As accepted to PD:TRT April 2018.

Running head: GENUINE AND POSED DISTRESS IN PSYCHOPATHY

All tears are crocodile tears:

Impaired perception of emotion authenticity in psychopathic traits

Amy Dawel^{1*}, Luke Wright², Rachael Dumbleton², Elinor McKone¹

- Amy Dawel: amy.dawel@anu.edu.au; ORCID iD: [0000-0001-6668-3121](https://orcid.org/0000-0001-6668-3121)
- Luke Wright: luke.wright@anu.edu.au
- Rachael Dumbleton: rachael.dumbleton@anu.edu.au
- Elinor McKone: elinor.mckone@anu.edu.au; ORCID iD: [0000-0003-1655-4297](https://orcid.org/0000-0003-1655-4297)

¹Research School of Psychology, and ARC Centre of Excellence in Cognition and its Disorders, The Australian National University, ACT 2600, Australia.

²Research School of Psychology, The Australian National University, ACT 2600, Australia.

*Corresponding author: Research School of Psychology (building 39), The Australian National University, Canberra, ACT 2600, Australia. tel. +61 2 6125 4106 e-mail: amy.dawel@anu.edu.au

Abstract

In everyday life, other peoples' distress is sometimes genuine (e.g., real sadness), and sometimes pretended (e.g., feigned sadness aimed at manipulating others). Here, we present the first study of how psychopathic traits affect responses to genuine versus posed distress. Using facial expression stimuli, and testing individual differences across the general population ($N = 140$), we focus on the affective features of psychopathy (e.g., callousness, poor empathy, shallow affect). Results show that, while individuals low on affective psychopathy report greater arousal and intent-to-help towards faces displaying genuine relative to posed distress, these differences weakened or disappeared with higher levels of affective psychopathy. Strikingly, a key theoretical prediction—that arousal should mediate the association between affective psychopathy and intent-to-help—was supported only for genuine distress, and not for posed distress. A further novel finding was of reduced ability to discriminate the authenticity of distress expressions with higher affective psychopathy, which, in addition to and independently of arousal, also mediated the association between affective psychopathy and reduced prosociality. All effects were specific to distress emotions (did not extend to happy, anger, disgust), and to affective psychopathy (did not extend to Factor 2 psychopathy, disinhibition, boldness). Overall, our findings are highly consistent with Blair's theorising that atypical processing of distress emotions plays a key etiological role in the affective aspects of psychopathy. We go beyond these ideas to add novel evidence that unwillingness to help others is also associated with a failure to fully appreciate the authenticity of their distress.

Word count = 250 words (max = 250)

Key words ($N = 5$ for PD:TRT): callous; psychopathy; genuine emotion; arousal; prosocial.

There is broad agreement across theorists and clinical experts that affective deficits are core to psychopathy (e.g., Kreis, Cooke, Michie, Hoff, & Logan, 2012; Miller, Lynam, Widiger, & Leukefeld, 2001). These include a lack of emotional responsiveness to other people (e.g., deficits in social emotions such as empathy, guilt and remorse) that manifests as callous, cold and uncaring behaviour, and deficient prosocial behaviour (Frick & White, 2008; Shirtcliff, Vitacco, Graf, Gostisha, Merz & Zahn-Waxler, 2009). Blair's (1995, 2005) influential theorising proposes these affective features develop through problems in processing other people's distress (e.g., fearful and sad facial expressions), and particularly a lack of aversive arousal to others' distress. Here, we address several interrelated issues relevant to evaluating Blair's theorising, and also propose a new idea that it may be atypical responses to others' *genuine* distress that are particularly important.

Blair's aetiological theory of the affective features of psychopathy

Psychopathy is a multidimensional and heterogeneous construct. The affective deficits fall within *Factor 1* (F1)¹ of Hare and colleagues' two-factor model of psychopathy (Hare, Harpur, Hakistan, Forth, Hart & Newman, 1990), *meanness* in Patrick, Fowles and Kruegers' (2009) triarchic psychopathy model, and *callous unemotional traits* (CU) in other work focussing on this set of traits alone (Frick & White, 2008). In comparison, *Factor 2* (F2) of Hare's model is composed of lifestyle (e.g., impulsivity and stimulation seeking) and antisocial (e.g., criminal behaviour) features, and the two remaining dimensions of Patrick and colleagues' triarchic model focus on *disinhibition* (i.e., impulsivity and poor behavioural control) and *boldness*, which encompasses traits of low fearfulness, stress tolerance and a dominant personality style.

Blair (1995, 2005; also cf. 2017) proposes that central to the affective features of psychopathy are impairments in processing other peoples' distress cues (e.g., fear and sad expressions; but not their non-distress emotions, e.g., happy, angry). Typical individuals

¹ Factor 1 in Hare and colleagues' (1990) model also incorporates other interpersonal features, including glib and superficial behaviour, grandiosity, and manipulativeness.

experience aversive, negative arousal in response to others' distress (e.g., Krebs, 1975; Hein, Lamm, Brodbeck & Singer, 2011). This aversive arousal is conceptualized as “punishing” any antisocial behavior that might be causing the distress, which motivates the person to stop the behaviour. The theory is then that this mechanism is impaired in psychopathy (potentially driven by abnormalities in amygdala function; Blair 2005, 2017), with lack of aversive arousal (e.g., Patrick, Cuthbert, & Lang, 1994) leading to failure to inhibit antisocial behaviour.

Blair's theoretical focus has been on explaining *antisocial* behaviour. Equally, however, we argue his proposed mechanisms have potential to explain the lack of caring, kind and helpful behaviour—that is, the lack of *prosocial* behaviour—that is also characteristic of the affective features of psychopathy (White, 2014). In typical individuals, aversive arousal to others' distress not only inhibits antisocial behaviour but can also potentially serve to motivate prosocial behaviour (e.g., via a “concern mechanism”; Nichols, 2001). Typical individuals express greater sympathy and intent to help towards faces showing distress expressions than non-distress expressions (Marsh & Ambady, 2007), and most people will offer help or comfort to a distressed person (Batson, Duncan, Ackerman, Buckley & Birch, 1981). Concerning arousal, individuals with stronger arousal to others' distress (measured via skin conductance responses, SCRs) are more likely to respond altruistically (Krebs, 1975; Hein, et al., 2011), and extremely prosocial individuals (e.g., willing to give a kidney to a stranger) show enhanced amygdala response to distress expressions (i.e., the opposite pattern to psychopaths; Marsh, Stoycos, Brethel-Haurwitz, Robinson, VanMeter, & Cardinale, 2014).

Affective psychopathy and prosocial behavior: Mediated by arousal to others' distress?

Our first two research questions derive directly from the ideas described above. First, we test the prediction prosocial responses will be reduced with higher affective psychopathy specifically in the context of others displaying distress (fearful and sad facial expressions), as assessed by ratings of intent-to-help. We could find no previous studies that have directly

tested this prediction. Although psychopathic traits have been associated with reductions in some types of prosocial behaviour, including particularly for the affective component (reduced self-reported prosocial behaviour in everyday life, for F1 but not F2, White, 2014; reduced reciprocity in the trust game, for F1 and also F2, Ibanez et al., 2016; more self-interest and reduced charitable donation in a money game task, CU in adolescents, Sakai et al., 2012, 2016) previous studies have not evaluated prosocial behaviour specifically to others displaying distress emotions (e.g., in the questionnaire used by White, 2014, most of the items referred to ideas such as donating money to the needy).

Second, we provide the first direct test of Blair's (1995) key mediation-by-arousal idea. Using mediation analyses in regression, we test the prediction that any association between affective psychopathy and reduced prosocial responses to others' distress should be *mediated by reduced arousal* to that distress. Again, we could find no previous literature explicitly testing this idea. Although psychopathic traits have been associated with reduced arousal to others' distress cues, including specifically for the affective component (reduced SCR to pictures of crying faces, for F1 but *not* F2, in children, Blair, 1999; reduced SCR to vocalized distress such as baby crying, for F1 but *not* F2, in adults, Verona, Patrick, Curtin, Bradley & Lang, 2004; reduced heart rate acceleration to a film showing a scared boy, for CU in children, Anastassiou-Hadjicharalambous & Warden, 2008), there have been no tests of whether this reduced arousal to distress acts as a *mediator* between affective psychopathy and reduced prosocial behaviour. Indeed, we were also unable to locate any formal tests of whether arousal mediates even the link between psychopathic traits and *antisocial* behaviour.

Potential importance of the authenticity of the distress emotion displayed

Our third research question raises a new theoretical possibility: that Blair's ideas, which focus on affective—that is, *emotional*—processing in psychopathy might receive stronger support when faces display genuinely-felt rather than posed distress (i.e., pretended emotion;

Figure 1). In everyday life, facial expressions sometimes reflect a genuinely-felt emotion (e.g., sadness when watching a sad movie, or when a pet dies) and are sometimes posed for other social reasons (e.g., deliberately feigning sadness to garner sympathy). An important component of social interaction is the ability to respond properly to the different social signals communicated by genuine and posed displays, including their implications for *prosocial behaviour*, such as "I really need help" (genuine sadness) versus "I am manipulating you to get what I want" (e.g., sadness posed to try to gain an advantage). Currently, nothing is known about responses to emotion authenticity in psychopathy.

Here, we provide the first direct test of responses to others' genuine versus posed distress in psychopathic traits. Concerning prosociality, low affective psychopathy individuals might be expected to show greater intent to help towards someone displaying genuine compared to posed distress. For high affective psychopathy individuals however, we hypothesized this prioritization towards helping authentically distressed individuals might be impaired. Turning to arousal, typical individuals report experiencing greater arousal to genuine than posed emotions, at least for happy faces (Krumhuber, Likowski & Weyers, 2014). Assuming arousal would also be stronger to genuine than posed distress in low affective psychopathy individuals, we hypothesized the typical pattern of increased arousal to genuine distress might also be impaired in high affective psychopathy. Note, for both arousal and intent-to-help, these ideas predict an affective psychopathy by genuineness interaction. Additionally, they imply that the predicted mediation of intent to help by arousal might be apparent only for genuine distress.

Our fourth research question goes beyond Blair's theorizing, to test basic perceptions of emotion authenticity in psychopathy, via ratings of whether each expression displayed genuine versus fake (pretended) emotion (Figure 1). Success in social interactions also requires us to distinguish accurately between authentic and posed emotion—that is, to *perceive* genuineness accurately. Impaired authenticity discrimination has been reported in other clinical disorders

(depression, autism spectrum disorder; Blampied, Johnston, Miles & Liberty, 2010; Boraston, Corden, Miles, Skuse & Blakemore, 2008; Douglas, Porter & Johnston, 2012). Here, we tested the novel hypotheses that: (a) higher affective psychopathy individuals might have problems telling apart genuine from posed distress (e.g., whether sadness is real or “crocodile tears”), as measured by the difference in genuineness ratings for genuine *minus* posed expressions; and (b) any initial failure to properly perceive whether a person is authentically distressed or “faking it” might contribute to reduced willingness to help, and thus reduced genuineness discrimination ability might be an additional mediator (over and above effects of reduced arousal) of reduced intent-to-help.

Specificity to distress and to affective features of psychopathy

Finally, we test the specificity of our major findings to *distress* emotions, and to the *affective* features of psychopathy as compared to other components such as disinhibition and boldness. Blair's theory predicts that associations between psychopathy and arousal and intent-to-help will be specific to distress expressions (fear and sadness), and should be absent for other facial expressions such as non-distress emotions (we test happy, anger, disgust) or neutral expressions. His ideas also apply specifically to the affective features of psychopathy rather than psychopathy as a whole (particularly in more recent instantiations; Blair, 2017).

Study design

We tested individual differences in a large undergraduate sample, and obtained ratings to facial expression stimuli in three tasks. For prosocial intent, participants rated how much they would want to be able to help each person (face), for genuine and posed expressions of distress (fear, sadness) and neutral faces. For arousal, participants rated how aroused each facial expression made them feel using the Self-Assessment Manikin (SAM; Bradley & Lang, 1994; Lang, 1980); note, SAM arousal ratings correlate highly with SCRs (Greenwald, Cook & Lang, 1989; Lang, Greenwald, Bradley & Hamm, 1993). Finally, to assess ability to tell apart

genuine from posed expressions, we used Dawel et al.'s (2017) -7 to +7 rating scale which centers around a neutral midpoint of 0 labelled "don't know", with positive numbers indicating an expression is perceived as showing a genuinely-felt experience of emotion and negative numbers indicating an expression is perceived as pretending emotion (Figure 1).

Psychopathic traits were assessed using three common measures: the Levenson Self-Report Psychopathy Scale (LSRPS; Levenson, Kiehl & Fitzpatrick, 1995); the Triarchic Psychopathy Measure (TriPM; Patrick, 2010); and the Inventory of Callous Unemotional traits (ICU; Frick, 2004). All three provide a measure of the affective component (LSRPS-F1 subscale; TriPM-meanness subscale; the ICU itself), enabling us to test the robustness of our results across different measures which were originally derived from somewhat different theoretical conceptualisations of psychopathy. The LSRPS and TriPM also measure other psychopathy components (LSRPS-F2 subscale; TriPM-disinhibition and -boldness subscales), enabling us to test the specificity of our findings to the affective component. All psychopathy scales were analysed as continuous variables, consistent with current conceptualisations of psychopathy as a continuum of traits distributed across the general population (e.g., Edens, Marcus, Lilienfeld & Polythress, 2006; Kimonis, Branch, Hagman, Graham & Miller, 2013). Where relevant, we confirmed relationships remained after accounting for covariates, comprising sex (which correlates with psychopathic traits), state and trait anxiety (which can correlate with arousal; e.g., Raskin, 1975), and depression (which affects perception of emotion authenticity for sad faces; Douglas et al., 2012).

METHOD

Participants

Data analysed were from 140 young adults (62% female, 38% male; *M* age = 20.6 years, *SD* = 3.3, range = 18-40) from the Australian National University community, recruited using a mix of general advertising plus flyers targeting individuals likely to have higher-than-average

levels of psychopathic traits (see Supplement S1 for details). All reported normal or corrected-to-normal vision. None reported autism spectrum disorder, ADD/ADHD, epilepsy, or other major neurological disorder. All were Caucasian to match the face stimuli (noting cultural differences in perception of emotional genuineness in faces; Thibault, Levesque, Gosselin, & Hess, 2012). Participants received course credit or \$22.50. The study was approved by the ANU Human Research Ethics Committee.

Session structure

Tasks were completed in a single 1.5 hr session in the following order: facial expression tasks (SAM arousal ratings, then intent-to-help ratings, then genuineness ratings); demographic questionnaire; psychopathy and covariate questionnaires.

Facial expression stimuli

Facial expression stimuli (Figure 1; details in Supplement 2) were photographs from genuine and posed sets recently collated and normed by Dawel et al. (2017). The genuine expressions were *elicited* in response to events where there was good reason to think the person displaying the expression would be feeling that emotion (e.g., a fear expression in response to a scary movie, or a sad expression displayed while recalling a sad memory) and, congruently, were *perceived* by the “average observer” in Dawel et al. (2017) to be showing genuinely-felt emotion.² Conversely, the posed expressions were elicited by posing instructions (e.g., to produce particular facial muscle configurations as in Ekman & Friesen, 1976) that were, congruently, perceived, on average, as showing “fake” (pretended) emotion.

Across our various tasks, we included all five emotions in the Dawel et al. (2017) set, including distress emotions (fear, sadness), plus non-distress emotions (happy, anger, disgust) (see Dawel et al., 2017 Supplement S3 for full details of stimuli). In Dawel et al. (2017), we

² There was some overlap ($N=93$) between present participants and “the average observer” study (Dawel et al., 2017, Experiment 3) used to validate the emotional genuineness of the present stimuli. Note Dawel et al. (2017) analysed genuineness ratings *only* at the group-average level, and did *not* analyse psychopathic traits nor intent-to-help nor arousal, nor the extra 47 participants added here.

described a four-stage pilot testing and validation process to put together, after a wide search of multiple databases and other images to provide potential genuine expressions, the best possible set of genuine expressions and matched posed expressions. “Best possible” was defined as the genuine expressions being most clearly perceived as genuine, and the posed expressions most clearly perceived as posed, in addition to: the genuine and posed sets being matched for sex, viewpoint (frontal looking towards the camera, or three-quarter looking away), and as closely as possible on rated intensity; and all emotions having good labelling accuracy (i.e., as showing fear, happy etc.). The final sets contain 98 fully validated stimuli (49 genuine, 49 posed) covering the 5 emotions we used here (fear, sadness, happiness, anger, disgust). Faces were shown in a rectangular frame with white background (Figure 1), sized 6.9° x 9.1° at viewing distance of 50 cm.

Facial expression ratings tasks

For the three face rating tasks, each trial presented a face in the centre of the screen until response. Each face was shown once per task, in a different random order for each participant. Full task instructions are in Supplements S3-S5. In all tasks, ratings for distress faces showed high reliability (Cronbach’s $\alpha > .8$; see Table 1).

Prosociality: Intent-to-help ratings. Participants were instructed to imagine each person (face) was someone they knew (e.g., a colleague at university or work) who had just had something bad happen to them, and rate how much they would want to be able to help them from 0 (*not at all*) to 7 (*very much*). Faces were all female: societal norms imply intent-to-help may be stronger for female than male faces in *low* affective psychopathy individuals, thus allowing maximum room for reductions in high-trait individuals to show up. Faces were: 6 genuine fear items (5 frontal, 1 three-quarter view) and 12 genuine sad items (6 frontal, 6 three-quarter), plus an equal number of viewpoint matched items for posed; and 10 neutral expressions (5 frontal, 5 three-quarter). Mean intensity of the expressions was matched across

genuine and posed distress ($M_{\text{genuine}} = 4.9$, $M_{\text{posed}} = 4.7$, on 0-9 scale; $t(34) = .31$, $p = .760$, data from Dawel et al., 2017). Neutral was included to assess whether any reduced intent-to-help was specific to the two distress emotions. Non-distress emotions were not included due to the implausibility of asking participants how much they wanted to help someone displaying happiness (given the "something bad has happened" context) or anger (which even low-psychopathic-trait participants would likely give very low intent-to-help ratings).

Arousal ratings using the Self-Assessment Manikin (SAM; Bradley & Lang, 1994).

Participants were instructed to “rate how aroused each facial expression makes you feel” using the SAM arousal scale. Responses were scored from 0 to 8, corresponding to the five “manikins” that make up the scale plus the four intermediate boxes; the completely unaroused manikin was scored 0, the box next to it 1, and so on (Supplement S4). There were 103 stimuli: 18 genuine distress (fear+sadness), 18 posed distress (fear+sad), 31 genuine non-distress (happy+anger+disgust), 31 posed non-distress (happy+anger+disgust), and 5 neutral.

Genuineness ratings (for correctly labelled expressions). For each face, participants first chose an emotion category label (anger, disgust, fear, happy, sad, neutral). Then, as long as “neutral” was not selected, the face stayed onscreen and the participant rated emotion genuineness, from -7 to $+7$ (Figure 1). Task instructions (identical to those in Dawel et al.’s 2017 Exp. 3) included a detailed explanation of the meaning of “genuine” (i.e., emotion genuinely-felt by the displayer) and “fake” (i.e., pretended). There were 108 stimuli in total: the 103 used in the arousal rating task, plus 5 extra neutral faces.

To calculate genuineness scores, we excluded trials where participants had incorrectly labelled the emotion (i.e., to avoid the situation where participants were, say, rating a sad stimulus item for the genuineness of its “disgust”). This excluded 15% of trials across fear and sad expressions (and fewer of the other expressions). Overall, labelling accuracy was high (fear genuine = 71%, posed = 83%; sad genuine = 88%, posed = 90%; see Supplemental Table

S6 for other emotions). Note, even participants at the upper end of our affective psychopathy range (i.e., upper quartile on a combined scale measure of affective psychopathy described below), who might be expected to have reduced labelling accuracy for distress expressions (Dawel et al., 2012; replicated here in Supplemental Table S6), still had sufficient correctly-labelled-genuineness-ratings trials to analyse (accuracy >75%).

Psychopathy questionnaire measures: LSRPS, TriPM, ICU, Affective Combined

Table 1 reports descriptive statistics for our three psychopathy questionnaire measures. All demonstrated acceptable reliability and good variation in our sample (consistent with previous findings in undergraduates; Kimonis, et al., 2013; Levenson et al., 1995; Patrick, 2010; Sellbom, 2011; Sellbom & Phillips, 2013). The *Levenson Self-Report Psychopathy Scale* (LSRPS; Levenson et al., 1995) comprises two subscales that measure F1 (16 items) and F2 (10 items) psychopathy; each item is rated from 1 (*disagree strongly*) to 4 (*agree strongly*). The *Triarchic Psychopathy Measure* (TriPM; Patrick, 2010; also cf. Sellbom & Phillips, 2013) comprises three subscales measuring meanness (19 items), boldness (19 items), and disinhibition (20 items); rating options for each item are *true—mostly true—mostly false—false* (scored from 1 to 4). The *Inventory of Callous-Unemotional Traits* (ICU; Frick, 2004) comprises 24 items that target callous and unempathic behaviour (e.g., “I do not care who I hurt to get what I want”) and unemotionality (“I seem very cold and uncaring to others”) rated from 0 (*not at all*) to 3 (*definitely true*).

The three measures of affective psychopathy were, as expected, moderately correlated ($r_s = .65, .64$, and $.45$ from Supplemental Table S7, with upper bound correlations of $.82-.86$ calculated from alpha values in Table 1). Additionally, correlations with the measures of interest (intent-to-help, arousal, genuineness discrimination) were generally similar across the three measures (see Results). We thus also created an *Affective Psychopathy Combined* score for each participant, by converting participants’ raw scores on each measure to z -scores, and

then averaging them together for each participant (i.e., $[z_{\text{LSRPS-F1}} + z_{\text{TriPM-mean}} + z_{\text{ICU}}]/3$).

Covariate questionnaire measures: State-trait anxiety, depression

We used the State-Trait Anxiety Inventory (STAI; Spielberger et al., 1983) to assesses state and trait anxiety (20 items each), from 1 (*not at all/almost never*) to 4 (*very much so/almost always*). Depression was assessed with the Depression Anxiety Stress Scale—depression subscale (DASS-depression; Lovibond & Lovibond, 1995), with 14 items that measure depression symptoms over the past week, from 0 (*never*) to 3 (*almost always*).

RESULTS

Initial analyses revealed very similar results for fear and sadness (Supplement S8), and thus, to maximise power, we focus on ratings for all distress items combined. Patterns of correlations were also highly similar across our three affective questionnaire measures (i.e., LSRPS-F1, TriPM-meanness, ICU); to streamline reporting, we focus on the Affective Psychopathy Combined measure (Figure 2 and Supplement S8 give results for individual measures). Analyses addressed *a priori* theory-derived predictions or research questions.

Is higher affective psychopathy associated with reduced intent-to-help?

To test the prediction that higher affective psychopathic traits should be associated with reduced *prosociality* towards others displaying distress, we correlated affective psychopathy scores with intent-to-help ratings separately for genuine, posed, and neutral expressions. Higher Affective Psychopathy Combined scores were significantly correlated with reduced intent-to-help ratings: for genuinely distress, $r(138) = -.467, p < .001$ (and $p < .001$ for each of the three individual affective measures, see Figure 2Ai); for posed distress, $r(138) = -.215, p = .011$; and also for neutral expressions, $r(138) = -.191, p = .023$ (consistent with previous findings of reduced *prosociality* outside the distress context; White, 2014). These correlations were significant and of equivalent magnitude for frontal and three-quarter viewpoint faces (Supplement S9), and were not due to our covariates: partial correlations with Affective

Psychopathy Combined, controlling for gender, state and trait anxiety and depression scores, were very similar in strength to the original Pearson's r values (original r s with partial r s in [] for comparison: genuine distress = $-.467$ [$-.440$], both $ps < .001$; posed distress = $-.215$ [$-.284$], both $ps < .05$; neutral = $-.191$ [$-.223$], both $ps < .05$).

Is this reduction in intent-to-help stronger for genuine than for posed distress?

Blair's theorising concerns how individuals high on affective psychopathic traits respond to others' *emotions*, and not to their facial expressions per se, which in everyday life sometimes display real emotion and sometimes only pretend emotion. Our next question was thus whether the relationship revealed above—of reduced intent-to-help with increasing affective psychopathy scores—is strongest when the face stimuli display genuine emotion; it was. For Figure 2Ai, a global GLM revealed a significant Affective Psychopathy Combined $\times$ expression type interaction (genuine distress, posed distress, neutral), Greenhouse-Geisser corrected $F(1.9, 258.3) = 4.10$, $MSE = .666$, $p = .020$, $\eta_p^2 = .029$. Further, the correlation with intent-to-help was significantly stronger for genuine distress than posed distress, at more than twice the strength (genuine $r = -.467$ vs. posed $r = -.215$; $z = 3.24$, $p < .001$ using Steiger's z -test for comparing dependent correlations; Steiger, 1980; Hoerger, 2013).

The plot of lines of best fit (Figure 2Aii) further clarifies the nature of the interaction. First, consider the low end of the affective psychopathy range on the left hand side of the graph. These individuals reported substantially greater intent-to-help people showing genuine distress relative to posed distress or neutral expressions, while posed distress elicits only a little more intent-to-help than neutral expressions. Then, consider how intent-to-help changes as affective psychopathy increases. While intent-to-help reduces to some extent for all expression types (genuine and posed distress, and neutral), the most striking reduction is for genuine distress. Indeed, at the highest end of our affective psychopathy range the lines of best fit indicate that genuine distress elicited only a little more intent-to-help than posed distress

(which again elicited only a little more intent-to-help than neutral expressions).

Affective psychopathy and intent-to-help: Mediated by arousal only for genuine distress?

Blair's (1995, 2005) theorising predicts a key mediating variable between affective psychopathy and reduced intent-to-help should be reduced *arousal* to distress. Figure 2D shows mediation model results, using Preacher and Hayes' (2008) macro (n bootstraps = 10,000). Outcomes supported the predicted mediating role of arousal in explaining the association between higher affective psychopathy and reduced intent-to-help—but *only in the case of genuine distress expressions*. For genuine distress, significant mediation by arousal was evident in the predicted, negative, direction, $b = -.065$, $CI_{95} [-.134, -.011]$. For posed distress, however, the small (but significant) association between affective psychopathy and intent-to-help was not mediated by arousal, $b = -.009$, $CI_{95} [-.076, .048]$. The same was true for neutral expressions, $b = -.026$, $CI_{95} [-.095, .012]$.

Bivariate associations between affective psychopathy and arousal cast further light on the difference between genuine and posed distress. Affective psychopathy showed a significant association with reduced arousal for genuine distress (Figure 2Bi; $r(138) = -.176$, $p < .038$ for Affective Psychopathy Combined), but not for posed distress ($r(138) = -.022$, $p = .794$). Lines of best fit (Figure 2Bii) indicate individuals at the low end of the affective psychopathy range found genuine distress more arousing than posed distress; those at the high end of the range reported genuine distress to be no more arousing than posed distress. Reduced arousal was specific to distress, with no correlation between affective psychopathy and arousal for the non-distress emotions (happy, anger, and disgust combined: genuine, $r(138) = -.028$, $p = .743$; posed, $r(138) = .067$, $p = .434$; see Supplement S10 for each emotion separately).

Affective psychopathy and reduced ability to perceive the authenticity of distress

One hypothesis concerning why individuals who are higher on affective psychopathy show reduced intent-to-help for genuine distress emotions is that, in addition to effects of reduced

arousal, they might also fail to properly *perceive* the authenticity or otherwise of the distress. Results supported this novel idea. To measure ability to tell apart genuine from posed distress, we used difference scores (genuineness rating for the face stimuli displaying genuine distress *minus* genuineness rating for stimuli displaying posed distress), where a lower score indicates poorer authenticity discrimination. Figure 2Ci shows higher Affective Psychopathy Combined was significantly correlated with poorer authenticity discrimination for distress, $r(138) = -.216, p = .010$ (also $p < .05$ for each affective measure separately). This association remained significant in partial correlation analyses accounting for sex, depression, state and trait anxiety, $r(138) = -.193, p = .024$, indicating atypical genuineness perception in depression (Douglas et al., 2012) cannot explain away our findings. The result was also specific to distress expressions. For non-distress emotions, Affective Psychopathy Combined showed no correlation with genuineness discrimination ($r(138) = -.055, p = .526$ for happy, anger, disgust combined; see Supplement S10 for each emotion separately).

In Figure 2Cii, we plot lines of best fit separately for genuine distress and posed distress. These show that reduced genuineness discrimination reflected a combination of rating genuine expressions as less genuine (i.e., ratings get closer to towards the central zero value, indicating "don't know", as affective psychopathy increases) and rating posed expressions as less fake (again, ratings get closer to central zero). Importantly, they also demonstrate that while discrimination is *reduced* with higher affective psychopathy, the ability to correctly recognise authenticity certainly did not disappear altogether: even at the highest end of our affective psychopathy range, highly significant genuineness discrimination remained (i.e., the 95% CIs on the two lines of best fit do not come close to overlapping), plus also the normal pattern remained in which genuine faces were rated as significantly genuine (i.e., 95% CI lines are *above* zero) and posed as significantly fake (i.e., 95% CI lines are *below* zero).

Together, these novel data on percepts of emotion genuineness are highly consistent with

the Blair's general idea that atypical processing of *distress* emotions plays a key etiological role in the affective aspects of psychopathy. Our data extend this idea, however, to add that this atypical processing includes a failure in high affective psychopathy to fully appreciate the authenticity of others' distress emotions (while being normal at appreciating authenticity of non-distress emotions).

Is authenticity discrimination an additional mediator of prosocial intent?

Our earlier mediation tests revealed arousal only partially mediated the association between higher affective psychopathy and reduced prosociality; that is, in Figure 2Di, a significant direct-effect link remains between Affective Psychopathy Combined and intent-to-help faces showing genuine distress ($b = -.499, p < .001$). Given the evidence of poorer authenticity discrimination, we tested whether this factor might be an additional mediator; it was. In a double-mediation model, we found independent mediating effects of arousal, $B = -.057$, $CI_{95} [-.131, -.014]$, and authenticity discrimination, $B = -.043$, $CI_{95} [-.118, -.004]$, on the association between Affective Psychopathy Combined scores and intent-to-help ratings for genuine distress (see Supplement S11 for model illustration, and details). However, even with both of these mediators included, there is still a significant direct link in the model between affective psychopathy and intent-to-help ($b = -.465$, $CI_{95} [-.638, -.292]$). This may mean that, in reality, there is a direct link. Alternatively, it may mean that there are one or more additional mediating variables yet to be identified.

Specificity of key findings to affective psychopathy

To assess the specificity of our findings to the affective features of psychopathic traits, we ran a series of multiple regressions to test whether there were any *independent* associations with affective psychopathy when other psychopathy components were included as predictors, and also whether those other psychopathy components themselves had any significant independent effect. Each regression entered (simultaneously) three predictor variables:

Affective Psychopathy Combined; a combined "secondary psychopathy" score derived from LSRPS-F2 and TriPM-disinhibition (i.e., $z_{\text{LSRPS-F2}} + z_{\text{TriPM-disinhibition}}/2$); and TriPM-boldness. The combined secondary psychopathy score was used because there is substantial theoretical overlap between the LSRPS-F2 and TriPM-disinhibition, and these scales correlated highly in our sample ($r(138) = .611$ with $r_{\text{upperbound}} = .746$).

Results (Table 2) demonstrate specificity to the affective component, for all face tasks. Affective psychopathy had significant independent effects, over and above other components, in all cases of relevance (i.e., genuine distress for intent-to-help and arousal; and genuine-versus-posed discrimination for distress). Concerning the other psychopathy factors (secondary psychopathy, boldness), although some of the simple *bivariate* correlations with intent-to-help and arousal (but not genuineness discrimination) were significant (Supplement S8), none of these associations survived once affective psychopathy was included as a predictor, indicating the other psychopathy components had no *independent* effect.

DISCUSSION

The present study provided the first test of association between higher affective psychopathy and reduced prosocial responses specifically in the context of distress facial expressions. Results show that, while prosocial behaviour (intent-to-help) was reduced even without distress (i.e., for neutral expressions), it was most strongly reduced for faces displaying authentic distress. We then tested the Blair-derived prediction that this should be mediated by reduced arousal to distress. We found a significant mediating role for arousal only for genuine distress expressions, and not for posed distress or neutral expressions. Overall then, Blair's ideas, which focus on *affective*—that is, *emotional*—processing in psychopathy, received stronger support when faces showed genuinely-felt emotions rather than only posed expressions (pretended emotion). Additionally, our results revealed a novel finding that people high on affective psychopathic traits misperceive the authenticity of other's distress emotions,

showing reduced ability to tell genuine from posed distress, as compared to individuals low on these traits. This reduced perception of authenticity also provided an additional mediator of the relationship between affective psychopathy and reduced prosocial behavior, contributing independently of reduced arousal.

Findings also demonstrated Blair's predicted pattern of specificity. This included specificity to the *affective* features of psychopathy (F1, meanness, CU), with only these, and not other psychopathy components (F2, disinhibition, boldness), having unique independent associations with reductions for distress in any of our face judgement tasks (intent-to-help, arousal, genuineness discrimination). It also included specificity to *distress emotions*, where we found affective psychopathy to correlate with arousal and genuineness discrimination, for distress emotions (fear, sad) but not other emotions (happy, anger, disgust).

Theoretically, we view our findings as compatible with, and providing an extension of, Blair's theorising. Blair's ideas predict reduced prosociality, mediation of this by reduced arousal, and specificity of these effects to affective psychopathy and to distress emotions; our results supported all these predictions. Additionally, our results extend Blair's theorising by arguing that his arousal-mediation idea applies only to others' *authentic* distress, and that one reason for this is that individuals high on affective psychopathic traits fail to fully appreciate, or experience enhanced arousal to, the authenticity of others' distress emotions (while being normal at appreciating authenticity of non-distress emotions).

Finally, our results also have implications for earlier versus later instantiations of Blair's ideas. In his early model (the VIM; Blair, 1995), failure of aversive arousal was included explicitly in the causal pathway leading to the development of psychopathic traits. In the more recent neurobiological instantiation (IES model, 2005), however, arousal is present only implicitly, via reduced amygdala responses. Our present results imply that explicit reference to arousal is worth reinstating in Blair's model, consistent with classic ideas of reduced arousal in

psychopathy (Patrick et al., 1994). Ideally, work could be done to test a more complete model that links neural activation with peripheral arousal, and also behaviour. For example, does reduced neural activation to others' distress (e.g., as evidenced by fMRI; Blair, 2013; Thornton et al., 2017; or EEG, Brislin et al., 2017) mediate the association between high affective psychopathic traits and reduced arousal? What contributions do neural activation and arousal then make to prosocial (and antisocial) behavioural outcomes?

Future directions concerning impaired authenticity perception for distress

Our novel finding of reduced ability to tell apart genuine from posed expressions of distress with higher affective psychopathy promotes several interesting lines of inquiry.

In addition to fear and sadness, pain is another important case of others' distress relevant to psychopathy (Wolf & Centifanti, 2014). Good face stimulus sets showing authentic versus pretended pain are available (Bartlett, Littlewort, Frank & Lee, 2014). It would be valuable to test whether individuals high on affective psychopathic traits show reduced ability to perceive the difference between others' genuine versus pretended pain, and whether, for example, this might mediate antisocial behaviour.

Another question concerns what might happen at the top-most end of affective psychopathy. The present study covered a wide range of trait levels (Table 2; scatterplot in Figure 2), but probably did not include many clinical-level “psychopaths”. Given that psychopathy is understood theoretically to be continuous within the general-population range (Edens et al., 2006), we expect true “psychopaths” (diagnosed, for example, via Hare’s PCL-R clinical interview procedure; Hare, 2003) would show an *exaggeration* of the pattern of findings observed here. For example, authenticity discrimination ability for distress emotions might disappear altogether (i.e., in Figure 2Cii, the lines would join up if extended out to the right, to more extreme affective psychopathy values).

Finally, a question of strong theoretical interest is what role poor authenticity perception

might play in the *development* of psychopathy. This might include using longitudinal studies to examine whether there is a causal, longitudinal role of poor authenticity perception in the emergence of callous-unemotional traits (the precursor to psychopathy) in children. For example, does reduced ability to recognise whether another child's distress is authentic, as opposed to just pretended in play, appear *early* in development? Does its appearance *precede* or *follow* the emergence of callous-unemotional traits across time in individual children?

Potential limitations: Use of ratings

In all tasks, we used rating measures. Addressing the validity of our conclusions from ratings is perhaps particularly important concerning *arousal*. Previous studies of arousal to distress in psychopathy have all used SCR, which assesses physiological arousal, and future corroboration of the present findings with physiological measures would be valuable. SCR assesses actual arousal, independent of the participant's conscious awareness of their arousal level. In contrast, ratings require participants to have conscious insight. It might be suggested that our reduced ratings of arousal to distress could be explained by high-affective psychopathy individuals underestimating their actual arousal relative to low-affective psychopathy individuals because of lack of insight, rather than because of reduced actual arousal. We argue, however, this is unlikely for several reasons: reduced physiological arousal (and reduced amygdala response) specific to distress is well established in affective psychopathy (e.g., Blair, 2013; Patrick et al., 1994); there seems no obvious reason why a lack of *awareness* of one's reduced physiological arousal would apply only to genuine distress, and not to posed distress nor to non-distress (happiness, anger, disgust), as would be needed to explain our results; and ratings on the SAM arousal task we used correlate well with SCR across the general population (Lang, et al., 1993; Greenwald et al., 1989).

Turning to our other tasks, distress specificity again supports validity. For *genuineness ratings*, our finding that high-affective psychopathy individuals rated genuine distress

emotions as less genuine and posed distress emotions as less fake cannot be attributed to, say, being simply less willing to use numbers further away from zero on the rating scale than low-affective psychopathy. This would predict a correlation with reduced genuineness discrimination for all emotions, which did not occur (i.e., no reduction for happy, anger, disgust). And, for *intent-to-help*, the reduction in ratings for authentic distress can also not be attributed merely to general changes in use of the scale, given the reduction for authentic distress was significantly stronger than for posed distress and for neutral (and also that for posed distress, ratings at the high-affective psychopathy end of our range remained comfortably above the scale minimum, i.e., mean rating of approximately 2.2 where scale minimum is 0, with plenty of room to give lower ratings had participants wished to).

Broader implications: The benefits of using genuine emotion face stimuli

Our findings highlight the theoretical and practical benefits to researchers of choosing facial expression stimuli that show genuinely-felt, authentic emotion. The posed facial expressions available in many popular databases (e.g., RaFD, Langner et al., 2010; Pictures of Facial Affect, Ekman & Friesen, 1976) are widely used in affective and clinical research (e.g., depression, Douglas et al., 2012; psychopathy, Dawel, et al., 2012; schizophrenia, Russell et al., 2007; social anxiety, Kivity & Huppert, 2016; amygdala damage, Adolphs et al., 1999; empathy, Besel & Yuille, 2010), but are commonly perceived as showing pretended rather than real emotion (Dawel et al., 2017). In contrast, a focus on using expressions of naturalistic, genuine emotion is present in some research traditions (notably arousal research, e.g., using the IAPS stimuli showing crying babies in psychopathy arousal studies, Blair et al., 1997). Our results argue there would be value in extending more broadly this focus on genuine emotion.

In the present study, the core theoretical idea that reduced prosocial behaviour in affective psychopathy should be mediated by reduced arousal to distress (Blair, 1995, 2005) was supported *only* using genuine distress expressions and not using posed ones. Indeed, had we

used only posed facial expression stimuli, we would have been using our results to, wrongly, argue against Blair's theory. This joins a previous case where, in schizophrenia, a common theoretical view that impaired affective processing is a core feature of the disorder (e.g., Kohler, Walker, Martin, Healey & Moberg, 2009) has been supported by studies testing ability to label posed expressions, but controverted by studies using genuine expressions where *superior* emotion recognition was observed in schizophrenia relative to healthy controls (Davis & Gibson, 2000; La Russo, 1978). Overall, it is clear that theoretical conclusions reached using genuine emotion stimuli can differ substantively from those reached with posed expressions.

Our results also show testing both genuine *and* posed expressions can sometimes cast light on interesting issues, and highlight that researchers using facial expression stimuli need to be cognisant of the different social meanings of expressions conveyed by expressions showing genuine versus pretended emotion. For example, although observers might agree the label “sad” applies equally to two different expression stimuli—one genuine, one posed—the *social message* being sent by these expressions can be quite different, such as “I’m in distress, I need help” versus “I’m feigning sadness to elicit your sympathy”. Our present results argue that ability to read these messages varies. Observers low on affective psychopathic traits (who would also be expected to be high in empathy), can correctly read the social signal, which in turn affects their intended behaviour: specifically, they desire to help others in genuine distress substantially more than others merely pretending distress. In contrast, individuals high on affective psychopathy appear to fail to respond appropriately to the different social message communicated by authentic relative to posed distress.

Conclusion

The present study is the first to show that higher levels of affective psychopathy are associated with reduced prosocial intent specifically in the context of others’ distress expressions, and that arousal mediates this association—but only when faces display genuine

distress. For posed distress, the evidence was much weaker and only supported a bivariate association between affective psychopathy and prosocial intent, with no mediating role for arousal. The present study is also the first to examine percepts of emotion authenticity in psychopathy, finding that ability to tell genuine from posed distress is impaired, and that this impairment was an additional, independent mediator of the association between affective psychopathy and prosocial intent. All findings demonstrated specificity to the affective features of psychopathy and to distress expressions. Overall, results were highly consistent with Blair's theorising about how the affective features of psychopathy arise, and demonstrate the value of using facial expression stimuli that show genuine emotion in research on affective processing. Finally, concerning normal processing of social signals, a novel finding was that individuals with *low* levels of affective psychopathy reported greater intent-to-help and arousal towards people showing genuine relative to posed distress.

Acknowledgements

Research funded by: Australian Research Council Discovery Project grant to E.M. (DP110100850); Australian Research Council Centre of Excellence for Cognition and Its Disorders (CE110001021) www.ccd.edu.au. Tamara Gradden tested some participants, Dr Rachel Robbins helped with data screening, Dr Martin Sellbom provided recruitment advice.

References

- Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., ... & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologia*, 37(10), 1111-1117. doi:10.1016/S0028-3932(99)00039-1
- Anastassiou-Hadjicharalambous, X., & Warden, D. (2008). Physiologically-indexed and self-perceived affective empathy in conduct-disordered children high and low on callous-unemotional traits. *Child Psychiatry & Human Development*, 39(4), 503-517. doi:10.1007/s10578-008-0104-y
- Bartlett, M. S., Littlewort, G. C., Frank, M. G., & Lee, K. (2014). Automatic decoding of facial movements reveals deceptive pain expressions. *Current Biology*, 24(7), 738-743. doi:10.1016/j.cub.2014.02.009
- Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of altruistic motivation? *Journal of personality and Social Psychology*, 40(2), 290. doi:10.1037/0022-3514.40.2.290
- Besel, L. D., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences*, 49(2), 107-112. doi:10.1016/j.paid.2010.03.013
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1-29. doi:10.1016/0010-0277(95)00676-P
- Blair, R. J. R. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Personality and Individual Differences*, 27(1), 135-145. doi:10.1016/S0191-8869(98)00231-1
- Blair, R. J. R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Development and psychopathology*, 17(3), 865-891. doi:10.1017/S0954579405050418

- Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience*, 14(11), 786. doi:10.1038/nrn3577
- Blair, R. J. R. (2017). Emotion-based learning systems and the development of morality. *Cognition*, 167, 38-45. doi:10.1016/j.cognition.2017.03.013
- Blair, R. J. R., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 34(2), 192-198. doi:10.1111/j.1469-8986.1997.tb02131.
- Blampied, M., Johnston, L., Miles, L. K., & Liberty, K. (2010). Sensitivity to differences between enjoyment and non-enjoyment smiles in children with autism spectrum disorder. *British Journal of Developmental Psychology*, 28, 483–489. doi:10.1348/026151009X467621
- Boraston, Z. L., Corden, B., Miles, L. K., Skuse, D. H., & Blakemore, S. J. (2008). Brief report: Perception of genuine and posed smiles by individuals with autism. *Journal of Autism and Developmental Disorders*, 38(3), 574–580. doi:10.1007/s10803-007-0421-1
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. doi:10.1016/0005-7916(94)90063-9
- Brislin, S. J., Yancey, J. R., Perkins, E. R., Palumbo, I. M., Drislane, L. E., Salekin, R. T., ... & Patrick, C. J. (2017). Callousness and Affective Face Processing in Adults: Behavioral and Brain-Potential Indicators. Advance online publication. doi:10.1037/per0000235
- Davis, P. J., & Gibson, M. G. (2000). Recognition of posed and genuine facial expressions of emotion in paranoid and nonparanoid schizophrenia. *Journal of Abnormal Psychology*, 109(3), 445. doi:10.1037/0021-843X.109.3.445
- Dawel, A., O’Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience & Biobehavioral Reviews*, 36(10), 2288-2304.

doi:10.1016/j.neubiorev.2012.08.006

- Dawel, A., Wright, L., Irons, J., Dumbleton, R., Palermo, R., O’Kearney, R., & McKone, E. (2017). Perceived emotion genuineness: normative ratings for popular facial expression stimuli and the development of perceived-as-genuine and perceived-as-fake sets. *Behavior Research Methods*, 1-24. doi:10.3758/s13428-016-0813-2
- Douglas, K. M., Porter, R. J., & Johnston, L. (2012). Sensitivity to posed and genuine facial expressions of emotion in severe depression. *Psychiatry Research*, 196(1), 72-78. doi:10.1016/j.psychres.2011.10.019
- Edens, J. F., Marcus, D. K., Lilienfeld, S. O., & Poythress Jr, N. G. (2006). Psychopathic, not psychopath: Taxometric evidence for the dimensional structure of psychopathy. *Journal of Abnormal Psychology*, 115(1), 131. doi:10.1037/0021-843X.115.1.131
- Ekman, P., & Friesen, W. V. Pictures of facial affect. (1976). *Palo Alto: Consulting Psychologists*.
- Frick, P. J. (2004). The inventory of callous-unemotional traits. *Unpublished rating scale*.
- Frick, P. J., & White, S. F. (2008). Research review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry*, 49(4), 359-375. doi:10.1111/j.1469-7610.2007.01862.x
- Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology*, 3(1), 51-64.
- Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., ... & Gur, R. E. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of Neuroscience Methods*, 115(2), 137-143. doi:10.1016/S0165-0270(02)00006-7
- Hare, R. D. (2003). *The Psychopathy Checklist–Revised*. Toronto: ON.

- Hare, R. D., Harpur, T. J., Hakstian, A. R., Forth, A. E., Hart, S. D., & Newman, J. P. (1990). The revised Psychopathy Checklist: Reliability and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(3), 338. doi:10.1037/1040-3590.2.3.338
- Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach*. New York: Guilford.
- Hein, G., Lamm, C., Brodbeck, C., & Singer, T. (2011). Skin conductance response to the pain of others predicts later costly helping. *PloS One*, 6(8), e22759. doi:10.1371/journal.pone.0022759
- Hoerger, M. (2013). ZH: An updated version of Steiger's Z and web-based calculator for testing the statistical significance of the difference between dependent correlations. *Retrieved March 11, 2014*.
- Ibáñez MI, Sabater-Grande G, Barreda-Tarrazona I, Mezquita L, López-Ovejero S, Villa H, Perakakis P, Ortet G, García-Gallego A and Georgantzis N (2016). Take the Money and Run: Psychopathic Behavior in the Trust Game. *Frontiers in Psychology*, 7:1866. doi:10.3389/fpsyg.2016.01866
- Kimonis, E. R., Branch, J., Hagman, B., Graham, N., & Miller, C. (2013). The psychometric properties of the Inventory of Callous–Unemotional Traits in an undergraduate sample. *Psychological Assessment*, 25(1), 84. doi:10.1037/a0029024
- Kivity, Y., & Huppert, J. D. (2016). Emotional reactions to facial expressions in social anxiety: a meta-analysis of self-reports. *Emotion Review*, 8(4), 367-375. doi:10.1177/1754073915594436
- Kochanska, G., Barry, R. A., Jimenez, N. B., Hollatz, A. L., & Woodard, J. (2009). Guilt and effortful control: Two mechanisms that prevent disruptive developmental trajectories. *Journal of Personality and Social Psychology*, 97(2), 322. doi:10.1037/a0015471
- Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2009). Facial emotion perception in schizophrenia: a meta-analytic review. *Schizophrenia Bulletin*, 36(5),

1009-1019. doi:10.1093/schbul/sbn192

Krebs, D. (1975). Empathy and altruism. *Journal of Personality and Social Psychology*, 32(6), 1134. doi:10.1037/0022-3514.32.6.1134

Kreis, M. K., Cooke, D. J., Michie, C., Hoff, H. A., & Logan, C. (2012). The Comprehensive Assessment of Psychopathic Personality (CAPP): content validation using prototypical analysis. *Journal of Personality Disorders*, 26(3), 402-413. doi:10.1521/pedi.2012.26.3.402

Krumhuber, E. G., Likowski, K. U., & Weyers, P. (2014). Facial mimicry of spontaneous and deliberate Duchenne and non-Duchenne smiles. *Journal of Nonverbal Behavior*, 38(1), 1-11. doi:10.1007/s10919-013-0167-8

Lang, P. J. (1980). Self-assessment manikin. *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*.

Lang, P. J., & Greenwald, M. K. (1988). The international affective picture system standardization procedure and initial group results for affective judgments: Technical report 1A. *The Center for Research in Psychophysiology, University of Florida*.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, 30(3), 261-273. doi:10.1111/j.1469-8986.1993.tb03352.x

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8), 1377-1388. doi:10.1080/02699930903485076

LaRusso, L. (1978). Sensitivity of paranoid patients to nonverbal cues. *Journal of Abnormal Psychology*, 87(5), 463. doi:10.1037/0021-843X.87.5.463

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. doi:10.1037/0022-3514.68.1.151

- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3), 335-343. doi:10.1016/0005-7967(94)00075-U
- Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*.
- Marsh, A. A., & Ambady, N. (2007). The influence of the fear facial expression on prosocial responding. *Cognition and Emotion*, 21(2), 225-247. doi:10.1080/02699930600652234
- Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W., & Cardinale, E. M. (2014). Neural and cognitive characteristics of extraordinary altruists. *Proceedings of the National Academy of Sciences*, 111(42), 15036-15041. doi:10.1073/pnas.1408440111
- Miller, J. D., & Lynam, D. (2001). Structural models of personality and their relation to antisocial behavior: A meta-analytic review. *Criminology*, 39(4), 765-798. doi:10.1111/j.1745-9125.2001.tb00940.x
- Nichols, S. (2001). Mindreading and the cognitive architecture underlying altruistic motivation. *Mind & Language*, 16(4), 425-455. doi:10.1111/1468-0017.00178
- Patrick, C. J. (2010). Operationalizing the triarchic conceptualization of psychopathy: Preliminary description of brief scales for assessment of boldness, meanness, and disinhibition. *Unpublished manual, Department of Psychology, Florida State University, Tallahassee, FL*.
- Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and psychopathology*, 21(03), 913-938. doi:10.1017/S0954579409000492

- Patrick, C. J., Cuthbert, B. N., & Lang, P. J. (1994). Emotion in the criminal psychopath. *Journal of Abnormal Psychology, 103*(3), 523-534. doi: 10.1037/0021-843X.103.3.523
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891. doi:10.3758/BRM.40.3.879
- Raskin, M. (1975). Decreased skin conductance response habituation in chronically anxious patients. *Biological Psychology, 2*, 309-319. doi:10.1016/0301-0511(75)90039-3
- Righi, G., Peissig, J. J., & Tarr, M. J. (2012). Recognizing disguised faces. *Visual Cognition, 20*(2), 143-169. doi:10.1080/13506285.2012.654624
- Russell, T. A., Reynaud, E., Kucharska-Pietura, K., Ecker, C., Benson, P. J., Zelaya, F., ... & Phillips, M. L. (2007). Neural responses to dynamic expressions of fear in Schizophrenia. *Neuropsychologia, 45*(1), 107-123. doi:10.1016/j.neuropsychologia.2006.04.026
- Sakai, J. T., Dalwani, M. S., Gelhorn, H. L., Mikulich-Gilbertson, S. K., & Crowley, T. J. (2012). A behavioral test of accepting benefits that cost others: associations with conduct problems and callous-unemotionality. *PLoS One, 7*(4), e36158. doi:10.1371/journal.pone.0036158
- Sakai, J. T., Dalwani, M. S., Mikulich-Gilbertson, S. K., McWilliams, S. K., Raymond, K. M., & Crowley, T. J. (2016). A behavioral measure of costly helping: replicating and extending the association with callous unemotional traits in male adolescents. *PloS one, 11*(3), e0151678. doi:10.1371/journal.pone.0151678
- Sellbom, M. (2011). Elaborating on the construct validity of the Levenson Self-Report Psychopathy Scale in incarcerated and non-incarcerated samples. *Law and Human Behavior, 35*(6), 440-451. doi:10.1007/s10979-010-9249-x
- 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. doi:10.1037/a0029306

- Shirtcliff, E. A., Vitacco, M. J., Graf, A. R., Gostisha, A. J., Merz, J. L., & Zahn-Waxler, C. (2009). Neurobiology of empathy and callousness: implications for the development of antisocial behavior. *Behavioral Sciences & the Law*, 27(2), 137-171. doi:10.1002/bsl.862
- Sigman, M. D., Kasari, C., Kwon, J. H., & Yirmiya, N. (1992). Responses to the negative emotions of others by autistic, mentally retarded, and normal children. *Child Development*, 63(4), 796-807. doi:10.1111/j.1467-8624.1992.tb01662.x
- Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory STAI (form Y) ("self-evaluation questionnaire").
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245. doi:10.1037/0033-2909.87.2.245
- Thibault, P., Levesque, M., Gosselin, P., & Hess, U. (2012). The Duchenne marker is not a universal signal of smile authenticity—but it can be learned!. *Social Psychology*, 43, 215-221. doi:10.1027/1864-9335/a000122
- Thornton, L. C., Penner, E. A., Nolan, Z. T., Adalio, C. J., Sinclair, S., Meffert, H., ... White, S. F. (2017). The processing of animacy information is disrupted as a function of callous-unemotional traits in youth with disruptive behavior disorders. *NeuroImage: Clinical*, 16, 498-506. doi:10.1016/j.nicl.2017.08.024
- Verona, E., Patrick, C. J., Curtin, J. J., Bradley, M. M., & Lang, P. J. (2004). Psychopathy and physiological response to emotionally evocative sounds. *Journal of Abnormal Psychology*, 113(1), 99. doi:10.1037/0021-843X.113.1.99
- White, B. A. (2014). Who cares when nobody is watching? Psychopathic traits and empathy in prosocial behaviors. *Personality and Individual Differences*, 56, 116-121. doi:10.1016/j.paid.2013.08.033
- Wolf, S., & Centifanti, L. C. M. (2014). Recognition of pain as another deficit in young males with high callous-unemotional traits. *Child Psychiatry & Human Development*, 45(4), 422-432. doi:10.1007/s10578-013-0412-8

A. Genuine distress expressions

- Expressions elicited by emotion-inducing events, e.g., recalling a sad event, or watching a scary or sad movie (e.g., for sad, the scene in "My Girl" where boy dies).

Genuine fear

Genuine sad

B. Posed distress expressions

- Expressions elicited by instructions to 'pose sad/fear', or configure facial muscles in a particular way (e.g., 'widen eyes', 'turn corners of your mouth down', etc.) with no reference to the name of the target emotion.

Posed fear

Posed sad

Figure 1. Genuine and posed expression sets used in the present study, showing example stimuli for the distress emotions (fear and sad). Inclusion in genuine or posed sets was based on meeting criteria for both elicitation method (i.e., how the face expressions images were obtained from displayers) *plus* how the expression is perceived by 'the average observer' (i.e., the mean genuineness rating from n=94 undergraduates, Dawel et al, 2017). Precise elicitation method varies across databases from which images collated. Database source shown in italics: *FacePlace* = spontaneous expressions elicited while watching movies (Righi, Peissig & Tarr, 2012); *Gur-M* = method-acted expressions (referred to as 'evoked' in Gur et al., 2002); *RaFD* = people asked to pose specific facial muscle configurations (Langner et al., 2010); *KDEF* = actors given description of expression and asked to pose (Lundqvist, Flykt & Ohman, 1998). All faces were female. To view images at the same size and resolution as the participants, expand in electronic version to 7.1 cm tall and view from 50 cm distance.

Figure 2. Results. A-C. Direct associations between affective psychopathy scores and facial expression ratings. (i) Top panels show bivariate correlations (Pearson's r , except TriPM = Spearman's ρ) between affective psychopathy measures and each participants' mean rating across items for distress expressions (sad & fear combined) for **A.** intent-to-help, **B.** arousal ratings, and **C.** genuineness *discrimination* scores (i.e., genuine-minus-posed difference scores; not the raw genuine and posed mean rating scores).

(ii) Bottom panels show lines of best fit for affective psychopathy combined scores; shaded regions above and below show 95% CI on the line of best fit. **D. Arousal mediation models for intent-to-help.** Models use Preacher and Hayes' (2008) macro (this method was preferred over the Sobel test because it has greater power and fewer assumptions about distribution shape; Hayes, 2013). Path strength coefficients are unstandardised weights. As the bootstrapping method does not produce an exact p -value for indirect paths, these are interpreted as significant when 95% confidence intervals do not cross zero. All significant paths are bolded. * $p < .05$. ** $p < .01$. *** $p < .001$. All unmarked correlation coefficient bars are non-significant at $p > .05$.

Table 1

Descriptive statistics for face rating tasks and questionnaire measures (N = 140 participants)

Measure	M (SD)	Score range across participants	Scale range	Cronbach's α
Face rating tasks ¹				
Intent-to-help ratings				
Genuine distress	4.3 (1.0)	1.4-6.9	0 to +7	.89
Posed distress	2.8 (1.1)	0.3-6.7	0 to +7	.93
SAM arousal ratings				
Genuine distress	3.9 (1.2)	0.1-6.3	0 to +8	.90
Posed distress	3.2 (1.3)	0.1-6.4	0 to +8	.93
Genuineness ratings				
Genuine distress	3.6 (1.6)	-2.2-6.7	-7 to +7	.81
Posed distress	-4.2 (1.5)	-7.0-0.3	-7 to +7	.82
Questionnaires				
Affective psychopathy subscales				
LSRPS-F1	29.3 (7.0)	16-51	16-64	0.84
TriPM-meanness	31.5 (8.4)	19-56	19-76	0.88
ICU	23.8 (5.9)	13-42	0-72	0.81
Other psychopathy subscales				
LSRPS-F2	20.6 (4.4)	12-31	10-40	0.67
TriPM-disinhibition	36.9 (8.5)	22-62	20-80	0.83
TriPM-boldness	48.4 (9.4)	20-72	19-76	0.85
Covariate questionnaire measures				
STAI-state	36.6 (10.4)	20-68	20-80	0.92
STAI-trait	43.5 (10.9)	20-75	20-80	0.93
DASS-depression	8.5 (7.1)	0-37	0-42	0.93

Note. ¹In the ratings tasks, high alpha values demonstrate high internal consistency in ratings across face items

Table 2

Multiple regression models showing unique contributions of affective features, but not other components, of psychopathy.

	Intent-to-help				Arousal				Genuineness	
	Genuine		Posed		Genuine		Posed		Discriminability	
	r_{sp}	p	r_{sp}	p	r_{sp}	p	r_{sp}	p	r_{sp}	p
Affective Psychopathy Combined	-.361	<.001	-.116	.166	-.198	.019	-.054	.520	-.211	.013
Secondary Psychopathy	-.018	.809	-.065	.440	.132	.114	.118	.160	.074	.378
TriPM-Boldness	.050	.514	-.072	.392	-.067	.425	-.155	.067	.010	.907

Notes. Affective Psychopathy Combined scores = $(z_{\text{LSRPS-F1}} + z_{\text{TriPM-meanness}} + z_{\text{ICU}})/3$. Secondary Psychopathy scores = $(z_{\text{LSRPS-F2}} + z_{\text{TriPM-disinhibition}} + z_{\text{ICU}})/3$. r_{sp} = semi-partial correlations. Note the key result (independent effect only of affective aspects) was replicated in a four-predictor model variant entering the two secondary psychopathy scales separately