Original Article

Corresponding author:
Peter Fonagy;
Email: p.fonagy@ucl.ac.uk

CAMBRIDGE
UNIVERSITY PRESS

The mediating role of reflective functioning and general psychopathology in the relationship between childhood conduct disorder and adult aggression among offenders

Karen Yirmiya[1,2] , Matthew Constantinou[1,2], Elizabeth Simes[1,2], Anthony Bateman[1,2], James Wason[3], Jessica Yakeley[4], Mary McMurran[5], Mike Crawford[6], Alison Frater[7], Paul Moran[8], Barbara Barrett[9], Angus Cameron[10], Zoe Hoare[11], Elizabeth Allison[1,2], Stephen Pilling[1], Stephen Butler[1,2] and Peter Fonagy[1,2]

[1]Research Department of Clinical, Educational and Health Psychology, University College London, London, UK; [2]Anna Freud National Centre for Children and Families, London, UK; [3]Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK; [4]Portman Clinic, Tavistock and Portman NHS Foundation Trust, London, UK; [5]Institute of Mental Health, University of Nottingham, Nottingham, UK; [6]Division of Psychiatry, Imperial College, London, UK; [7]School of Law, Royal Holloway, University of London, London, UK; [8]Centre for Academic Mental Health, Population Health Sciences Department, Bristol Medical School, University of Bristol, Bristol, UK; [9]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [10]National Probation Service London Division, London, UK and [11]NWORTH Clinical Trials Unit, School of Health Sciences, Bangor University, Bangor, UK

Abstract

Background. The nature of the pathway from conduct disorder (CD) in adolescence to antisocial behavior in adulthood has been debated and the role of certain mediators remains unclear. One perspective is that CD forms part of a general psychopathology dimension, playing a central role in the developmental trajectory. Impairment in reflective functioning (RF), i.e., the capacity to understand one's own and others' mental states, may relate to CD, psychopathology, and aggression. Here, we characterized the structure of psychopathology in adult male-offenders and its role, along with RF, in mediating the relationship between CD in their adolescence and current aggression.

Methods. A secondary analysis of pre-treatment data from 313 probation-supervised offenders was conducted, and measures of CD symptoms, general and specific psychopathology factors, RF, and aggression were evaluated through clinical interviews and questionnaires.

Results. Confirmatory factor analyses indicated that a bifactor model best fitted the sample's psychopathology structure, including a general psychopathology factor (p factor) and five specific factors: internalizing, disinhibition, detachment, antagonism, and psychoticism. The structure of RF was fitted to the data using a one-factor model. According to our mediation model, CD significantly predicted the p factor, which was positively linked to RF impairments, resulting in increased aggression.

Conclusions. These findings highlight the critical role of a transdiagnostic approach provided by RF and general psychopathology in explaining the link between CD and aggression. Furthermore, they underscore the potential utility of treatments focusing on RF, such as mentalization-based treatment, in mitigating aggression in offenders with diverse psychopathologies.

Introduction

Conduct disorders (CD) represent some of the most pervasive mental and behavioral disturbances, observed in 3–5% of children and adolescents globally (Erskine et al., 2017; Sadler et al., 2018). The role of CD symptoms in precipitating antisocial aggressive behavior in adulthood has been debated within the developmental psychopathology literature (Fairchild et al., 2019). CD serves as a criterion for diagnosing antisocial personality disorder (ASPD), a persistent pattern of disregard for and violation of others' rights. This pattern suggests some homotypic continuity in antisocial difficulties from childhood to adulthood. However, relatively high proportions of children and adolescents diagnosed with CD do not exhibit ASPD in adulthood (Copeland, Shanahan, Costello, & Angold, 2009). Instead, a significant number have been found to develop other disorders, including depression and anxiety (Kim-Cohen et al., 2003). Indeed, the prognostic significance of CD for ASPD and aggressive

behavior remains unclear because CD may overlap with broader risk factors that are more predictive of poorer mental health (Krueger, Markon, Patrick, & Iacono, 2005).

Existing longitudinal and cross-sectional studies support the claim that CD is part of a broader dimension of externalizing behavioral problems (Fairchild et al., 2019), which in turn are part of a broader continuum of mental ill health. Researchers have highlighted the high prevalence of comorbidity among psychiatric diagnoses of various kinds, leading to the identification of a latent general psychopathology factor, colloquially termed the 'p factor' (Caspi et al., 2014; Caspi, Houts, Fisher, Danese, & Moffitt, 2023). This factor is posited to play a key role in the emergence of various mental health difficulties (Caspi & Moffitt, 2018). The p factor can be estimated using different factor analytic models that have different theoretical implications, but are statistically similar in practice (Hyland et al., 2021; van Bork, Epskamp, Rhemtulla, Borsboom, & van der Maas, 2017), in which the overlap in item responses on self-report measures is partitioned into that which all items share (i.e. a general factor), and that which subsets of items have in common, after considering the shared variance (i.e. specific factors). When applied to broadband psychopathology measures, the general p factor reflects individual differences in overall vulnerability to, or severity of, psychopathology (Caspi & Moffitt, 2018), whilst specific internalizing and externalizing factors reflect unique styles of expressing difficulties (Caspi et al., 2014).

The principles of the p factor have been extrapolated to CD research, advocating for the application of developmental-ontogenetic and hierarchical-dimensional models to investigate psychopathic traits, externalizing problems, and CD more comprehensively (Hoffmann et al., 2023; Perkins et al., 2022). Grounded in these theoretical considerations, we propose that the relationship between CD in adolescence and subsequent aggressive behavior in adult male offenders is mediated by the general psychopathology factor.

Another potential mediator between adolescent CD and adulthood aggressive behavior is impairment in reflective functioning (RF). RF, referring to an individual's capacity to understand their own and others' mental states – including emotions, desires, intentions, and beliefs – might at least partly explain the link between psychopathology and aggression (Fonagy & Target, 1997; Fonagy, Target, Steele, & Steele, 1998; Luyten, Campbell, Allison, & Fonagy, 2020). RF serves as the operationalization of mentalization theory (Fonagy et al., 2022; Fonagy & Target, 1997) and impairments in this function can lead to disrupted recognition and regulation of emotions, as well as difficulties comprehending and interpreting empathetic responses, resulting in impulsivity and aggression. The role of mentalization, or RF, as a mediator of the association between psychopathology or personality traits and behavioral outcomes related to violence and aggression, has been highlighted in previous studies (Bo, Sharp, Lind, Simonsen, & Bateman, 2023; Parada-Fernández, Herrero-Fernández, & Rodríguez-Arcos, 2023; Taubner, White, Zimmermann, Fonagy, & Nolte, 2013). Impairments in mentalization have been studied in men who offended with ASPD, and while the effectiveness of mentalization-based treatment (MBT) interventions for PD is recognized, the literature on MBT for adult who offended remains limited (Bateman, O'Connell, Lorenzini, Gardner, & Fonagy, 2016; Newbury-Helps, Feigenbaum, & Fonagy, 2017; Thomas & Jenkins, 2019).

In this nested study, our primary aim was to investigate the role of RF impairment and general psychopathology as potential mediators of the association between CD in adolescence and adult aggression among male offenders. By building upon prior research on RF as a marker of resilience (Bo et al., 2023; Kealy, Rice, Seidler, Oliffe, & Ogrodniczuk, 2021; Parada-Fernández et al., 2023; Taubner et al., 2013), we aimed to elucidate the mechanisms that could be instrumental in reducing aggressive behaviors and diminishing the stigma associated with individuals who have a history of CD. As not all individuals with CD develop into aggressive adults (Fairchild et al., 2019), it is crucial to explore factors that contribute to this heterogeneity. Our study utilizes baseline data from participants recruited for the Mentalization for Offending Adult Males (MOAM) trial (Fonagy et al., 2020), before any intervention took place. Two main objectives were addressed: First, we sought to evaluate the structure of psychopathology in men who offended, and second, to determine the role of their current psychopathology and RF in mediating the relationship between CD in adolescence and aggressive behavior. We hypothesized that in addition to a direct pathway from CD to aggressive behavior, CD will also predict general psychopathology, which in turn will predict aggression via RF. In such a model, RF would mediate the impact of psychopathology, as represented by the p factor, on aggression, with individuals exhibiting lower levels of RF demonstrating a greater propensity for aggression. Alternative hypotheses were that CD alone contributes to reduced RF (irrespective of the p factor), or that the p factor completely mediates the relationship between CD and aggression (without the involvement of RF). To supplement these hypotheses, we conducted a further analysis to determine the specific factors contributing to the formation of the general p factor as a sensitivity analysis. By examining these hypotheses, we aimed to determine the interplay between CD, RF, the general psychopathology factor, and their contribution to aggressive tendencies in adult who offended.

Methods

Participants

This study conducted a secondary analysis of pre-treatment data from 313 adult male offenders under the National Probation Service (NPS) supervision. All participants met the criteria for personality difficulties and were invited to join the MOAM trial. The MOAM trial, a multisite randomized controlled study in an NHS setting, assessed the efficacy and cost-effectiveness of integrating MBT with probation as usual (PAU) against PAU alone (Fonagy et al., 2020). The sample's sociodemographic details are in Table 1.

The trial evaluated MBT's role in diminishing aggressive and antisocial behaviors and enhancing health status and quality of life. To qualify, participants were male, 21 or older, supervised by NPS with at least six months left on their sentence. They fulfilled the DSM-IV criteria for ASPD, scored ≥15 on the Modified Overt Aggression Scale (OAS-M; Endicott, Tracy, Burt, Olson, and Coccaro, 2002), and had adequate English and cognitive skills for trial processes. Those convicted of child sexual offenses or primarily diagnosed with psychotic or neurodevelopmental disorders were excluded. Out of 364 potential participants, 313 were recruited due to refusal or imprisonment. All gave written consent, with approvals from the London–South East Research Ethics Committee (reference: 14/LO/1696) and the National Offender Management Service (reference: 2014-315). Assessment methods

Table 1. Sociodemographic and main study variables

	N = 313 (%)	M	S.D.
Age		34.24	9.27
Education			
O levels / CSEs / GSCEs school certificate	41 (16.4)		
A / AS levels / Higher school certificate	12 (4.8)		
NVQ / GNVQ HCN / HND	186 (74.4)		
Academic degrees	11 (4.4)		
Marital status			
Never married / separated / divorced / annulled	288 (92.3)		
Married / engage / live-in partner / civil union	24 (7.7)		
Ethnicity			
White	247 (79.2)		
Black	30 (9.3)		
Asian	6 (1.9)		
Mixed	29 (9.6)		
Type of sentence			
Community sentence	33 (10.5)		
On license after prison release	280 (89.5)		
Amount of time left on license/community sentence			
12 months or longer	208 (66.5)		
Less than 12 months	105 (33.5)		
CD symptoms (1–15)		7.66	3.46
OAS-M (weighted aggression score)		322.14	177.46

and participant selection are detailed in the trial protocol paper (Fonagy et al., 2020).

Measures

Conduct disorder (CD)

CD was retrospectively assessed using 15 questions from the ASPD section of the Structured Clinical Interview for DSM-IV Disorders (SCID-II; Gibbon, Spitzer, Williams, First, and Benjamin, 1997). Among these questions, specific behaviors such as 'Before the age of 15 often bullied, threatened, or intimidated others', 'Before the age of 15 was physically cruel to people', and 'Before the age of 15 often lied to obtain goods or favors or to avoid obligations (i.e. 'cons' others)' were included to assess CD. Participants meeting the criteria for conduct problems were assigned a score of 'threshold or true,' with an index ranging from 0 to 15. This index indicates the number of CD symptoms experienced before the age of 15. The SCID-II is a semi-structured diagnostic interview with demonstrated validity and reliability (Ullrich et al., 2008), notably within offenders populations (Lobbestael, Leurgans, & Arntz, 2011). In our current sample, the mean (average) and standard deviation for CD scores was 7.66 (S.D. = 3.46).

Adult psychopathology

Adult psychopathology was evaluated at the symptom level using several self-report items that align with the spectral level factors detailed in the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017). For Internalizing difficulties, the 12-item CORE-OM Problems and Symptoms subscale (Barkham et al., 2001) was utilized, which has been validated and found reliable (Evans et al., 2002). Disinhibition, Detachment, Antagonism, and Psychoticism were each assessed using five items from the Personality Inventory for DSM-5, contributing to an additional 20 items (PID-5; Krueger, Derringer, Markon, Watson, and Skodol, 2012). In the present study, the PID-5 and CORE-OM demonstrated excellent internal consistency, with Cronbach's alphas of 0.914 and 0.915, respectively. The measures' average scores were 18.41 (S.D. = 12.40) and 22.89 (S.D. = 12.91) for the CORE-OM and the PID-5, respectively.

Reflective functioning (RF)

RF was measured as part of baseline assessment using the Brief Reflective Functioning Questionnaire (BRFQ; Luyten, under development). This 14-item self-report scale evaluates an individual's capacity to reflect on their own and others' mental states. Higher scores indicate more significant difficulties in RF. The BRFQ is derived from the Reflective Functioning Questionnaire (Fonagy et al., 2016), which shows good test-retest reliability and convergent validity within clinical and non-clinical samples (Euler et al., 2021). Briefer versions of the RFQ have been modeled with a single latent dimension (Müller et al., 2022). In the present study, the BRFQ demonstrated good internal consistency, with a Cronbach's alpha of 0.834. The mean score of this measure was 46.31 (S.D. = 15.22).

Aggressive behavior

Aggression was assessed using the Modified Overt Aggression Scale (OAS-M; Endicott et al., 2002), a clinician-administered semi-structured interview that rates aggression over the past week across four scales: verbal aggression, aggression against objects, aggression against others, and auto-aggression (i.e. aggression towards oneself). The total weighted score is calculated by multiplying the sum of each subscale by its corresponding severity weight and then adding the weighted subscale scores together. The OAS-M was found to have robust validity, reliability, and commendable psychometric properties (Coccaro, 2020). In the present study, the OAS-M mean weighted total score was 322.14 (s.d. = 177.46).

Statistical analyses

Structure of psychopathology and RF

Confirmatory factor analysis was employed to estimate four structural models of psychopathology utilizing items from both the 12 CORE-OM and 20 PID-5 items:

1. *Single-Factor Model*: All 32 items loaded onto a single factor.
2. *Correlated Factors Model*: Consisting of five correlated factors, each item loaded onto one of the following: internalizing, disinhibition, detachment, antagonism, or psychoticism.
3. *Higher-Order Model:* This was statistically equivalent to the correlated factors model but included a higher-order factor to predict the covariances between correlated factors.
4. *Bifactor Model*: All items loaded onto a general p factor as well as specific factors related to internalizing, disinhibition, detachment, antagonism, or psychoticism. An alternative bifactor model with three specific factors (internalizing, externalizing, and psychoticism) was also examined. Covariances were set to zero between the general p factor and each specific factor, and between each specific factor.

Reflective functioning assessment

The structure of the BRFQ has not been established. In line with previous studies, we utilized a single factor model to assess global RF (Schwarzer, Nolte, Fonagy, & Gingelmaier, 2022; Stagaki et al., 2022). Two items found to have less adequate fit to the RF factor were subsequently excluded from the final measurement.

We used the maximum likelihood estimator with robust standard errors in all analyses. Model fit was evaluated using several fit indices, including the Chi-square (χ^2), Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). Acceptable fit was indicated by CFI and TLI values $\geqslant 0.90$, and RMSEA values $\leqslant 0.06$ (Hu & Bentler, 1999). Although the χ^2 statistic is sensitive to sample size, we report it for completeness. To compare models, we employed the Akaike information criteria (AIC) and Sample-size adjusted Bayesian information criteria (BIC). These indices impose penalties for model complexity and are suitable for comparing non-nested models. The evidence for preferring one model over another was determined by the difference in AIC or BIC between models (Murray & Johnson, 2013). A difference greater than 10 (in Both AIC and BIC) was taken as very strong evidence in favor of the competing model (Fabozzi, Focardi, Rachev, & Arshanapalli, 2014).

Reliability Analysis

We applied model-based reliability estimates to evaluate the reliability of the general and specific psychopathology factors, including:

1. Omega Hierarchical ($\omega H / \omega Hs$): Estimates the proportion of raw test score variance explained by the general factor/ a given specific factor with a cut-off of >0.08 considered as high.
2. Explained Common Variance (ECV/ECVs): Assesses the multidimensionality of a measure based on the proportion of modeled variance explained by the general factor/specific factor relative to the total modeled variance. ECV > 0.70 suggests a strong general factor (Dueber, 2016).
3. H Index: A measure of construct reliability based on the proportion of variance in indicators explained by the factor, divided by the variance in indicators unexplained by that factor (Hancock & Mueller, 2001).
4. Factor Determinacy (FD): Examines the precision of factor scores based on the correlation between factor scores and the actual factors. It is recommended that factor score estimates should only be used when FD > 0.90 (Gorsuch, 2013).

Additionally, we determined the total information functions for the general and specific psychopathology factors, employing weighted least squares means and variances adjusted estimates of the chosen bifactor model.

Mediation model

To determine the extent to which general psychopathology (M1) and RF (M2) mediate the association between CD during adolescence (X) and aggression in adulthood (Y), we used structural equation modeling (SEM). Our double mediation model utilized the OAS-M total weighted scores as the primary outcome variable (Y), regressing them on CD total scores (X) through a p factor (M1) and RF (M2). Additionally, in a sensitivity analysis, we inverted this relationship, regressing RF on the p factor, to evaluate the robustness of our directional assumption. Supplementary analyses included the examination of five alternative models, wherein the general p factor in the final SEM was replaced with each of the specific factors.

To determine the significance of both direct and indirect effects, 5000 bias-corrected and accelerated bootstrapping analyses were performed to yield 95% confidence intervals (CIs) (Hayes, 2013; Tofighi & MacKinnon, 2016). An effect was deemed significant if the CI excluded zero. Model parameters were adjusted if modification indices exceeded 10 in the BRFQ or 30 in the final SEM, and if there was substantive basis. Model fit was assessed based on previously defined criteria, and all analyses were run with the Lavaan package for RStudio (version 2023.06.0 + 421).

Results

Objective 1: structure of psychopathology and RF

An inter-item polychoric correlation matrix of the items included in the psychopathology factor is summarized in online Supplementary Table S1. The psychopathology single-factor model showed a poor fit to the data (see Table 2) with moderate to strong positive factor loadings (see online Supplementary Table S2). The correlated factors model displayed an acceptable absolute fit (RMSEA = 0.055). However, relative fit indices such as the TLI was marginally below the acceptable threshold of 0.09 (see Table 2 and online Supplementary Table S3 for the model's factor loadings). The correlated factors model fit better than the single-factor model (ΔAIC = 871.01, ΔBIC = 865.82).
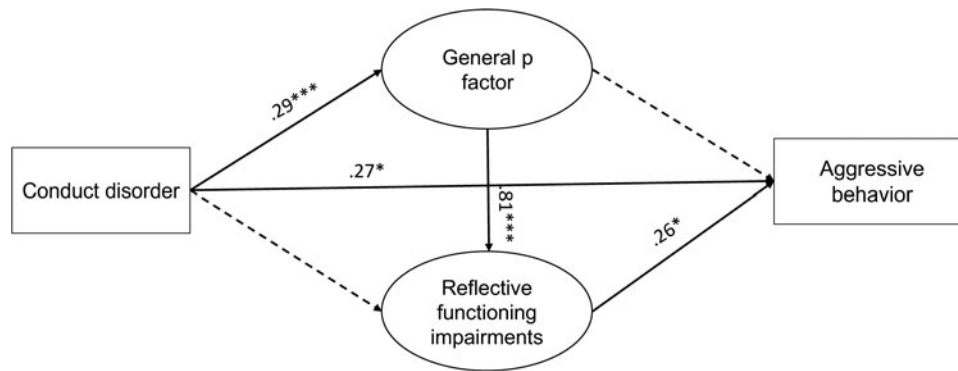
Figure 1. Mediation model leading from conduct disorder in adolescence to aggressive behavior via general p factor and reflective functioning.
Note: Coefficients represent standardized regression weights and standard errors. OAS-M, Overt Aggression Scale – Modified. *$p = 0.05$. **$p = 0.01$. *** $p = 0.001$.

Table 2. Model fit parameters for CFA and SEM models

Model	Fit parameters	AIC	BIC
CFA			
Single factor model	$\chi^2_{(464)} = 1746$, $p < 0.001$, TLI = 0.674, CFI = 0.695, RMSEA = 0.097	27 452.05	27 485.27
Correlated model	$\chi^2_{(454)} = 855$, $p < 0.001$, TLI = 0.896, CFI = 0.905, RMSEA = 0.055	26 581.04	26 619.45
Higher order model	$\chi^2_{(459)} = 893$, $p < 0.001$, TLI = 0.888, CFI = 0.897, RMSEA = 0.057	26 608.83	26 644.64
Bifactorial model (5 factors)	$\chi^2_{(432)} = 779$, $p < 0.001$, TLI = 0.905, CFI = 0.917, RMSEA = 0.052	26 549.05	26 598.88
Bifactorial model (3 factors)	$\chi^2_{(525)} = 1067$, $p < 0.001$, TLI = 0.856, CFI = 0.873, RMSEA = 0.061	30 341.90	30 389.77
SEM			
Final model with general p	$\chi^2_{(942)} = 1471.62$, $p < 0.001$, CFI = 0.901, TLI = 0.891, RMSEA = 0.045	38 886.25	39 385.36
Only internalizing	$\chi^2_{(942)} = 1665.75$, $p < 0.001$, CFI = 0.864, TLI = 0.851, RMSEA = 0.053	39 080.37	39 579.49
Only disinhibition	$\chi^2_{(942)} = 1599.17$, $p < 0.001$, CFI = 0.877, TLI = 0.865, RMSEA = 0.050	39 013.80	39 512.92
Only detachment	$\chi^2_{(942)} = 1661.01$, $p < 0.001$, CFI = 0.865, TLI = 0.852, RMSEA = 0.053	39 075.64	39 574.76
Only antagonism	$\chi^2_{(942)} = 1657.53$, $p < 0.001$, CFI = 0.866, TLI = 0.853, RMSEA = 0.053	39 072.16	39 571.27
Only psychoticism	$\chi^2_{(942)} = 1642.09$, $p < 0.001$, CFI = 0.869, TLI = 0.856, RMSEA = 0.052	39 056.72	39 555.83
Alternative model	$\chi^2_{(942)} = 1471.62$, $p < 0.001$, CFI = 0.901, TLI = 0.891, RMSEA = 0.045	38 886.25	39 385.36

Note: AIC, Akaike information criterion; BIC, Bayesian information criterion; BRFQ, Brief Reflective Functioning Questionnaire; CFI, Comparative Fit Index; RMSEA, Root Mean Square Error of Approximation; TLI: Tucker–Lewis Index.

The higher-order model showed acceptable absolute fit RMSEA = 0.057) and near-acceptable relative fit (CFI and TLI < 0.09) as well (see Table 2 and online Supplementary Table S4). The higher-order model demonstrated a superior fit compared to the single-factor model (ΔAIC = 843.22, ΔBIC = 840.63), but fit worse than the correlated model (ΔAIC = −27.79, ΔBIC = −25.19). The bifactor model with a p factor and five specific factors (internalizing, disinhibition, detachment, antagonism, and psychoticism) provided an excellent fit to the data ($\chi^2_{(432)} = 779$, $p < 0.001$, TLI = 0.905, CFI = 0.917, RMSEA = 0.052), and outperformed the single-factor model (ΔAIC = 903.00, ΔBIC = 886.39), the correlated-factors model (ΔAIC = 31.99, ΔBIC = 20.57), and the higher-order model (ΔAIC = 59.78, ΔBIC = 45.76). Table 3 illustrates the standardized factor loadings for this bifactor model. In contrast, an alternative bifactor model that consolidated the three externalizing subscales into a single factor, fit worse than the bifactor model with five specific factors (ΔAIC = 3792.85, ΔBIC = 3790.89). It had the worst fit parameters in comparison with the single-factor, correlated, and higher-order models

($\chi^2_{(525)} = 1067$, $p < 0.001$, TLI = 0.856, CFI = 0.873, RMSEA = 0.061, see items loadings in online Supplementary Table S5). We therefore used the bifactor model with a p factor and five specific factors in further SEM analyses.

The p factor showed moderate positive factor loadings ($M = 0.502$, S.D. = 0.13) and high levels of information (see information function, Table 2 and online Supplementary Table S6), which is consistent with its high raw score reliability (ωH = 0.80), construct reliability ($H = 93$), and factor determinacy (FD = 0.94). In contrast, specific factors showed weak-to-moderate factor loadings, low information functions and subthreshold reliability indices (see online Supplementary Table S6). Whilst most reliability metrics indicated the dominance of the p factor over specific factors, ECV was split between the p factor (ECV = 0.57) and specific factors (ECVs = 0.43), with a bias towards the general factor. This indicates that the measurement model was multidimensional, making it important to model both general and specific factors in our SEM.

In the RF model, all items were loaded positively onto a single factor ($M = 0.567$, S.D. = 0.15), which showed a good fit to the

Table 3. Standardized factor loadings for the bifactor model with cross-loadings

Variable name	Question	General p	Internalizing	Disinhibition	Detachment	Antagonism	Psychoticism
CORE_q2	I have felt tense, anxious or nervous.	0.582***	0.494***				
CORE_q5	I have felt totally lacking in energy and enthusiasm.	0.423***	0.405***				
CORE_q8	I have been troubled by aches, pains or other physical symptoms.	0.31***	0.314***				
CORE_q11	Tension and anxiety have prevented me from doing important things.	0.542***	0.494***				
CORE_q13	I have been disturbed by unwanted thoughts and feelings.	0.643***	0.434***				
CORE_q15	I have felt panic or terror.	0.459***	0.451***				
CORE_q18	I have had difficulty getting to sleep or staying asleep.	0.587***	0.44***				
CORE_q20	My problems have been impossible to put to one side.	0.545***	0.514***				
CORE_q23	I have felt despairing or hopeless.	0.581***	0.57***				
CORE_q27	I have felt unhappy.	0.53***	0.647***				
CORE_q28	Unwanted images or memories have been distressing me.	0.558***	0.481***				
CORE_q30	I have thought I am to blame for my problems and difficulties.	0.438***	0.423***				
PID5BF_q1	People would describe me as reckless.	0.494***		0.492***			
PID5BF_q2	I feel like I act totally on impulse.	0.539***		0.57***			
PID5BF_q3	Even though I know better, I can't stop making rash decisions.	0.574***		0.501***			
PID5BF_q5	Others see me as irresponsible.	0.535***		0.447***			
PID5BF_q6	I'm not good at planning ahead.	0.439***		0.126			
PID5BF_q4	I often feel like nothing I do really matters.	0.593***			0.073		
PID5BF_q13	I'm not interested in making friends.	0.312***			0.329***		
PID5BF_q14	I get irritated easily by all sorts of things.	0.313***			0.599***		
PID5BF_q16	I don't like to get too close to people.	0.528***			0.475***		
PID5BF_q18	I rarely get enthusiastic about anything.	0.532***			0.13		
PID5BF_q17	It's no big deal if I hurt other peoples' feelings.	0.438***				0.372***	
PID5BF_q19	I crave attention.	0.221***				0.402***	
PID5BF_q20	I often have to deal with people who are less important than me.	0.281***				0.43***	
PID5BF_q22	I use people to get what I want.	0.348***				0.751***	
PID5BF_q25	It is easy for me to take advantage of others.	0.358***				0.663***	
PID5BF_q7	My thoughts often don't make sense to others.	0.727***					−0.06
PID5BF_q12	I steer clear of romantic relationships.	0.564***					0.319***
PID5BF_q21	I often have thoughts that make sense to me but that other people say are strange.	0.715***					0.024

PID5BF_q23	I often 'zone out' and then suddenly come to and realize that a lot of time has passed.	0.66***				0.485***
PID5BF_q24	Things around me often feel unreal, or more real than usual.	0.688***				0.479***
M		0.502	0.427	0.321	0.524	0.249
S.D.		0.130	0.174	0.223	0.171	0.254
ω/ω_s		0.948	0.811	0.710	0.765	0.856
ω_H/ω_{Hs}		0.800	0.329	0.236	0.548	0.104
ECV/ECV$_s$		0.567 / 0.182	0.068	0.047	0.098	0.038

Note: CORE, Clinical Outcomes in Routine Evaluation – Outcome Measure; ECV, Explained Common Variance; ECV$_s$, Explained Common Variance subscale; M, mean; PID5BF, The Personality Inventory for DSM-5 – Brief Form; S.D., standard deviation; ω, Omega; ω_s, Omega subscale; ω_H, Omega hierarchical; ω_{Hs}, Omega hierarchical subscale.
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

data. Certain error covariances were freed to account for item overlap, as indicated by modification indices (e.g. 'When I get angry, I say things without really knowing why I am saying them' and 'When I get angry, I say things that I later regret'). Online Supplementary Tables S7 and S8 present the RF factor analysis modification indices and item loadings, respectively. The single factor BRFQ model showed good fit of the data ($\chi^2_{(45)} = 88.91$, $p < 0.001$, CFI = 0.959, TLI = 0.940, RMSEA = 0.057; AIC = 14 547.53, BIC = 14 669.53)

Objective 2: mediation models

Our hypothesized SEM, in which CD scores during adolescence predicted OAS-M aggression scores via the p factor and RF factor, provided good fit to the data: $\chi^2_{(942)} = 1471.62$, $p < 0.001$, CFI = 0.901, TLI = 0.891, RMSEA = 0.045 (see Table 2 and Fig. 1; model modification indices are available in online Supplementary Table S9). CD during adolescence positively predicted aggression in adulthood, both directly and indirectly, through pathways involving the p factor and RF. In this mediation model, CD had a direct effect on OAS-M aggression scores, indicating that 1 point increase in CD symptoms was associated with a 0.27 increase in aggressive behavior score in adulthood. Additionally, each increase in CD scores was associated with a 0.29 increase in p factor scores, but not RF. A strong positive relationship was observed between the p factor and RF ($\beta = 0.81$), which subsequently associated with increased aggression scores ($\beta = 0.26$). Notably, the link between the p factor and the OAS-M scores was not independently significant within this model. The indirect pathway was statistically significant, where CD predicted p, which in turn positively associated with RF, which finally linked with elevated levels of aggression ($B = 0.032$, 95% CI [0.0002–0.071]). However, the indirect paths involving only the p factor (CD → p factor → OAS-M) and RF (CD → RF → OAS-M) as mediating the link between CD and aggression scores was not statistically significant ($B = 0.012$ and $B = 0.011$ respectively, both pathways had non-significant 95% CI).

Alternative models

In a sensitivity analysis where the direction between the p factor and RF factor was reversed (e.g. CD → RF → p → OAS-M), the mediation effect was not significant ($B = 0.011$, 95% CI [−0.023 to 0.047]). In supplementary analyses, using specific factors as mediators instead of the p factor, only the specific psychoticism factor was a significant mediator between CD, RF, and OAS-M, but no model showed acceptable fit (see Table 2 and online Supplementary Figs S2–S6). Higher CD scores in adolescence predicted elevated psychoticism scores in adulthood, which in turn was linked with greater RF scores, and RF scores were associated with increased OAS-M scores ($B = 0.015$, 95% CI [0.003–0.033]; see online Supplementary Fig. S6).

Discussion

In this study, we examined the structure of psychopathology and RF in adult male who offended, and explored how these factors mediated the relationship between CD in adolescence and aggression in adulthood. Our results highlight that, of the models assessed, a bifactor model demonstrated the most optimal fit. This model integrated a general p factor and specific factors including internalizing, disinhibition, detachment, antagonism,

and psychoticism. Its superiority in fit was evident when compared against correlated factors model, higher-order model, and single-factor model, based on criteria such as the AIC and BIC. Consistent with our expectations, CD during adolescence was associated with aggressive behavior in adulthood. In addition to this direct relationship, the pathway from CD to aggression was mediated by both the p factor and RF. Individuals with more CD symptoms, which were linked with higher general psychopathology and consequently with difficulties in RF, were more likely to display elevated aggression scores. When examined in isolation, neither the p factor nor the RF factor mediated the link between CD and aggression.

In relation to the first study objective, which focused on the structure of psychopathology in our samples, our findings support a transdiagnostic approach in assessing psychopathology among offenders. Data supported a bifactor model, with a p factor explaining the shared variance across domains and specific factors like internalizing, disinhibition, detachment, antagonism, and psychoticism explaining the variance amongst subsets of items. Although few studies have used the bifactor model in forensic groups, existing research recognizes a p factor alongside distinct psychopathology elements (Constantinou et al., 2019; Ignatyev, Baggio, & Mundt, 2019). Concerns exist over the bifactor model's tendency to overfit noise in the data and unclear definitions of the p factor and specific factors (Bonifay & Cai, 2017; Greene et al., 2019; Watts, Lane, Bonifay, Steinley, & Meyer, 2020). There are also concerns that the 'p factor' is merely a methodological artifact. Specifically, critics argue that individuals without co-occurring problems or disorders tend to inflate correlations among various psychopathologies and that the p factor is weaker in models that incorporate data from multiple informants (Watts et al., 2021, 2022). We would argue that none of these critiques rule out a p factor, since it has been validated against external criteria and shows stability like that of psychological traits. Rather, researchers should be more cautious in their interpretations of the p factor. We remain agnostic about the causal basis of the p factor in our sample, since we do not validate it against external measures or analyze longitudinal data which would speak to causal mechanisms. Instead, we use it as a statistical tool to summarize individual differences in the overlap amongst multiple conditions, which provides some index of severity, and helps us delve deeper into the constructs that tie together CD during adolescence and adult aggression.

Regarding the study's second objective, our mediation model revealed that adolescent CD positively influences adult aggression, both directly and mediated by the p and RF factors. The established link between CD and an increased risk for aggressive behaviors is well-documented (Fairchild et al., 2019; Hodgins, Cree, Alderton, & Mak, 2008; North, Kotamarti, & Pollio, 2022). In our study, this direct link between CD and aggressive behavior remained consistently strong across all main and alternative models, thereby bolstering the robustness of this established link. The significant indirect path from CD to aggression via the p factor underscores the important role of general psychopathology. The link between CD and the p factor was evident both in the main and the specific models, albeit weaker in the latter. This aligns with studies suggesting CD may underpin multiple psychopathologies (Kim-Cohen et al., 2003; Krueger et al., 2005) and stresses the need to explore broader risk factors like parental behavior and childhood adversities when studying issues like CD or aggression (Boden, Fergusson, & Horwood, 2010). Importantly, the influence of p factor scores on the link between conduct issues

and aggression was mediated by RF levels. While previous research linked various mental health difficulties, particularly severe personality disorders and comorbidity, to increased aggression (Girasek, Nagy, Fekete, Ungvari, & Gazdag, 2022; Weltens et al., 2021; Whiting & Fazel, 2020) our model did not identify a significant link between p factor and aggressive behavior. Instead, the connection was mediated by RF impairments, indicating a noteworthy relationship between RF and aggressive behavior.

Our finding that RF mediates the relationship between general psychopathology and aggression underscores the importance of an individual's ability to reflect on their own mental states and those of others in the expression of aggressive behavior. Fonagy and Luyten (2018) argued that conduct problems can lead to temporary or chronic difficulties in mentalizing, impairing the inhibition of interpersonal violence through empathy and perspective-taking. Our study supports this hypothesis and broadens it to include general psychopathology, especially among male offenders who exhibited prevalent CD symptoms during adolescence. The presence of CD during childhood and adolescence has been identified as a risk factor for a broad array of mental health issues and psychosocial difficulties (i.e. p), both of which can disrupt the development of RF and heighten the risk of later aggression (Fergusson, Horwood, & Lynskey, 1994). For instance, executive functioning impairments, common to various childhood and adolescent mental health conditions such as oppositional defiant disorder (ODD), attention deficit hyperactivity disorder (ADHD), and CD (Figueiredo, Ramião, Barroso, & Barbosa, 2022), have been linked to RF deficits in both healthy and clinical populations (Ogilvie, Stewart, Chan, & Shum, 2011; Rutherford et al., 2018). Constructs that are related to RF, such as empathy and theory of mind (ToM) during childhood and adolescence, have been also found to predict later symptoms of ASPD, psychopathy, violence, and offending behavior (Frick & White, 2008; Galán, Choe, Forbes, & Shaw, 2017; Karoğlu, Ferguson, & Ó Ciardha, 2022; Rhee et al., 2021). However, while ToM represents the cognitive capacity to understand others' mental states, and empathy incorporates an emotional dimension to this understanding, RF encompasses both cognitive and affective processes, enabling individuals to reflect upon and interpret behaviors considering underlying mental states. Consequently, RF is viewed as a more operationalized, comprehensive, integrative construct essential for effective interpersonal relationships and self-regulation, making it an especially pertinent target for therapeutic interventions aimed at offenders with ASPD traits.

At the neurocognitive level, RF is understood as reprocessing mental state information, enhancing empathy, and interpreting social cues, which are also essential to ToM. These functions may be diminished due to neurodevelopmental challenges, leading to difficulty in regulating emotions and impulses, especially in contexts related to aggression. This regulation requires 'hot executive functions' for adaptive responses and better understanding of oneself and others (Zelazo, 2020). This perspective is consistent with research suggesting that the general psychopathology factor (p) mirrors impulsivity issues, explaining the correlation between psychopathology and RF (Carver, Johnson, & Timpano, 2017). Furthermore, interrelations between CD, general psychopathology, RF, and aggression may arise from shared brain circuitry, especially within the orbitofrontal and ventromedial prefrontal cortex (PFC), areas associated with CD and key to mentalization processes (Bateman, Ryle, Fonagy, & Kerr, 2007;

Rubia, 2011). Abnormalities in these regions, stemming from early-life risk factors, can contribute to heightened adult psychopathology, RF challenges, and aggression (De Brito, Viding, Kumari, Blackwood, & Hodgins, 2013; Hiser & Koenigs, 2018).

Among five alternative models with each specific factor added as a mediator, only the specific psychoticism factor significantly mediated the relationship between CD and RF, subsequently linking to aggression. Psychoticism, reflecting the severity of mental health issues, is sometimes disregarded in assessing mental health difficulties, despite its prevalence in both clinical and community samples (Singh & Gupta, 2021; Van Os, Linscott, Myin-Germeys, Delespaul, & Krabbendam, 2009). This distinctive nature of psychoticism within the broader construct of psychopathology in our results highlights the need to consider its singular contribution to our comprehension of mental health and aggressive behavior. Specific factors might represent psychopathological dimensions devoid of overlap with other dimensions, effectively studying 'pure' problem domains (Caspi et al., 2014). In this case, 'pure' psychoticism might reflect the degree of disorganized thought, a feature that parallels RF since both involve metacognitive abilities (Weijers et al., 2020). Nonetheless, the mediation model did not show an acceptable fit to the data, nor did the specific psychoticism factor show acceptable levels of reliable, so we caution these interpretations.

In the context of treating conduct problems or antisocial/aggressive behaviors, practitioners often focus on specific diagnoses like CD, ODD, or ASPD, viewing these conditions as root causes. However, consistent with transdiagnostic approaches, psychological treatments might target both specific and broad underlying mechanisms (Barlow, Ellard, & Fairholme, 2010; Bateman, Campbell, Luyten, & Fonagy, 2018). For example, in the START trial, both specific conduct difficulties and general psychopathology decreased throughout a parent-based intervention targeting teens at risk of antisocial behavior (Constantinou et al., 2019). Similarly, interventions that emphasize externalizing mechanisms, such as impulsivity, self-control, goal setting and planning skills, and anger management, have been found not only to reduce criminal and aggressive behavior but also to enhance quality of life and social relationships (Elyasi, Bijandi, Bijandi, & Abdolmaleki, 2017; Ferguson, Conway, Endersby, & MacLeod, 2009; Shahsavarani et al., 2016). Our model thus underscores the importance of addressing a spectrum of impairments and difficulties by concentrating on RF, a mental ability posited to underpin all human interaction. As such, it forms the foundation for the well-being of individuals across various mental health populations. MBT, in particular, seeks to cultivate healthier thought patterns, emotions, and behaviors, ultimately leading to a reduction in delinquent behavior.

Our findings should be interpreted with the following limitations in mind. First, although childhood conduct problems were identified as a significant risk factor for aggressive difficulties, not all children with conduct problems progress to adult psychopathology. Our mediation model merely outlines one potential developmental pathway. Second, we relied on self-report measures for psychopathology assessment and cross-sectional data, thus causality cannot be established. Third, the retrospective assessment of CD might be subject to memory and reporting biases. While some studies have reported a high concordance between prospective and retrospective assessments of childhood maltreatment, psychopathology (Scott, McLaughlin, Smith, & Ellis, 2012), and the accuracy of self- reported offending history (Jolliffe et al., 2003), the validity of retrospectively assessing CD in adult

offenders remains understudied. Future research should employ longitudinal designs and objective measures to assess the retrospective validity of psychopathology in offenders and to evaluate the pathway identified in the current study. Longitudinal studies could also address a significant methodological limitation in our study, arising from the reliance on cross-sectional data in the mediation analysis, which precludes establishing causal relationships among the observed variables. Although our alternative models partially mitigate concerns of reverse causation, longitudinal data would provide a more effective mean of establishing temporal precedence, an approach we intend to employ in future studies with this cohort. Considering the well-documented adverse life-course outcomes for individuals who follow antisocial developmental pathways (Healey, Knapp, & Farrington, 2004), it is crucial for future research to investigate the interplay between factors such as impaired fear conditioning, callous-unemotional traits, brain variations linked to empathy and ToM, and environmental influences including unstable family backgrounds, as potential mediators and moderators of our identified pathways.

In conclusion, this investigation contributes valuable insights into psychopathology's structure in a forensic sample and underscores the mediating roles of the p factor and RF between CD and aggression. It emphasizes the importance for clinicians to assess general psychopathology moving beyond categorical diagnoses. Incorporating the theoretical perspective of general psychopathology with the operationalization concept of RF into future research could enhance our understanding of the underlying mechanisms connecting various psychiatric disorders to aggression. This approach could lead to more effective, holistic (or transdiagnostic) clinical strategies, especially in managing high-risk individuals with complex comorbidities, thereby improving intervention outcomes and reducing aggression in psychiatric settings. Our findings have substantial implications for the design and implementation of efficacious interventions for aggression and psychopathology among men who offended, such as MBT.

References

Barkham, M., Leach, C., Lucock, M., Evans, C., Margison, F., Mellor-Clark, J., … McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, 69(2), 184–196. https://doi.org/10.1037/0022-006X.69.2.184

Barlow, D. H., Ellard, K. K., & Fairholme, C. P. (2010). *Unified protocol for transdiagnostic treatment of emotional disorders: Workbook* (pp. 145–158). New York: Oxford University Press.

Bateman, A., Campbell, C., Luyten, P., & Fonagy, P. (2018). A mentalization-based approach to common factors in the treatment of borderline personality disorder. *Current Opinion in Psychology*, 21, 44–49. https://doi.org/10.1016/j.copsyc.2017.09.005

Bateman, A., O'Connell, J., Lorenzini, N., Gardner, T., & Fonagy, P. (2016). A randomised controlled trial of mentalization-based treatment versus structured clinical management for patients with comorbid borderline personality disorder and antisocial personality disorder. *BMC Psychiatry*, 16(1), 304. https://doi.org/10.1186/s12888-016-1000-9

Bateman, A. W., Ryle, A., Fonagy, P., & Kerr, I. B. (2007). Psychotherapy for borderline personality disorder: Mentalization based therapy and cognitive analytic therapy compared. *International Review of Psychiatry*, 19(1), 51–62.

Bo, S., Sharp, C., Lind, M., Simonsen, S., & Bateman, A. (2023). Mentalizing mediates the relationship between psychopathy and premeditated criminal offending in schizophrenia: A 6-year follow-up study. *Nordic Journal of Psychiatry*, 77(6), 547–559. https://doi.org/10.1080/08039488.2023.2186483

Boden, J. M., Fergusson, D. M., & Horwood, L. J. (2010). Risk factors for conduct disorder and oppositional/defiant disorder: Evidence from a New Zealand birth cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49(11), 1125–1133. https://doi.org/10.1016/j.jaac.2010.08.005

Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research*, 52(4), 465–484. https://doi.org/10.1080/00273171.2017.1309262

Carver, C. S., Johnson, S. L., & Timpano, K. R. (2017). Toward a functional view of the p factor in psychopathology. *Clinical Psychological Science*, 5, 880–889. https://doi.org/10.1177/2167702617710037

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137. https://doi.org/10.1177/2167702613497473

Caspi, A., Houts, R. M., Fisher, H. L., Danese, A., & Moffitt, T. E. (2023). The general factor of psychopathology (p): Choosing among competing models and interpreting p. *Clinical Psychological Science*, 12(1), 53–82. https://doi.org/10.1177/21677026221147872

Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, 175(9), 831–844. https://doi.org/10.1176/appi.ajp.2018.17121383

Coccaro, E. F. (2020). The Overt Aggression Scale Modified (OAS-M) for clinical trials targeting impulsive aggression and intermittent explosive disorder: Validity, reliability, and correlates. *Journal of Psychiatric Research*, 124, 50–57. https://doi.org/10.1016/j.jpsychires.2020.01.007

Constantinou, M. P., Goodyer, I. M., Eisler, I., Butler, S., Kraam, A., Scott, S., … Fonagy, P. (2019). Changes in general and specific psychopathology factors over a psychosocial intervention. In *Journal of the American Academy of Child & Adolescent Psychiatry*, 58, 776–786. www.jaacap.org.

Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of General Psychiatry*, 66(7), 764–772. https://doi.org/10.1001/archgenpsychiatry.2009.85

De Brito, S. A., Viding, E., Kumari, V., Blackwood, N., & Hodgins, S. (2013). Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. *PLoS ONE*, 8(6), e65566. https://doi.org/10.1371/journal.pone.0065566

Dueber, D. M. (2016). Bifactor indices calculator: A microsoft excel-based tool to calculate various indices relevant to bifactor CFA models. *Educational, School, and Counseling Psychology Research Tools*. https://doi.org/10.13023/edp.tool.01

Elyasi, S., Bijandi, M., Bijandi, G., & Abdolmaleki, S. (2017). The effectiveness of anger control training on resilience; quality of life; and reducing student's misbehavior. *Quarterly Journal of Education Studies*, 3(10), 101–122.

Endicott, J., Tracy, K., Burt, D., Olson, E., & Coccaro, E. F. (2002). A novel approach to assess inter-rater reliability in the use of the overt aggression scale-modified. *Psychiatry Research*, 112(2), 153–159. https://doi.org/10.1016/S0165-1781(02)00185-3

Erskine, H. E., Baxter, A. J., Patton, G., Moffitt, T. E., Patel, V., Whiteford, H. A., … Scott, J. G. (2017). The global coverage of prevalence data for mental disorders in children and adolescents. *Epidemiology and Psychiatric Sciences*, 26(4), 395–402. https://doi.org/10.1017/S2045796015001158

Euler, S., Nolte, T., Constantinou, M., Griem, J., Montague, P. R., & Fonagy, P. (2021). Interpersonal problems in borderline personality disorder: Associations with mentalizing, emotion regulation, and impulsiveness. *Journal of Personality Disorders*, 35(2), 177–193. https://doi.org/10.1521/pedi_2019_33_427

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., … Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, 180(JAN.), 51–60. https://doi.org/10.1192/bjp.180.1.51

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The Basics of Financial Econometrics*. Hoboken, NJ: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118856406

Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., … De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers*, 5(1), 43. https://doi.org/10.1038/s41572-019-0095-y

Ferguson, G., Conway, C., Endersby, L., & MacLeod, A. (2009). Increasing subjective well-being in long-term forensic rehabilitation: Evaluation of well-being therapy. *The Journal of Forensic Psychiatry & Psychology*, 20(6), 906–918. https://doi.org/10.1080/14789940903174121

Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1994). Structure of DSM-III-R criteria for disruptive childhood behaviors: Confirmatory factor models. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33(8), 906–918. https://doi.org/10.1097/00004583-199410000-00010

Figueiredo, P., Ramião, E., Barroso, R., & Barbosa, F. (2022). Executive (dys) functions and ODD and CD: A systematic review and meta-analysis. *Neuropsychology*, 37(2), 133–156. https://doi.org/10.1037/neu0000881

Fonagy, P., Campbell, C., Constantinou, M., Higgitt, A., Allison, E., & Luyten, P. (2022). Culture and psychopathology: An attempt at reconsidering the role of social learning. *Development and Psychopathology*, 34(4), 1205–1220. https://doi.org/10.1017/S0954579421000092

Fonagy, P., & Luyten, P. (2018). Conduct problems in youth and the RDoC approach: A developmental, evolutionary-based view. *Clinical Psychology Review*, 64, 57–76.

Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y. W., Warren, F., Howard, S., … Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS ONE*, 11(7), e0158678. https://doi.org/10.1371/journal.pone.0158678

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology*, 9(4), 679–700. https://doi.org/10.1017/s0954579497001399

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual version 5 for application to adult attachment interviews*.

Fonagy, P., Yakeley, J., Gardner, T., Simes, E., McMurran, M., Moran, P., … Bateman, A. (2020). Mentalization for Offending Adult Males (MOAM): Study protocol for a randomized controlled trial to evaluate mentalization-based treatment for antisocial personality disorder in male offenders on community probation. *Trials*, 21(1), 1001. https://doi.org/10.1186/s13063-020-04896-w

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 and Allied Disciplines*, 49, 359–375. https://doi.org/10.1111/j.1469-7610.2007.01862.x

Galán, C. A., Choe, D. E., Forbes, E. E., & Shaw, D. S. (2017). Interactions between empathy and resting heart rate in early adolescence predict violent behavior in late adolescence and early adulthood. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 58(12), 1370–1380. https://doi.org/10.1111/jcpp.12771

Gibbon, M., Spitzer, R. L., Williams, J. B. W., First, M. B., & Benjamin, L. S. (1997). *User's guide for the structured clinical interview for DSM-IV axis II personality disorders: SCID-II*. Washington, DC: American Psychiatric Press.

Girasek, H., Nagy, V. A., Fekete, S., Ungvari, G. S., & Gazdag, G. (2022). Prevalence and correlates of aggressive behavior in psychiatric inpatient populations. *World Journal of Psychiatry*, 12(1), 1–23. https://doi.org/10.5498/wjp.v12.i1.1

Gorsuch, R. L. (2013). *Factor analysis*. New York: Psychology Press. https://doi.org/10.4324/9780203781098

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., … Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, 128(7), 740–764. https://doi.org/10.1037/abn0000434

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sorbom (Eds.), *Structural Equation Modeling: Present und Future—A Festschrift in Honor of Karl Joreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford, pp. 3–4. https://doi.org/978-1-60918-230-4

Healey, A., Knapp, M., & Farrington, D. P. (2004). Adult labour market implications of antisocial behaviour in childhood and adolescence: Findings from a UK longitudinal study. *Applied Economics*, 36(2), 93–105. https://doi.org/10.1080/0003684042000174001

Hiser, J., & Koenigs, M. (2018). The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biological Psychiatry*, 83, 638–647. https://doi.org/10.1016/j.biopsych.2017.10.030

Hodgins, S., Cree, A., Alderton, J., & Mak, T. (2008). From conduct disorder to severe mental illness: Associations with aggressive behaviour, crime and victimization. *Psychological Medicine*, 38(7), 975–987. https://doi.org/10.1017/S0033291707002164

Hoffmann, M. S., Evans-Lacko, S., Collishaw, S., Knapp, M., Pickles, A., Shearer, C., … Maughan, B. (2023). Parent- and teacher-reported associations from adolescent bifactor models of psychopathology: An outcome-wide association study of 26 outcomes in mid-life. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 64(3), 397–407. https://doi.org/10.1111/jcpp.13707

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Hyland, P., Murphy, J., Shevlin, M., Bentall, R. P., Karatzias, T., Ho, G. W. K., … McElroy, E. (2021). On top or underneath: Where does the general factor of psychopathology fit within a dimensional model of psychopathology? *Psychological Medicine*, 51(14), 2422–2432. https://doi.org/10.1017/S003329172000104X

Ignatyev, Y., Baggio, S., & Mundt, A. P. (2019). The underlying structure of comorbid mental health and substance use disorders in prison populations. *Psychopathology*, 52(1), 679–700. https://doi.org/10.1159/000495844

Jolliffe, D., Farrington, D. P., Hawkins, J. D., Catalano, R. F., Hill, K. G., & Kosterman, R. (2003). Predictive, concurrent, prospective and retrospective validity of self-reported delinquency. *Criminal Behaviour and Mental Health*, 13(3), 179–197. https://doi.org/10.1002/cbm.541

Karoğlu, N., Ferguson, H. J., & Ó Ciardha, C. (2022). Theory of mind in offending: A systematic review. *Trauma, Violence, and Abuse*, 23, 1610–1628. https://doi.org/10.1177/15248380211013143

Kealy, D., Rice, S. M., Seidler, Z. E., Oliffe, J. L., & Ogrodniczuk, J. S. (2021). Reflective functioning and men's mental health: Associations with resilience and personal growth initiative. *Stress and Health*, 37(4), 706–714. https://doi.org/10.1002/smi.3030

Kim-Cohen, J., Caspi, A., Moffitt, T. E., Harrington, H. L., Milne, B. J., & Poulton, R. (2003). Prior juvenile diagnoses in adults with mental disorder: Developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry*, 60(7), 709–717. https://doi.org/10.1001/archpsyc.60.7.709

Kotov, R., Waszczuk, M. A., Krueger, R. F., Forbes, M. K., Watson, D., Clark, L. A., … Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. https://doi.org/10.1037/abn0000258

Krueger, R. F, Derringer, J., Markon, K. E., Watson, D., & Skodol, A. E.. (2012). Personality inventory for DSM-5. Psychiatry Research. https://doi.org/10.1037/t30042-000

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, 114, 537–550. https://doi.org/10.1037/0021-843X.114.4.537

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). Inter-rater reliability of the structured clinical interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). *Clinical Psychology and Psychotherapy*, 18(1), 75–79. https://doi.org/10.1002/cpp.693

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). CP16CH13_Luyten ARjats.cls The mentalizing approach to psychopathology: State of the art and future directions CP16CH13_Luyten ARjats.cls. *Annual Review of Clinical Psychology*, 2020(16), 297–325. https://doi.org/10.1146/annurev-clinpsy-071919

Müller, S., Wendt, L. P., Spitzer, C., Masuhr, O., Back, S. N., & Zimmermann, J. (2022). A critical evaluation of the Reflective Functioning Questionnaire (RFQ). *Journal of Personality Assessment*, 104(5), 613–627. https://doi.org/10.1080/00223891.2021.1981346

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, 41(5), 407–422.

Newbury-Helps, J., Feigenbaum, J., & Fonagy, P. (2017). Offenders with antisocial personality disorder display more impairments in mentalizing. *Journal of Personality Disorders*, 31(2), 232–255. https://doi.org/10.1521/pedi_2016_30_246

North, C. S., Kotamarti, V., & Pollio, D. E. (2022). Deconstructing childhood conduct and adult antisocial criteria for the diagnosis of antisocial personality disorder. *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists*, 34(2), 97–105. https://doi.org/10.12788/acp.0063

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., & Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavioural basis of health program behavioural basis of health program. *Criminology; An interdisciplinary Journal*, 49(4), 1063–1108.

Parada-Fernández, P., Herrero-Fernández, D., & Rodríguez-Arcos, I. (2023). The moderation effect of mentalization in the relationship between impulsiveness and aggressive behavior. *Scandinavian Journal of Psychology*, 64(6), 794–801.

Perkins, E. R., Joyner, K. J., Foell, J., Drislane, L. E., Brislin, S. J., Frick, P. J., … Patrick, C. J. (2022). Assessing general versus specific liability for externalizing problems in adolescence: Concurrent and prospective prediction of symptoms of conduct disorder, ADHD, and substance use. *Journal of Psychopathology and Clinical Science*, 131(7), 793–807. https://doi.org/10.1037/abn0000743

Rhee, S. H., Woodward, K., Corley, R. P., Du Pont, A., Friedman, N. P., Hewitt, J. K., … Zahn-Waxler, C. (2021). The association between toddlerhood empathy deficits and antisocial personality disorder symptoms and psychopathy in adulthood. *Development and Psychopathology*, 33(1), 173–183. https://doi.org/10.1017/S0954579419001676

Rubia, K. (2011). "Cool" inferior frontostriatal dysfunction in attention-deficit/hyperactivity disorder versus "hot" ventromedial orbitofrontal-limbic dysfunction in conduct disorder: A review. *Biological Psychiatry*, 69, e69–87. https://doi.org/10.1016/j.biopsych.2010.09.023

Rutherford, H. J. V., Byrne, S. P., Crowley, M. J., Bornstein, J., Bridgett, D. J., & Mayes, L. C. (2018). Executive functioning predicts reflective functioning in mothers. *Journal of Child and Family Studies*, 27(3), 944–952. https://doi.org/10.1007/s10826-017-0928-9

Sadler, K., Vizard, T., Ford, T., Marchesell, F., Pearce, N., Mandalia, D.. (2018). *Mental health of children and young people in England, 2017*. UK: NHS Digital, part of the Government Statistical Service.

Schwarzer, N. H., Nolte, T., Fonagy, P., & Gingelmaier, S. (2022). Self-rated mentalizing mediates the relationship between stress and coping in a non-clinical sample. *Psychological Reports*, 125(2), 742–762. https://doi.org/10.1177/0033294121994846

Scott, K. M., McLaughlin, K. A., Smith, D. A. R., & Ellis, P. M. (2012). Childhood maltreatment and DSM-IV adult mental disorders: Comparison of prospective and retrospective findings. *British Journal of Psychiatry*, 200(6), 469–475. https://doi.org/10.1192/bjp.bp.111.103267

Shahsavarani, A. M., Noohi, S., Heyrati, H., Mohammadi, M., Mohammadi, A., & Sattari, K. (2016). Anger management and control in social and behavioral sciences: A systematic review of literature on biopyschosocial model. *International Journal of Medical Reviews Systematic Review International Journal of Medical Reviews*, 3(1), 355–364.

Singh, M. K., & Gupta, U. (2021). Psychoticism and mental health. *Journal of Psychosocial Research*, 16(1), 143–151.

Stagaki, M., Nolte, T., Feigenbaum, J., King-Casas, B., Lohrenz, T., Fonagy, P., … Montague, P. R. (2022). The mediating role of attachment and mentalising in the relationship between childhood maltreatment, self-harm and suicidality. *Child Abuse and Neglect*, 128, 105576. https://doi.org/10.1016/j.chiabu.2022.105576

Taubner, S., White, L. O., Zimmermann, J., Fonagy, P., & Nolte, T. (2013). Attachment-related mentalization moderates the relationship between psychopathic traits and proactive aggression in adolescence. *Journal of Abnormal Child Psychology*, 41(6), 929–938. https://doi.org/10.1007/s10802-013-9736-x

Thomas, N., & Jenkins, H. (2019). The journey from epistemic vigilance to epistemic trust: Service-users experiences of a community mentalization-based treatment programme for anti-social personality disorder (ASPD). *Journal of Forensic Psychiatry and Psychology*, 30(6), 909–938. https://doi.org/10.1080/14789949.2019.1670856

Tofighi, D., & MacKinnon, D. P. (2016). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2), 194–205.

Ullrich, S., Deasy, D., Smith, J., Johnson, B., Clarke, M., Broughton, N., … Coid, J. (2008). Detecting personality disorders in the prison population of England and Wales: Comparing case identification using the SCID-II screen and the SCID-II clinical interview. *Journal of Forensic Psychiatry and Psychology*, 19(3), 301–322. https://doi.org/10.1080/14789940802045182

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. J. (2017). What is the *p*-factor of psychopathology? Some risks of general factor modeling. *Theory and Psychology*, 27(6), 759–773. https://doi.org/10.1177/0959354317737185

Van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. *Psychological Medicine*, 39(2), 179–195. https://doi.org/10.1017/S0033291708003814

Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. C. (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological Inquiry*, 31, 310–320. https://doi.org/10.1080/1047840X.2020.1853476

Watts, A. L., Makol, B. A., Palumbo, I. M., De Los Reyes, A., Olino, T. M., Latzman, R. D.…Sher, K. J. (2022). How robust is the p factor? Using multi-trait-multimethod modeling to inform the meaning of general factors of youth psychopathology. *Clinical Psychological Science*, 10(4), 640–661.

Watts, A. L., Meyer, F. A., Greene, A. L., Wood, P. K., Trull, T. J., Steinley, D., & Sher, K. J. (2021). Spurious empirical support for the p-factor arises with the inclusion of undiagnosed cases.

Weijers, J. G., Ten Kate, C., Debbané, M., Bateman, A. W., De Jong, S., Selten, J.-P. C. J., … Eurelings-Bontekoe, E. H. M. (2020). Mentalization and psychosis: A rationale for the use of mentalization theory to understand and treat non-affective psychotic disorder. *Journal of Contemporary Psychotherapy*, 50, 223–232.

Weltens, I., Bak, M., Verhagen, S., Vandenberk, E., Domen, P., van Amelsvoort, T., … Drukker, M. (2021). Aggression on the psychiatric ward: Prevalence and risk factors. A systematic review of the literature. *PLoS ONE*, 16(10), e0258346. https://doi.org/10.1371/journal.pone.0258346

Whiting, D., & Fazel, S.. (2020). Epidemiology and Risk Factors for Violence in People with Mental Disorders. In B. Carpiniello, A Vita, & AC Mencacci (Eds.), *Violence and Mental Disorders. Comprehensive Approach to Psychiatry* (Vol. 1). Cham: Springer.

Zelazo, P. D. (2020). Executive function and psychopathology: A neurodevelopmental perspective. *Annual Review of Clinical Psychology*, 16, 431–454. https://doi.org/10.1146/annurev-clinpsy-072319-024242