATs showed concurrent validity. The magnitude of associations between each IAT and its corresponding explicit measures were comparable with, or more than, many reported within the relevant literature. De Cuyper et al. (2017) calculated a median correlation of 0.20 between implicit and explicit measures.

The psychopathy IATs also appear to have achieved discriminant validity in that each IAT was only significantly correlated to its counterpart scale in the explicit measures and not to the other scales. This is important as it might be argued that such IATs might merely be tapping some “valence” dimension with the words being simply classified by the participants are either “good” words or “bad” words and effectively turning each IAT into a form of “self-esteem” IAT (for a discussion of this point see Grumm and von Collani, 2007). This “problem” may be exaggerated in the domain of psychopathy as, almost by its nature, the words used to describe psychopathic traits such as Meanness are universally seen as negative (e.g., “cruel”). Though we attempted when choosing the stimulus items for the IATs to have both “positive” and “negative” words to describe each psychopathic trait, it did not prove possible. However, the finding that each IAT was only associated with its own explicit scale would seem to argue that participants were not simply recoding the words into “good vs bad” as this would produce the same levels of correlation for each of the IATs with the explicit scales.

4.3. Predictive validity

The pattern of associations that the IATs held with explicit measures of aggression was not wholly as predicted, and where they emerged, they were of small effect size. In line with predictions, the Meanness IAT scores were positively associated with scores on the Proactive scale of the RPQ. This is consistent with previous findings where TriPM Meanness has related to proactive forms of aggression (e.g., Donnellan & Burt, 2016; Gray et al., 2019). Also as predicted, the Meanness-IAT and Reactive RPQ were related, however this was only a weak association ($p = .05$). Previous studies of TriPM Meanness have shown a relationship to reactive aggression (Donnellan & Burt, 2016; Gray et al., 2019) and this association was present between these two self-report measures within our data. However, it is important to highlight that other research (Pink et al., 2022) with a mixed gender sample such as this, has found no relationship between TriPM Meanness and Reactive RPQ aggression. Consistent with predictions, Disinhibition-IAT scores were associated with both Reactive and Proactive RPQ aggression scales, although the relationship with Reactive aggression scores was weak. This concurs with previous findings using explicit measures of Disinhibition (e.g., Donnellan & Burt, 2016; Gray et al., 2019; Pink et al., 2022) and the corresponding relationship with Disinhibition TriPM scores in this study. The hypothesised association between the Boldness-IAT and

Proactive RPQ aggression did not emerge, while it was evident for TriPM Boldness scores.

That the predicted associations with aggression were either not present, or of small effect size when present, merits some consideration. Unsurprisingly, in our community sample, low levels of all three triarchic psychopathy domains emerged. In a similar sample, levels of aggressive, violent, and antisocial behaviours (which are typically associated with psychopathy) only increased rapidly after a certain threshold of psychopathy was reached (Coïd & Yang, 2008). Thus, it is possible that in our low psychopathy sample, reactive and proactive aggression was at a low level, and therefore not associated with the low levels of implicit psychopathy detected by the IATs. An alternative consideration is whether a strong association would be expected between IATs based upon the triarchic model of psychopathy and antisocial behaviours such as proactive and reactive aggression. Some authors argue that the triarchic model of psychopathy diverges away from its core traits of antagonism and impulsivity, by its inclusion of Boldness (see Sleep et al., 2019 for a discussion). Rather than conceptualising the psychopath as a malign and malevolent criminal with self-regulation issues, the triarchic model places less weight on criminality and considers the psychopath to be a dominant, insensitive, impulsive risk taker with a lack of regard or empathy for others (Patrick et al., 2009). However, this is a matter of debate, and research with serious offenders has found associations between Boldness and proactive aggression (e.g., Gray et al., 2019) and correlations between the Boldness scale and the Interpersonal and Antisocial facets of the PCL-R (e.g., Venables et al., 2014).

The relationship between the IATs and prosocial behaviour, measured by the PSA, partly aligned with the hypotheses. As expected, and in line with Gatner et al. (2016), the Meanness-IAT was negatively associated with prosocial behaviour. This is perhaps unsurprising given that the PSA indexes a range of positive actions that are inconsistent with the callousness considered an important component of TriPM Meanness (Patrick et al., 2009) and which is reflected in the items included within the Mean-IAT. Against hypotheses, there was no association found between TriPM Boldness and PSA scores, and this was repeated in a lack of association between the Bold-IAT and the PSA. This lack of association between Boldness in both explicit and implicit measures is inconsistent with previous findings (Gatner et al., 2016).

4.4. Social desirability

As expected, no association emerged between each of the IATs and social desirability (measured by the SDS), while all the three explicit TriPM scales were strongly related. As minimising desirable responding is one of the key motivators for the development of implicit measures, this is a further positive outcome for the psychopathy IATs. TriPM Meanness and Disinhibition were negatively associated with social desirability, while it was positively associated with TriPM Boldness. These associations mirror those found by Kelley et al. (2018), where the Positive Impression Management Scale (PIM; Morey, 1991) was used to index socially desirable responding. As Boldness in part represents a positive sense of self with outgoing and socially engaging behaviours, there may be shared characteristics with the extraverted and emotionally stable attitudes and behaviours which are captured by measures of social desirability (Kelley et al., 2018).

4.5. Limitations and future directions

Considering the findings together, the psychopathy IATs designed here offer a starting point from which to develop further IATs to tap into traits associated with the psychopathic disorder. The reliability was good, they each produced an explicit-implicit association of either equivalent or larger effect size to others published in the implicit personality literature, and each showed discriminant validity. However, given the novel nature of this study and the weak associations with self-

reported external correlates of psychopathy (e.g., aggression, a lack of prosocial behaviour), there are several limitations and considerations for future directions.

Firstly, while we based our IATs on the triarchic psychopathy model, there are other conceptualisations of psychopathy with associated self-report measures which researchers might prefer using. These include the Self-Report Psychopathy Scale (Paulhus et al., 2014), the Levenson Self Report Psychopathy scale (Levenson et al., 1995) and the PPI-R (Lilienfeld & Widows, 2005). Some of these alternatives may be more suitable for exploring the development of IATs specifically for offending populations, given any potential limitations of the triarchic model mentioned above. The methods used in developing our psychopathy IATs could be easily applied to others which represent alternative models of psychopathy, as required. Secondly, in developing the IATs, we used words and concepts included in the TriPM in order to reflect the explicit measure within the implicit task as closely as possible. Due to the nature of psychopathy, it was challenging to give a balance of positive and negative word stimuli to reflect both psychopathic and non-psychopathic traits. Thirdly, an important limitation of the study in assessing the external validity of the IATs is that we used self-report measures of past behaviours associated with psychopathic traits. It may be argued that the best validation of implicit measures of psychopathy would be to measure more spontaneous and implicit behaviour (e.g., microaggressions; Sue, 2010) rather than explicitly stated self-report measures of aggression or prosocial behaviour. It would be interesting to use behavioural measures to index a range of types of aggression and prosocial/antisocial behaviours to assess the predictive utility of the three psychopathy-IATs more robustly. Lastly, as mentioned previously, the study was conducted on a community sample that, presumably, contained few individuals who might be classified as “psychopathic” (e.g., a PCL-R score > 30). Future studies are needed to explore the utility of these, or similar implicit measures, with individuals who have higher levels of psychopathy, and a range of external correlates.

4.6. Conclusion

Three IATs have been created to index each of the three Triarchic psychopathy domains. They have been shown to be internally reliable, display concurrent validity, to have discriminant validity to explicit measures of the same constructs, and have some predictive validity to external criteria. Their associations with corresponding domain scores from a validated explicit measure of psychopathy (TriPM) and with personality thermometers which reflected the triarchic model, met and exceeded those generally reported in the personality IAT literature. Furthermore, each IAT was unrelated to social desirability, while the TriPM scales were all strongly related to this measure of desirable responding. With further development and extensive evaluation to ascertain if they are predictive of aggressive or other problematic behaviours associated with psychopathy, these implicit tools may prove useful in research settings as an adjunct to self-report measures of psychopathy.

Funding

Jennifer Pink's work was funded by the Economic and Social Research Council (UK).

CRedit authorship contribution statement

Jennifer Pink: Investigation, Formal analysis, Resources, Software, Writing – original draft. **Robert J. Snowden:** Methodology, Validation, Writing – review & editing, Supervision. **Nicola S. Gray:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are very grateful to Francisca Aiyejuro, Georga De Freitas Ludlow, Laurinda Doodoo, Heather Gilmour, Connor Goodall, Catrin Hancock, Linny Jacob, Katiana Marrero-Lapinell, Joshua Rabosa and Holly Rees for their assistance in data collection as part of their MSC dissertations, under the supervision of Professor Gray.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jrp.2023.104339>.

References

- Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology*, 83(2), 380–393. <https://doi.org/10.1037//0022-3514.83.2.380>
- Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, 97(3), 533–548. <https://doi.org/10.1037/a0016229>
- Bluemke, M., & Frieze, M. (2012). On the validity of idiographic and generic self-concept Implicit Association Tests: A Core-Concept Model. *European Journal of Personality*, 26, 515–528. <https://doi.org/10.1002/per.850>
- Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *WIREs. Article e1501 Cognitive Science*, 10(5). <https://doi.org/10.1002/wcs.1501>
- Buttrick, N., Axt, J., Ebersole, C. R., & Huband, J. (2020). Reassessing the incremental predictive validity of Implicit Association Tests. *Journal of Experimental Social Psychology*, 88, e103941.
- Caprara, G. V., Steca, P., Zelli, A., & Capanna, C. (2005). A new scale for measuring adults' prosocialness. *European Journal of Psychological Assessment*, 21(2), 77–89. <https://doi.org/10.1027/1015-5759.21.2.77>
- Coid, J., & Yang, M. (2008). The distribution of psychopathy among a household population: Categorical or dimensional? *Soc Psychiatry Psychiatr Epidemiol*, 43, 773–781. <https://doi.org/10.1007/s00127-008-0363-8>
- Collison, K. L., Miller, J. D., & Lynam, D. R. (2021). Examining the factor structure and validity of the Triarchic Model of Psychopathy across measures. *Personality Disorders: Theory, Research, and Treatment*, 12(2), 115–126. <https://doi.org/10.1037/per0000394>
- De Cuyper, K., De Houwer, J., Vansteelandt, K., Perugini, M., Pieters, G., Claes, L., et al. (2017). Using indirect measurement tasks to assess the self-concept of personality: A systematic review and meta-analyses. *European Journal of Personality*, 31, 8–41. <https://doi.org/10.1002/per.2092>
- Dentale, F., Vecchione, M., Ghezzi, V., & Barbaranelli, C. (2019). Applying the latent state-trait analysis to decompose state, trait, and error components of the Self-Esteem Implicit Association Test. *European Journal of Psychological Assessment*, 35(1), 78–85. <https://doi.org/10.1027/1015-5759/a000378>
- Donnellan, M. B., & Burt, S. A. (2016). A further evaluation of the triarchic conceptualization of psychopathy in college students. *Journal of Psychopathology and Behavioral Assessment*, 38, 172–182. <https://doi.org/10.1007/s10862-015-9512-z>
- Esteller, A., Poy, R., & Moltó, J. (2016). Deficient aversive-potentiated startle and the triarchic model of psychopathy: The role of Boldness. *Biological Psychology*, 117, 131–140. <https://doi.org/10.1016/j.biopsycho.2016.03.012>
- Fernandez, D., Zabala, M. C., Ros, L., Martínez, M., Martínez, A., Latorre, J. M., et al. (2019). Testing the properties of the triarchic model of psychopathy in a community sample: Self-reported trait aggression and drug consumption associations. *Scandinavian Journal of Psychology*, 60, 377–385. <https://doi.org/10.1111/sjop.12542>
- Florez, G., Villa, X. A., Lado, M. J., Cuesta, P., Ferrer, V., Garcia, L. S., et al. (2017). Diagnosing psychopathy through emotional regulation tasks: Heart rate variability versus implicit association task. *Psychopathology*, 50, 334–341. <https://doi.org/10.1159/000479884>
- Gatner, D. T., Douglas, K. S., & Hart, S. D. (2016). Examining the incremental and interactive effects of Boldness with Meanness and Disinhibition within the triarchic model of psychopathy. *Personality Disorders: Theory, Research, and Treatment*, 7(3), 259–268. <https://doi.org/10.1037/per0000182>
- Glenn, A. L., Raine, A., & Schug, R. A. (2009). The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry*, 14(1), 5–6. <https://doi.org/10.1038/mp.2008.104>
- Gray, N. S., Blumenthal, S., Shuker, R., Wood, H., Fonagy, P., & Snowden, R. J. (2019). The Triarchic Model of Psychopathy and antisocial behavior: Results from an offender population with personality disorder. *Journal of Interpersonal Violence*, 36(17), 1–23. <https://doi.org/10.1177/0886260519853404>
- Gray, N. S., Knowles, J., George, D., Harvey, A., Powell, R., Zadeh, M. V., et al. (2021). Explicit and implicit hopelessness and self-injury. *Suicide and Life-Threatening Behavior*, 51(3), 606–615. <https://doi.org/10.1111/sltb.12743>
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., et al. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22(9), 1231–1238. <https://doi.org/10.1007/s11606-007-0258-5>
- Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, D. F., Frieze, M., et al. (2021). Best research practices for using the Implicit Association Test. *Behavior Research Methods*, 54, 1161–1180. <https://doi.org/10.3758/s13428-021-01624-3>
- Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, 71(1), 419–445. <https://doi.org/10.1146/annurev-psych-010419-050837>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216. <https://doi.org/10.1037/0022-3514.85.2.197>
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <https://doi.org/10.1037/a0015575>
- Grumm, M., & von Collani, G. (2007). Measuring Big-Five personality dimensions with the implicit association test—Implicit personality traits or self-esteem? *Personality and Individual Differences*, 43(8), 2205–2217. <https://doi.org/10.1016/j.paid.2007.06.032>
- Hare, R. D. (1991). *The Hare psychopathy checklist-revised: Manual*. Incorporated: Multi-Health Systems.
- Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology*, 4, 217–246. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091452>
- Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *But. Communication Methods and Measures*, 14(1), 1–21. <https://doi.org/10.1080/19312458.2020.1718629>
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31(10), 1369–1385. <https://doi.org/10.1177/0146167205275613>
- Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science*, 28(1), 10–19. <https://doi.org/10.1177/0963721418797309>
- Kimonis, E. R., Kidd, J., Most, S. B., Krynen, A., & Liu, C. (2020). An elusive deficit: Psychopathic personality traits and aberrant attention to emotional stimuli. *Emotion*, 20(6), 951–964. <https://doi.org/10.1037/emo0000601>
- Kelley, S. E., Edens, J. F., Donnellan, M. B., Mowle, E. N., & Sörman, K. (2018). Self-and informant perceptions of psychopathic traits in relation to the triarchic model. *Journal of Personality*, 86(4), 738–751. <https://doi.org/10.1111/jopy.12354>
- Kelsey, K. R., Rogers, R., & Robinson, E. V. (2015). Self-report measures of psychopathy: What is their role in forensic assessments? *Journal of Psychopathology and Behavioral Assessment*, 37(3), 380–391. <https://doi.org/10.1007/s10862-014-9475-5>
- Knack, N., Blais, J., Baglole, J. S., & Stevenson, A. (2021). Susceptibility of the Self-Report Psychopathy Scale (SRP 4) to Response Distortion and the Utility of Including Validity Indices to Detect Deception. *Psychological Assessment*, 33(12), 1181–1191. <https://doi.org/10.1037/pas0001044>
- Koenig, S., Wu, Z., Gao, Y., et al. (2020). Abnormal cortical activation in visual attention processing in sub-clinical psychopathic traits and traumatic brain injury: Evidence from an fNIRS study. *Journal of Psychopathology and Behavioral Assessment*, 42, 627–636. <https://doi.org/10.1007/s10862-020-09808-5>
- Kolnes, M., Uusberg, A., & Konstabel, K. (2021). On the relationship between explicit and implicit self-concept of extraversion and neuroticism. *Journal of Research in Personality*, 90. <https://doi.org/10.1016/j.jrp.2020.104061>
- Kung, F. Y., Kwok, N., & Brown, D. J. (2018). Are attention check questions a threat to scale validity? *Applied Psychology*, 62(2), 264–283. <https://doi.org/10.1111/apps.12108>
- Lazarević, L. B., Bjekić, J., & Knežević, G. (2021). Multitrait Multimethod validation of Implicit Association Test as a measure of individual differences in personality: Is there personality at all? *Journal of Research in Personality*, 93. <https://doi.org/10.1016/j.jrp.2021.104124>
- Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, 68(1), 151–158. <https://doi.org/10.1037/0022-3514.68.1.151>
- Lilienfeld, S. O., & Fowler, K. A. (2006). The self-report assessment of psychopathy: Problems, pitfalls, and promises. In C. J. Patrick (Ed.), *Handbook of Psychopathy* (pp. 107–132). The Guilford Press.
- Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathy Personality Inventory Revised (PPI-R)*. Professional manual. Lutz, FL: Psychological Assessment Resources.
- Liu, M., & Wang, Y. (2015). Data collection mode effect on feeling thermometer questions: A comparison of face-to-face and Web surveys. *Computers in Human Behavior*, 48, 212–218. <https://doi.org/10.1016/j.chb.2015.01.057>

- Miglin, R., Rodriguez, S., Bounoua, N., & Sadeh, N. (2021). A multidimensional examination of psychopathy traits and gray matter volume in adults. *Social Cognitive and Affective Neuroscience*, 17(7), 662–672. <https://doi.org/10.1093/scan/nsab131>
- Morey, L. C. (1991). *Personality Assessment Inventory professional manual*. Odessa, FL: Psychological Assessment Resources.
- Office for National Statistics. (2021). *List of ethnic groups*. List of ethnic groups - GOV.UK (ethnicity-facts-figures.service.gov.uk).
- Olson, M. A., Fazio, R. H., & Hermann, A. D. (2007). Reporting tendencies underlie discrepancies between implicit and explicit measures of self-esteem. *Psychological Science*, 18(4), 287–291. <https://doi.org/10.1111/j.1467-9280.2007.01890.x>
- Paiva, T. O., Pasion, R., Patrick, C. J., Moreira, D., Almeida, P. R., & Barbosa, F. (2020). Further evaluation of the triarchic psychopathy measure: Evidence from community adult and prisoner samples from Portugal. *Psychological Assessment*, 32, e1–e14. <https://doi.org/10.1037/pas0000797>
- Paulhus, D. L., Neumann, C. S., & Hare, R. D. (2014). *Manual for Self-Report Psychopathy (Scale(4th ed.))*. Multi-Health Systems.
- Patrick, C. J. (2010). Triarchic psychopathy measure (Tri-PM). In PhenX (Ed.), *Toolkit Online assessment catalog*.
- Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualisation of psychopathy: Developmental origins of Disinhibition, Boldness, and Meanness. *Development and Psychopathology*, 21, 913–938. <https://doi.org/10.1017/S0954579409000492>
- Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 94(1), 16–31. <https://doi.org/10.1037/0022-3514.94.1.16>
- Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory and applications* (pp. 255–277). The Guilford Press.
- Pink, J., Snowden, R. J., Price, M. J., Kocsondi, A., Lawrence, C., Stephens, P., et al. (2022). Refining the relationship between psychopathy, aggression, and rule-breaking by gender: A comparison of the triarchic and septarchic models of psychopathy. *Personality and Individual Differences*, 185. <https://doi.org/10.1016/j.paid.2021.111282>
- Puroil, M. F., Oh, J., Abrom, M. C., Bernard, H., Forest, A., Huhtala, S. D., et al. (2022). Age differences in implicit and explicit personality traits. *Personality and Individual Differences*, 197. <https://doi.org/10.1016/j.paid.2022.111765>
- Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., et al. (2006). The Reactive-Proactive Aggression Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, 32, 159–171. <https://doi.org/10.1002/ab.20115>
- Schimmack, U. (2021). The Implicit Association Test: A method in search of a construct. *Perspectives on Psychological Science*, 16(2), 396–414. <https://doi.org/10.1177/1745691619863798>
- Segarra, P., Poy, R., Branchadell, V., Ribes-Guardiola, P., & Moltó, J. (2022). Psychopathy and heart rate variability: A new physiological marker for the adaptive features of Boldness. *Personality Disorders: Theory, Research, and Treatment*. <https://doi.org/10.1037/per0000573>
- Sellbom, M., & Phillips, T. R. (2013). An examination of the triarchic conceptualization of psychopathy in incarcerated and nonincarcerated samples. *Journal of Abnormal Psychology*, 122(1), 208–214. <https://doi.org/10.1037/a0029306>
- Sleep, C. E., Weiss, B., Lynam, D. R., & Miller, J. D. (2019). An examination of the Triarchic Model of psychopathy's nomological network: A meta-analytic review. *Clinical Psychology Review*, 71, 1–26. <https://doi.org/10.1016/j.cpr.2019.04.005>
- Snowden, R. J., Amad, S., Morley, E., Butkute, N., Budd, R., Jackson, L., et al. (2021). Explicit and implicit self-esteem and aggression: Differential effects of agency and communion. NP10036-NP10059 *Journal of Interpersonal Violence*. <https://doi.org/10.1177/0886260520985490>
- Snowden, R. J., Frongillo Juric, A., Leach, R., McKinnon, A., & Gray, N. S. (2022). Automatic processing of emotional images and psychopathic personality traits. *Cognition and Emotion*, 1–15. <https://doi.org/10.1080/02699931.2022.2054780>
- Snowden, R. J., & Gray, N. S. (2013). Implicit sexual associations in heterosexual and homosexual women and men. *Archives of Sexual Behavior*, 42(3), 475–485. <https://doi.org/10.1007/s10508-012-9920-z>
- Snowden, R. J., Smith, C., & Gray, N. S. (2017). Risk taking and the triarchic model of psychopathy. *Journal of Clinical and Experimental Neuropsychology*, 39(10), 988–1001. <https://doi.org/10.1080/13803395.2017.1300236>
- Steffens, M. C., & König, S. S. (2006). Predicting spontaneous big five behavior with implicit association tests. *European Journal of Psychological Assessment*, 22(1), 13–20. <https://doi.org/10.1027/1015-5759.22.1.13>
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment*, 17(3), 222–232.
- Sue, D. W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation*. Hoboken, NJ: Wiley.
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, 481–498).
- Vanova, M., Aldridge-Waddon, L., Norbury, R., Jennings, B., Puzzo, I., & Kumari, V. (2022). Distinct neural signatures of schizotypy and psychopathy during visual word-nonword recognition. *Human Brain Mapping*, 43, 3620–3632. <https://doi.org/10.1002/hbm.25872>
- Vecchione, M., Dentale, F., Alessandri, G., Imbesi, M. T., Barbaranelli, C., & Schnabel, K. (2017). On the applicability of the big five implicit association test in organizational settings. *Current Psychology*, 36(3), 665–674. <https://doi.org/10.1007/s12144-016-9455-x>
- Venables, N. C., Hall, J. R., & Patrick, C. J. (2014). Differentiating psychopathy from antisocial personality disorder: A triarchic model perspective. *Psychological Medicine*, 44, 1005–1013. <https://doi.org/10.1017/S003329171300161X>
- Verschuere, B., Uzieblo, K., De Schryver, M., Douma, H., Onraedt, T., & Crombez, G. (2014). The inverse relation between psychopathy and faking good: Not response bias, but true variance in psychopathic personality. *The Journal of Forensic Psychiatry & Psychology*, 25(6), 705–713. <https://doi.org/10.1080/14789949.2014.952767>
- Vianello, M., Robusto, E., & Anselmi, P. (2010). Implicit conscientiousness predicts academic performance. *Personality and Individual Differences*, 48(4), 452–457. <https://doi.org/10.1016/j.paid.2009.11.019>
- Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences*, 55, 994–999. <https://doi.org/10.1016/j.paid.2013.08.008>