DISTINGUISHING BETWEEN CONDUCT DISORDER WITH HIGH VERSUS LOW LEVELS OF CALLOUS-UNEMOTIONAL TRAITS

RUTH PAULI

A thesis submitted to the University of Birmingham for the degree of Doctor of Philosophy

The Centre for Human Brain Health (CHBH)

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

September 2019

ABSTRACT

The aim of this thesis was to investigate differences between conduct disorder with high versus low levels of callous-unemotional traits. Differences in parenting, grey matter volume and facial emotion recognition ability were investigated using univariate and machine learning methods. In Chapter 3, youths with conduct disorder experienced more negative and less positive parenting than typically developing youths. The high callous-unemotional group also experienced less positive parenting than the low callous-unemotional group. All groups were classified with above-chance accuracy. In Chapter 4, when controlling for ADHD, youths with conduct disorder exhibited reduced grey matter volume in the insulae relative to typically developing youths. Youths with conduct disorder and high callous-unemotional traits exhibited additional reductions in the left orbitofrontal cortex. All groups were classified with above-chance accuracies. In Chapter 5, youths with conduct disorder – regardless of callous-unemotional traits – were poorer at recognising emotions than typically developing youths. Youths with conduct disorder were classified against typically developing youths at above-chance levels, but the classifier did not exceed chance when distinguishing between high and low callous-unemotional groups. Together, these findings indicate both similarities and differences in conduct disorder with high versus low levels of callous-unemotional traits, consistent with different developmental pathways to similar outcomes.

ACKNOWLEDGEMENTS

I am indebted to my supervisors, Dr Stephane De Brito, Dr Pia Rotshtein and Professor Peter Tino, for their commitment to overseeing this thesis over the last three years. They have all been generous with their time, guiding me in their different areas of expertise, and this thesis would not have been possible without all three of them.

My friends and colleagues at the University of Birmingham have also provided me with invaluable support and companionship. Among many others, I would especially like to thank Yidian Gao, Lizzie Martin, Jack Rogers, Maartje Spetter and Ross Wilson, whose company I have appreciated so much. Thanks also to my family, Diana, Roger, Daniel and Lucy, who have always supported and encouraged me in every way.

Finally, I am immensely grateful to the many children, families, teachers, other professionals, and members of the FemNAT-CD consortium, who contributed so much to this thesis. It has been a great privilege to work with this consortium, and with the children and families who shared their stories and gave their time so generously to make this research possible.

Funding

My PhD studentship was funded by the Biosciences and Biotechnology Research Council (BBSRC) as part of the Midlands Integrated Biosciences Training Partnership (MIBTP). The FemNAT-CD project was funded by the European Union's Seventh Framework Programme for research, technological development and demonstration, under grant agreement no 602407.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

KEY ABBREVIATIONS

Angle-GMLVQ	Angled-based Generalised Matrix Learning Vector Quantisation
ASPD	Antisocial Personality Disorder
CBCL	Child Behaviour Checklist
CD	Conduct disorder
CD/HCU	Conduct disorder with high levels of callous-unemotional traits
CD/LCU	Conduct disorder with low levels of callous-unemotional traits
CD/mixed	Conduct disorder with both high and low levels of callous-unemotional traits (undifferentiated/mixed group)
CU	Callous-unemotional
GEM	Griffiths Empathy Scale
HCU-TD	Model classifying CD/HCU and TD groups
HCU-LCU	Model classifying CD/HCU and CD/LCU groups
LCU-TD	Model classifying CD/LCU and TD groups
MCER	Macro-averaged classification error rate
Mixed-TD	Model classifying CD/mixed and TD groups
ICU	Inventory of Callous-unemotional Traits
K-SADS-PL	Schedule for Affective Disorders and Schizophrenia in School-Age Children: Present and Lifetime Version
ODD	Oppositional defiant disorder
PCL-R	Psychopathy Checklist-Revised
PDS	Pubertal Development Scale
RPQ	Reactive-Proactive Aggression Questionnaire
SVM	Support Vector Machine
TD	Typical development/typically developing
WASI	Wechsler Intelligence Scale

WAIS	Wechsler Intelligence Scale
WISC	Wechsler Intelligence Scale
YPI	Youth Psychopathy Inventory

CHAPTER 1: INTRODUCTION TO CONDUCT DISORDER AND CALLOUS-UNEMOTIONAL TRAITS

1.1 Overview and Research Questions

Youths who engage in antisocial behaviour are not all alike (Frick & White, 2008; Frick & Viding, 2009). Temperamental and genetic predispositions interact with early life experiences to create multiple, complex pathways to antisocial behaviour. Understanding these various aetiologies and their resultant presentations is essential for therapeutic success (*e.g.*, Högström, Enebrink, & Ghaderi, 2013). In this thesis, I investigate how youths with a severe form of conduct disorder (CD), characterised by high levels of callous-unemotional (CU) traits, differ from youths with a milder form of CD with lower levels of CU traits. These subtypes are referred to as CD/HCU and CD/LCU respectively. The thesis focuses on differences at both the group and individual level, in parenting, grey matter volume and emotion recognition abilities. In each case, traditional statistical methods are first used to investigate group level differences between CD/HCU, CD/LCU and typical development (TD). A multivariate machine learning classifier is then used to quantify the extent to which differences are predictive of CD/HCU, CD/LCU and TD at the individual level. If group level differences do not translate into reliable markers at the individual level, then their practical relevance will remain limited. The overarching aim of this research is thus to combine complementary methodologies that deepen our understanding of differences between CD/HCU, CD/LCU and TD.

Chapter 1 introduces the literature that motivated this thesis. First, an introduction to CD and CU traits is provided, as well as related concepts such as psychopathy. The

evidence for differences in the aetiology and neurobiology of these subtypes is also reviewed. Chapter 1 concludes with a brief overview of the three experimental chapters that form the main body of this thesis.

1.2 Conduct Disorder

1.2.1 Definition and Prevalence

CD is a diagnosis given to children and adolescents who engage in severe and persistent antisocial behaviour. CD is the most severe of the childhood behavioural disorders recognised by the American Psychiatric Association (APA), constituting persistent violation of the rights of others and failure to adhere to age-appropriate social norms (APA, 2013). The diagnostic criteria for CD include aggression to people and animals, deceitfulness or theft, destruction of property and serious violations of rules, accompanied by significant social or academic impairment (see **Figure 1**). Estimates of population prevalence vary (Fairchild *et al*., 2019). In the United Kingdom, prevalence has been estimated at 0.8% for girls and 2.1% for boys, with a combined prevalence of all behavioural disorders as high as 4.2% (Maughan, Rowe, Messer, Goodman, & Meltzer, 2004). In the United States, as many as 9.5% of children meet the criteria for CD (Nock, Kazdin, Hiripi, & Kessler, 2006). The societal impact is enormous. Childhood conduct problems are associated with school dropout, educational failure, reliance on welfare sources, substance abuse, unstable relationships and partner violence (Kessler, Foster, Saunders, & Stang, 1995; Bardone, Moffitt, Caspi, Dickson, & Silva, 1996). Therapeutic and social services interventions alone for children with conduct problems cost an estimated £5960 per child *per annum* (Romeo, Knapp, & Scott, 2006), while the weekly cost of crime committed by youths aged 10-21 years amounts to

approximately £23,000,000 in the UK (Prince's Trust, 2010). Prevention and treatment of conduct problems is thus a key priority in youth mental health.

Diagnostic Criteria for Conduct Disorder according to the DSM-5

A. At least three of the following 15 criteria must have been present during the last 12 months, with at least one present during the last six months:

Aggression to People and Animals

1. *Often bullies, threatens, or intimidates others.*
2. *Often initiates physical fights.*
3. *Has used a weapon that can cause serious physical harm to others (e.g., a bat, brick, broken bottle, knife, gun).*
4. *Has been physically cruel to people.*
5. *Has been physically cruel to animals.*
6. *Has stolen while confronting a victim (e.g., mugging, purse snatching, extortion, armed robbery).*
7. *Has forced someone into sexual activity.*

Destruction of Property

8. *Has deliberately engaged in fire setting with the intention of causing serious damage.*
9. *Has deliberately destroyed others' property (other than by fire setting).*

Deceitfulness or Theft

10. *Has broken into someone else's house, building, or car.*
11. *Often lies to obtain goods or favors or to avoid obligations (i.e., "cons" others).*
12. *Has stolen items of nontrivial value without confronting a victim (e.g., shoplifting, but without breaking and entering; forgery).*

Serious Violations of Rules

13. *Often stays out at night despite parental prohibitions, beginning before age 13 years.*
14. *Has run away from home overnight at least twice while living in the parental or parental surrogate home, or once without returning for a lengthy period.*
15. *Is often truant from school, beginning before age 13 years.*

B. The disturbance in behaviour causes clinically significant impairment in social, academic, or occupational functioning.

C. If the individual is age 18 years or older, criteria are not met for antisocial personality disorder.

Figure 1. Diagnostic criteria for conduct disorder

1.2.2 Heterogeneity

CD is a highly heterogeneous disorder. Youths with CD differ greatly in severity, presentation and prognosis, prompting questions as to whether CD can legitimately be considered a single disorder (Richers & Cicchetti, 1993). Indeed, there are over 32,000 unique combinations of symptoms that would qualify for a diagnosis of CD (Nock *et al*., 2006). This heterogeneity can have a serious impact on therapeutic success (Högström *et al*., 2013). Several subtyping schemes have consequently been suggested, with the aim of identifying CD subtypes that are more homogeneous. Proposed criteria include age of onset (Moffitt, Caspi, Dickson, Silva, & StanoLahey *et al*., 1998), presence of comorbid Attention Deficit Hyperactivity Disorder (ADHD; Faraone, Biedermann, Jetton, & Tsuang, 1997) and CU traits. Age of onset subtypes are now included in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013). However, probably the most widely researched subtyping method uses CU traits to identify two subgroups within CD (*e.g*., Frick & Ellis, 1999). This approach is respected because it appears to identify distinct developmental trajectories towards CD (Frick & Viding, 2009), as well as being a valid marker of severity and persistence of conduct problems (Frick, Stickle, Dandreaux, Farrell, & Kimonis, 2005). In response to a substantial evidence base, the Limited Prosocial Emotions (LPE) specifier was introduced to the DSM-5 (APA, 2013). The LPE specifier allows clinicians to distinguish between youths with CD/HCU and those with CD/LCU, within the broader diagnostic category of CD.

1.3 Callous-Unemotional Traits

1.3.1 Definition and Prevalence

CU traits consist of a limited capacity for empathy, shallow or superficial affect, unconcern about performance in important activities and lack of guilt or remorse for wrongdoing (*e.g.*, Frick & White, 2008). Together, they constitute the core affective and interpersonal features of psychopathy as defined by Cleckley (1941), and typically emerge in early-to-middle childhood (Frick, Kimonis, Dandreaux, & Farrell, 2003). Youths with CD must exhibit at least two of the four characteristics (*i.e.*, lack of remorse, low empathy, shallow affect and unconcern about performance) over a 12-month period, consistently and in multiple settings, to qualify for CD/HCU under the LPE specifier (see **Figure 2**). Using a criterion of two or more parent-endorsed CU criteria, Rowe, Maughan, Moran, Ford, Briskman and Goodman (2010) estimated the prevalence of CD/HCU within CD at 46%, equivalent to a population prevalence of approximately 0.5%. In a clinic-referred sample (Kahn, Frick, Youngstrom, Findling, & Youngstrom, 2012), 21-50% of youths with CD qualified as CD/HCU, compared to 10-32% in a community CD sample. In primary school children, the prevalence of CD/HCU was estimated at 4% (Seijas, Servera, García-Bandia, Barry, & Burns, 2018). Another clinical sample produced a more conservative estimate, although the proportion with CD/HCU was still substantial at approximately 9% (increasing to 43% when including youths with elevated CU traits but few conduct problems; Christian, Frick, Hill, Tyler, & Frazer, 1997). Children with high levels of CU traits do not inevitably develop CD (Fanti, 2013). However, they usually have at least subtle impairments in interpersonal functioning, and are at risk for other psychiatric disorders even when they

do not meet the clinical threshold for CD (Herpers, Klip, Romelse, Greven, & Buitelaar, 2016; Rowe *et al*., 2010). The risk for psychopathology conferred by CU traits is thus not limited to CD, suggesting that CU traits are indicative of a distinctive psychological profile even in the general population.

Limited Prosocial Emotions Specifier

To qualify for this specifier, an individual must have displayed at least two of the following characteristics persistently over at least 12 months and in multiple relationships and settings.

1. Lack of remorse or guilt: Does not feel bad or guilty when he or she does something wrong (exclude remorse when expressed only when caught and/or facing punishment). The individual shows a general lack of concern about the negative consequences of his or her actions. For example, the individual is not remorseful after hurting someone or does not care about the consequences of breaking rules.
2. Callous—lack of empathy: Disregards and is unconcerned about the feelings of others. The individual is described as cold and uncaring. The person appears more concerned about the effects of his or her actions on himself or herself, rather than their effects on others, even when they result in substantial harm to others.
3. Unconcerned about performance: Does not show concern about poor/problematic performance at school, at work, or in other important activities. The individual does not put forth the effort necessary to perform well, even when expectations are clear, and typically blames others for his or her poor performance.
4. Shallow or deficient affect: Does not express feelings or show emotions to others, except in ways that seem shallow, insincere, or superficial (*e.g*., actions contradict the emotion displayed; can turn emotions "on" or "off" quickly) or when emotional expressions are used for gain (*e.g*., emotions displayed to manipulate or intimidate others).

Figure 2. Criteria for the Limited Prosocial Emotions Specifier

1.3.2 Measurement in Research Settings

In research settings, CU traits are typically assessed with self-, parent- or teacher-report questionnaire measures, or a combination thereof. Examples include the Inventory of Callous-Unemotional Traits (ICU; Essau, Sasagawa, & Frick, 2006b) and the CU dimension of the Youth Psychopathic Traits Inventory (YPI; Andershed, Kerr, Stattin, & Levander, 2002). Scores above a certain threshold are defined as elevated, while scores below the threshold are considered normative or low. The threshold might be

defined based on the distribution of CU traits within the sample, for example by selecting the first quartile, tertile or median score for youths with CD within the sample (*e.g.*, Wootton, Frick, Shelton, & Silverthorn, 1997). Other studies use a fixed cut-off point (*e.g.*, Bowen, Morgan, Moore, & van Goozen, 2014). The exact criteria for CD/HCU therefore differ between studies. Minor differences in cut-off should not drastically affect research outcomes, especially since CU and psychopathic traits appear to be largely dimensional rather than taxonic (Edens, Marcus, Lilienfeld, & Poythress, 2006; Clark, 2007; Murrie, Marcus, Douglas, Lee, Salekin, & Vincent, 2007). Nonetheless, the development of a standard definition of CD/HCU is a priority now that the LPE specifier has been formally incorporated into the DSM-5 (see Kimonis, Fanti, Goldweber, Marsee, Frick, & Cauffmann, 2014). The recently developed semi-structured CAPE interview aims to address this need (Centifanti, Shaw, Atherton, Thomson, MacLellan, & Frick, 2019).

1.4 Evidence for Distinct Forms of Conduct Disorder in Youths with High versus Low Levels of Callous-Unemotional Traits

1.4.1 Differences in Aetiology

Evidence from behavioural genetics indicates that CU traits are highly heritable. Based on a sample of 3687 seven-year-old twin pairs drawn from the Twins Early Development Study (TEDS; Trouton, Spinath, & Plomin, 2002), Viding, Blair, Moffitt and Plomin (2005) estimated the heritability of CU traits to be 0.68. This means that approximately two thirds of the variation in CU traits between the children defined as having 'extreme CU traits' and the rest of the sample could be attributed to genetic differences. Antisocial behaviour in these children was also under strong genetic

influence, with 0.81 heritability compared to 0.30 for children without elevated CU traits. Heritability remained high in a follow-up study two years later when the twins were nine years old, even when controlling for ADHD (Viding, Jones, Paul, Moffitt, & Plomin, 2008). The genetic risk for CU traits remains evident in children who have been adopted at birth, although positive parenting by the adoptive parents has a clear ameliorative effect (Hyde *et al*., 2016). The associations between parenting behaviours, CD and CU traits is a very active area of research, and is the topic of Chapter 3 (*e.g*., Waller, Gardner, & Hyde, 2013; Högström *et al*., 2013; Crum, Waschbusch, Bagner, & Coxe, 2015; Waller, Gardner, Shaw, Dishion, Wilson, & Hyde, 2015; Waller *et al*., 2013; Viding, Fontaine, Oliver, & Plomin, 2009; Pasalich, Dadds, Hawes, & Brennan, 2011; Clark & Frick, 2018; Kochanska, Kim, Boldt, & Yoon, 2013; Waller *et al*., 2015; Wall, Frick, Fanti, Kimonis, & Lordos, 2016; Chinchilla & Kosson, 2016).

1.4.2 Differences in Clinical Presentation

CU traits are associated with an increased risk for CD and earlier onset of conduct problems, as well as greater severity and persistence (Jezior, McKenzie, & Lee, 2016; Rowe *et al*., 2010; Frick, Ray, Thornton, & Kahn, 2014; Frick *et al*., 2005). Even in youths without CD, high levels of CU traits predict future aggression and delinquency (Frick, Cornell, Barry, Bodin, & Dane, 2003). Youths with CD/HCU also differ qualitatively from those with CD/LCU. For example, there is evidence that youths with CD/HCU engage in more proactive aggression, harming their victims without provocation in order to achieve their goals (Frick *et al*., 2003; Caputo, Frick, & Brodsky, 1999). Youths with CD/LCU, by contrast, exhibit a more reactive style of aggression, frequently 'lashing out' in response to perceived provocation (Fanti, Frick,

& Georgiou, 2009). Punishment insensitivity has also been reported in CD/HCU (Blair, Colledge, & Mitchell, 2001; O'Brien & Frick, 1996; Fanti, Panayiotou, Lazarou, Michael, & Georgiou, 2016), with youths with CD/HCU displaying a more reward-orientated style of learning relative to TD youths (Scerbo, Raine, O'Brien, Chan, Rhee, & Smiley, 1990). Fearless temperament in early childhood has been associated with the subsequent development of CD/HCU (Barker, Oliver, Viding, Salekin, & Maughan, 2011). However, neither fearlessness nor punishment insensitivity are consistently linked to CU traits or CD/HCU (*e.g.*, Mills-Koonce, Wagner, Willoughby, Stifter, Blair, Granger, & The Family Life Project Key Investigators, 2015; Byrd, Hawes, Burke, Loeber, & Pardini, 2018).

1.4.3 Neurocognitive Theory

According to Blair (*e.g.*, Blair, 2013; Blair, 2001; Blair, Mitchell, & Blair, 2005), differences in proactive and reactive aggression are key to understanding the neurobiology of CD/HCU and CD/LCU (see **Figure 3**). Proactive aggression, which is elevated in CD/HCU but not in CD/LCU, is linked to dysfunction in a network of brain regions including the amygdala, caudate, orbitofrontal and ventromedial prefrontal cortex and anterior insula (Blair, 2013). In youths with high levels of psychopathic (including CU) traits, negative affective stimuli (particularly expressions of pain, fear and sadness) are not experienced as inherently aversive or salient. Consequently, the amygdala does not perform its normal role in stimulus-reinforcement learning. The striatum (which is implicated in reward learning) also fails to perform its role in both stimulus-reinforcement and response-outcome learning. As a result of the dysfunction of the amygdala and striatum, the expected value of social cues and associated

behavioural responses are poorly represented in the orbitofrontal cortex, and decision-making is impaired. Blair suggests that the anterior insula (via its involvement in response initiation) is also implicated in the expression of psychopathic traits, although this is secondary to the failure to associate negative affective stimuli with punishment (Blair, 2001). However, the evidence for a central role of the anterior insula in CD/HCU has grown stronger in recent years. The anterior insula is key to interoception and integration of bodily states with cognition and action (Namkung, Kim, & Sawa, 2017; Craig, 2009). It has also been consistently associated with norm compliance in trust games (Bellucci, Feng, Camilleri, Eickhoff, & Krueger, 2018), is often reduced in volume in CD (*e.g.*, Rogers & De Brito, 2016) and has been associated with affective introspection in CD (Sethi, O'Nions, McCrory, Bird, & Viding, 2018). Together, these studies suggest that the anterior insula might be more directly implicated in CD/HCU than previously thought.

Reactive aggression, by contrast, is elevated in both CD subtypes. It is associated with hyper-responsivity of the basic neural threat circuit, which includes the amygdala, hypothalamus and periaqueductal gray. Hyper-responsivity in this network might reflect priming caused by adverse events (*e.g.*, childhood abuse) or insufficient regulation by the orbitofrontal and anterior cingulate cortices (Blair, 2013; Blair *et al.*, 2005). Dysfunction in the threat circuit is hypothesised to result in increased startle response, heightened reactive aggression and difficulty recognising expressions of anger (Blair *et al.*, 2005). According to Blair, therefore, youths with CD/LCU are characterised by dysfunction in the basic threat circuit, while youths with CD/HCU are primarily characterised by dysfunction in the regions associated with proactive aggression, and additionally by dysfunction in the basic threat circuit. In the case of CD/HCU, however,

Blair argues that reactive aggression is more likely to occur in the context of frustration (*e.g.*, not achieving a desired response) than to threat (Blair, 2010).

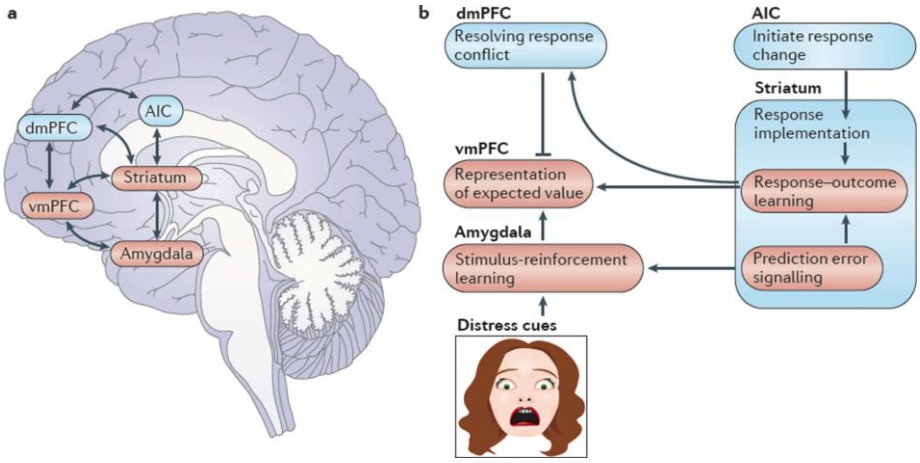


Figure 3. Blair's model of psychopathy (image reproduced from Blair, 2013). Regions implicated in psychopathy **(a)** and their associated functional impairments **(b)**

1.4.4 Neuroimaging Evidence

Functional magnetic resonance imaging (fMRI) studies commonly report reduced amygdala reactivity to fearful facial expressions in CD/HCU, and this has been linked to proactive aggression (Marsh *et al*., 2008; Jones, Laurens, Herba, Barker, & Viding, 2009; Lozier, Cardinale, VanMeter, & Marsh, 2014). Likewise, a negative association between CU traits and amygdala activity has been reported in an affective theory of mind task (Sebastian, McCrory, Cecil, Lockwood, De Brito, Fontaine, & Viding, 2012). There is also some evidence for differences in amygdala reactivity when youths with CD/HCU are compared to youths with CD/LCU. For example, using a different paradigm, Viding, Sebastian, Dadds, Lockwood, Cecil, De Brito and McCrory (2012) demonstrated a reduced amygdala response to pre-attentively presented fearful (versus

calm) faces in children with CD/HCU compared to those with CD/LCU. These faces were not consciously perceived, implying that reductions in amygdala response are driven by a relatively primitive 'bottom-up' attentional failure. By contrast, youths with CD/LCU exhibited an elevated amygdala response relative to the TD group (Viding *et al*., 2012), which might relate to their increased levels of reactive aggression (White *et al*., 2015) Contrary to these findings, in a subclinical population (Dotterer, Hyde, Swartz, Hariri, & Williamson, 2017), there was no evidence for an association between CU traits and reduced amygdala reactivity to angry or fearful expressions. However, antisocial behaviour was associated with increased amygdala reactivity to angry faces, as expected. Structurally, there is very little research directly comparing youths with CD/HCU versus CD/LCU, although there is some support for grey matter differences in the brain regions identified by Blair (Blair, 2013; Sebastian, De Brito, McCrory, Hyde, Lockwood, Cecil, & Viding, 2016; De Brito *et al*., 2009; Fairchild, Hagan, Walsh, Passamonti, Calder, & Goodyer, 2013; Sterzer, Stadler, Poustka, & Kleinschmidt, 2007; Fairchild, Toschi, Hagan, Goodyer, Calder & Passamonti, 2015; Raschle *et al*., 2018; Rogers & De Brito, 2016). The literature on structural differences is reviewed in Chapter 4. In summary, there is consistent evidence for clinical, psychological and neurobiological differences between youths with CD/HCU and CD/LCU, although these differences are not yet fully understood.

1.5 Childhood Comorbidities and Related Disorders in Adulthood

Finally, when discussing heterogeneity in CD and the differences between CD/HCU and CD/LCU, it is important to consider the relationships between CD and other forms of psychopathology. Although not the topic of this thesis, the overlap between CD and

other disorders – both comorbid and subsequently developing – is highly relevant to understanding differences between subtypes (see Chapter 6 for further discussion). There is often a clinical overlap with other childhood psychiatric disorders, as well as conceptual overlap with related disorders in adulthood. This literature is reviewed briefly here, to provide additional theoretical context for the research questions in this thesis.

1.5.1 Comorbidities

Comorbid externalising and internalising disorders are common in CD (Nock *et al*., 2006). In general (and possibly excepting anxiety), comorbid internalising disorders seem to develop subsequent to CD; for example, it is more common for a child with conduct problems to develop subsequent depression than for a child with depression to develop subsequent conduct problems (McDonough-Caplan, Klein, & Beauchaine, 2018; Loeber, Burke, Lahey, Winters, & Zera, 2000; Nock *et al*., 2006). This sequence of morbidities raises the possibility that CD evokes a negative environment (*e.g*., exposure to violence, poor social relationships) that in turn increases the risk for internalising problems. Consistent with this, comorbidity rates are higher in CD than in many other disorders (*e.g*., Maughan *et al*., 2004). However, CU traits have been shown to be negatively correlated with physiological indices of anxiety (Fanti *et al*., 2016; see also Chapter 6, section 6.4.4), suggesting that in this respect, youths with CD/HCU might in fact be less vulnerable than those with CD/LCU to comorbid internalising problems.

ADHD, by contrast, often seems to precede the onset of CD (Nock *et al*., 2006). While it is not completely clear whether ADHD is associated with increased risk for CD, it is

clear that comorbid ADHD – in particular the hyperactive-impulsive dimension – is associated with earlier onset CD, and more severe and persistent behavioural problems (Loeber *et al*., 2000). In a review, Beauchaine, Hinshaw and Pang (2010) argue that trait impulsivity – which underlies the hyperactive-impulsive component of ADHD – is common to most externalising disorders, and is highly heritable. However, an exacerbating social environment (*e.g*., poor parenting) is necessary for this underlying impulsivity to manifest as CD. In the absence of an adverse environment, the more likely outcome is 'pure' ADHD (Beauchaine *et al*., 2010). In summary, several disorders are commonly comorbid with – and perhaps not completely separate from – CD (*e.g*., ADHD, anxiety, mood disorders *etc*.). However, their relationships with CD differ, with internalising disorders perhaps resulting from CD, while ADHD is better described as a risk factor for more severe CD. The role of CU traits in shaping the relationship between these different disorders is not yet known, but is likely to be important, as discussed further in Chapter 6.

1.5.2 Antisocial Personality Disorder

Antisocial personality disorder (ASPD) is an approximately equivalent diagnosis to CD in adults. The criteria for ASPD include criminal behaviour, deception, impulsivity, aggression, irresponsibility and lack of remorse, and can only be diagnosed in individuals who had a history of CD before the age of 15 years (APA, 2013). The diagnosis of ASPD was historically strongly influenced by that of psychopathy. However, reflecting their pragmatic philosophy, the APA moved towards placing more weight on overt antisocial behaviour than on 'covert' psychopathic traits (Follette & Houts, 1996). In the DSM, the diagnosis of psychopathy was first subsumed, and then

replaced, by ASPD. However, psychopathy is still widely recognised, and is now generally regarded either as a qualitatively distinct disorder within ASPD (*e.g.*, Kosson, Lorenz, & Newman, 2006) or as a more severe sub-variant (*e.g.*, Coid & Ullrich, 2010). CD is a clear risk factor for ASPD (Copeland, Shanahan, Costello, & Angold, 2009). It has been estimated that between 33-88% of adults meeting the diagnostic criteria for ASPD[1] have a history of CD (Simonoff, Elander, Holmshaw, Pickles, Murray, & Rutter, 2004), while approximately 33% of youths with CD transition to ASPD in adulthood (Robins, 1978).

1.5.3 Psychopathy

Under various guises, psychopathy has been recognised as a mental disorder since the early 19th century (Pinel, 1806/2017). This recognition arose from the observation that the antisocial behaviour exhibited by some patients was not 'ordinary' antisocial behaviour, and did not arise from cognitive impairments or psychosis, but rather stemmed from an underlying affective pathology characterised by shallow, fleeting and volatile emotions. However, these early conceptualisations of psychopathy were extremely broad (Millon, Simonsen, Birket-Smith, & Davis, 2002). Psychopathy in the narrower modern sense was first defined by Cleckley (1941). Cleckley described a group of patients who were superficially high functioning but had severely impoverished emotional lives. These patients were not necessarily deeply sadistic, but were largely devoid of empathy and had a marked proclivity towards antisocial behaviour. They appeared unable to learn from experience, engaging in repeated, socially outrageous and self-damaging behaviour despite superficially good intellectual

[1] These estimates ignore the diagnostic criterion for CD to have been present before giving a diagnosis of ASPD.

insight. Their behaviour could not easily be explained by relatable motives, such as intense emotion, financial gain or social status. Indeed, it often appeared largely incomprehensible to observers, indicating a deeply pathological origin despite the absence of mental disorder as it was generally understood at the time. It was this inferred underlying pathology, rather than the concomitant antisocial behaviour *per se*, that was considered central to psychopathy as a disorder (Cleckley, 1941). Although more recent conceptualisations of psychopathy have placed a slightly greater emphasis on criminal offending (Hare, 1999), the primary emphasis remains on affective impoverishment and lack of empathy. The most widely used assessment tool for psychopathy in clinical and forensic settings – the Psychopathy Checklist-Revised (PCL-R; Hare, 2003) – is still largely based on Cleckley's work. The proportion of youths with CD/HCU who transition to psychopathy in adulthood is currently unknown, although there is evidence for moderate stability of psychopathic traits across the transition from childhood to adulthood (Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007).

1.6 Summary

In summary, there are differences in aetiology, severity, presentation and prognosis of CD/HCU and CD/LCU. Youths with elevated CU traits are at increased risk for CD due to a genetically influenced temperamental predisposition towards antisocial behaviour. CD/HCU is characterised by a distinctive style of antisocial behaviour, which is more callous and proactive than that seen in CD/LCU. While not immune to the effects of the environment (especially in early childhood), CD/HCU appears to be highly heritable (*e.g.*, Viding *et al.*, 2005). By contrast, youths with CD/LCU tend to engage in milder

and less persistent antisocial behaviour, typically starting at a later age. CD/LCU is characterised by 'hot-headed', reactive aggression, which might be explained by a heightened sensitivity to provocation. Youths with CD/LCU possess some degree of genetic vulnerability to antisocial behaviour, but CD/LCU appears to be less heritable than CD/HCU (Viding *et al*., 2005). The neurobiological factors that sensitise these youths to environmental risks are likely related to emotional over-reactivity rather than the primarily affective deficits seen in CD/HCU.

1.7 Thesis Outline

The following chapters focus on the relevant methodologies, three research areas where the differences between CD/HCU, CD/LCU and TD are not yet fully clear (parenting, grey matter volume and emotion recognition ability), and a general discussion of the experimental results. First, Chapter 2 describes the methods used throughout the thesis: machine learning, neuroimaging and the FemNAT-CD sample (from which all data in this thesis are taken). Three experimental chapters then follow. In Chapter 3, differences in positive and negative parenting are investigated. Chapter 4 focuses on grey matter volume differences, and Chapter 5 focuses on differences in facial emotion recognition ability. The aim in Chapters 3-5 is to elucidate the extent to which the three groups (CD/HCU, CD/LCU and TD) differ from each other, and the reliability of these differences when predicting the status of individual youths. Specific hypotheses regarding group differences are provided at the start of each chapter. Finally, the findings from these three experimental chapters are integrated and discussed in Chapter 6, along with their implications for future research in the field of CD and CU traits.

Please note that findings from Chapter 3 (differences in positive and negative parenting) have been submitted for publication, and findings from both Chapters 3 and 4 (grey matter volume differences) have previously been presented at international conferences.

CHAPTER 2: METHODOLOGIES AND THE FEMNAT-CD PROJECT

2.1 Overview

Chapter 2 focuses on the methodologies employed in the following experimental chapters. First, the principles of machine learning classification are described, with a particular focus on Angle-Based Generalised Matrix Learning Vector Quantisation (Angle-GMLVQ), which is the main classifier used in this thesis (Bunte, Baranowski, Arlt, & Tino, 2016). Second, magnetic resonance imaging (MRI) is introduced. Third, the FemNAT-CD project, from which all data presented in this thesis are drawn, is introduced, and the key measures relevant to all three experimental chapters are described.

2.2 Machine Learning Classification

2.2.1 Overview of Classification

Definitions, descriptions and terminology used in this section are based largely on introductory texts by Flach (2012) and James, Witten, Hastie and Tibshirani (2013). Machine learning is a field of computer science in which data are analysed using algorithms that 'learn' (*i.e.*, adapt in response to data) without explicit programming (Flach, 2012; James *et al*., 2013). Classification is a form of supervised learning, in which input variables are mapped onto an output using labelled training data. For example, a classifier is trained to predict class labels (*e.g.*, CD or TD) based on a set of input variables (*e.g.*, regional grey matter volumes). This differs from regression, which follows similar principles, but with real-valued 'labels', *e.g.*, number of CD symptoms. Crucially, in all forms of supervised learning, the labels of the data are pre-determined

and known during the training phase. This contrasts with unsupervised learning, where data are unlabelled and the purpose is to derive information from the structure of the data, for example by identifying important clusters (a third type, semi-supervised learning, combines supervised and unsupervised techniques when class labels are known for only a subset of the data). Classification is thus a family of supervised learning techniques characterised by the prediction of categorical class labels. Although classification is a diverse family of methods, the basic aim of all classifiers is to learn a rule that separates members of different classes with maximum accuracy (Flach, 2012).

Classifiers are highly sensitive to subtle and complex patterns in large data sets, which would be difficult to detect with traditional statistical methods. Being data-driven rather than theory-driven, machine learning classifiers are suited to making accurate predictions about individuals without relying on pre-existing theory. Given its potential applications in clinical diagnosis, classification has become ubiquitous in neuroimaging research. Prediction of depression, psychosis, Alzheimer's, Parkinson's and other neurological diseases are all active areas of research (Janssen, Mourão-Miranda, & Schnack, 2018; Orru, Petttersson-Yeo, Marquand, Sartori, & Mechelli, 2012). In recent years, a small number of researchers have applied machine learning classifiers to neuroimaging data in the field of psychopathy and conduct disorder (Cope *et al*., 2014; Steele, Rao, Calhoun, & Kiehl, 2017; Sato *et al*., 2011; Zhang, Cao, Wang, Wang, Yao, & Huang, 2018a; Zhang *et al*., 2018b). This literature is reviewed in Chapter 4.

2.2.2 Key Concepts and Terminology

2.2.2.1 Features

In machine learning contexts, the input variables used to predict class membership are known as features. Features are conceptually similar to independent variables in traditional statistics, *e.g.*, questionnaire items. There is some ambiguity in the literature as to whether the feature is the variable itself, *e.g.*, 'age', or its corresponding score, *e.g.*, '10 years'; see Kohavi & Provost, 1998). In this thesis, 'feature' is used as a synonym of 'variable' rather than as the associated value of the variable. When visualised geometrically, each feature constitutes a different spatial dimension. For example, a hypothetical classifier with two features can be visualised as a simple two-dimensional line graph, with one feature on the x-axis and one feature on the y-axis. Each additional feature will form an additional dimension in multidimensional space.

2.2.2.2 Labels

The output or dependent variables are the class labels, *e.g.*, presence or absence of CD. For classification tasks, labels are categorical, but not necessarily dichotomous. Labels are usually treated as nominal rather than ordinal, so that in multiclass problems, all misclassifications are penalised to the same extent. For example, if classifying a disease into severe, moderate and mild forms, then a misclassification of a mild case as a severe case is treated the same way as a misclassification into a moderate case. This differs from regression methods.

2.2.2.3 Instances

An instance refers to the vector of feature scores for each participant, *e.g.*, the set of questionnaire item responses associated with that participant. An instance is thus effectively a single data point in multidimensional space.

2.2.2.4 Training and Testing Sets

Data are divided into training and testing sets. Training data consist of labelled data (*i.e.*, the labels are known to the classifier) and are used to learn a function that maps the feature values to the class labels. Once this function is learned, its performance is tested on the previously unseen, unlabelled testing set. The classifier uses the function to generate a predicted class label for each instance in the testing set, and these are then compared to the real class labels.

2.2.2.5 Validation

Validation refers to the process of testing the trained classifier on previously unseen data, to check how well the learned function generalises. Most simply, a proportion of the data (*e.g.*, 20%) is 'held out' for testing. This is known as a holdout validation, and is the approach used in Chapters 3-5. Other techniques include leave-one-out validation, in which the classifier is trained multiple times (once per instance) with only one instance set aside for testing each time, and *k*-fold cross-validation, in which the data are divided into *k* folds of equal size, and the classifier is trained and tested *k* times with each fold reserved for testing once (see **Figure 4**). Classifier performance is then averaged across the folds. These latter techniques are computationally more expensive, but they are useful in smaller samples where a single trained classifier might not generalise well to new data.

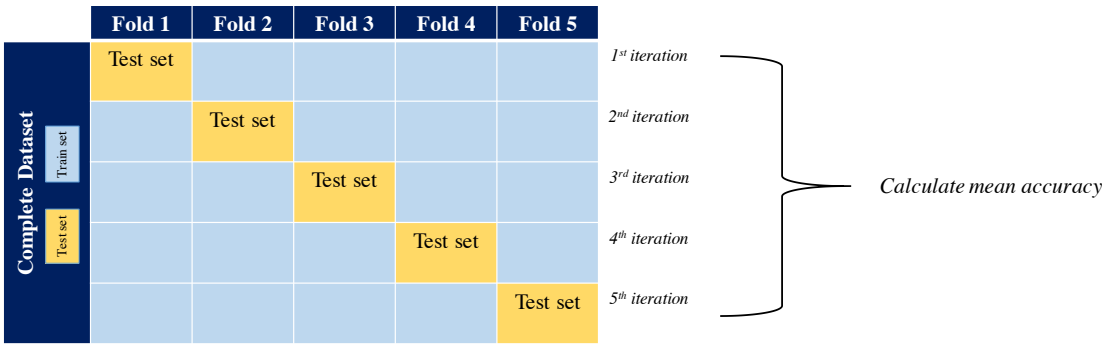


Figure 4. An illustration of *k*-fold cross validation where $k = 5$. A simple holdout validation is equivalent to the 1st iteration of *k*-fold, using the whole dataset

2.2.3 Support Vector Machine (SVM)

SVMs are relatively simple binary classifiers for use with two-class problems. SVM is a geometric model; each feature constitutes a dimension in multidimensional space (the instance space). Instances are described by a set of Cartesian coordinates within the instance space. A decision boundary, or separating hyperplane, is then constructed, which best separates instances of different classes. The decision boundary is linear in the basic SVM, but non-linear transformation kernels can be applied.

Linearly separable data (*i.e.*, when a decision boundary can be placed so that all instances belonging to a class are on the same side of the boundary) result in multiple possible decision boundaries. In this case, the optimal boundary will be the one that maximises the distances from the boundary to the closest instances of each class. In SVM, the decision boundary is thus determined only by those instances – known as support vectors – that lie closest to it in the instance space (see **Figure 5**). Since in practice most data are not linearly separable, a 'soft' margin is generally used; this allows some instances to lie on the wrong side of the decision boundary, but applies a penalty so that misclassifications are minimised. Non-linear transformations (using, for

example, a 2nd or 3rd order polynomial transformation kernel) can also be applied to the decision boundary, which can improve classifier performance if the classes are not well separated by a simple linear boundary.

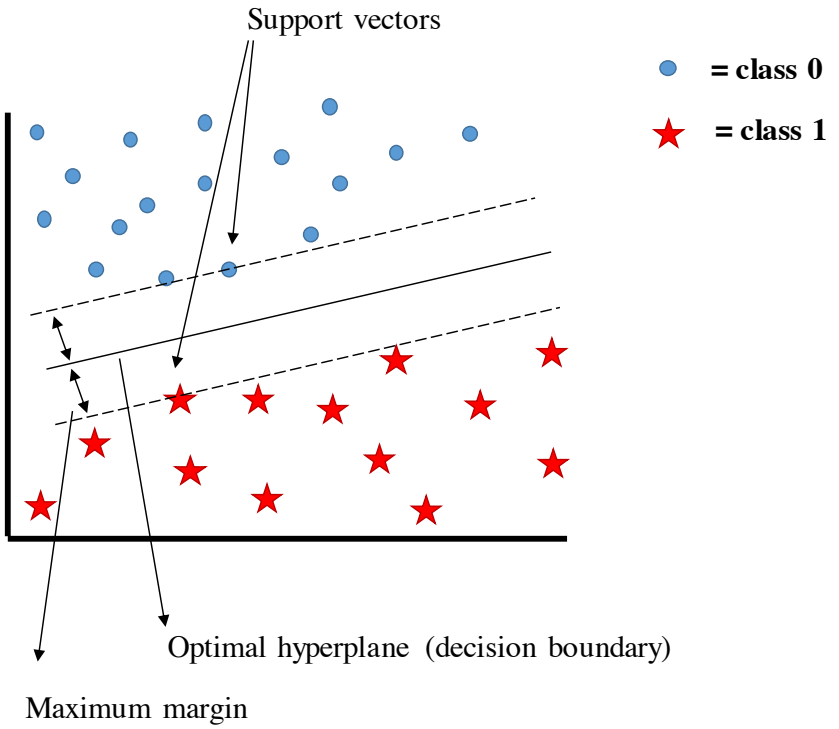


Figure 5. The decision boundary in a support vector machine (SVM) classifier

2.2.4 Angle-Based Generalised Matrix Learning Vector Quantisation (Angle-GMLVQ)

Angle-GMLVQ (Bunte *et al*., 2016) is an extension to the Learning Vector Quantisation (LVQ) group of classifiers (Kohonen, 1995). These are prototype-based classifiers, which classify data based on similarity to constructed representative prototypes ('codebook vectors'). In LVQ, a number of prototypes – at least one for each class – are

generated at some initial locations within the instance space at the initiation of the training phase. Each instance in the training set is then considered in turn. First, the closest (most similar) prototype of the same class and the closest prototype of the other class are selected. The same-class prototype is then moved closer to the instance, while the other-class prototype is moved further away. At the completion of the training phase, each prototype should thus be located in a representative location for its class (or sub-sample of its class, when there are multiple prototypes per class). In the testing phase, each instance is predicted to share the class label of the nearest (most similar) prototype. This is a 'winner-takes-all' rule, because each instance is classified as the class of the nearest prototype without any weighting or competition from surrounding prototypes (Kohonen, 1995)[2]. A simplified version of this process is illustrated in **Figure 6**.

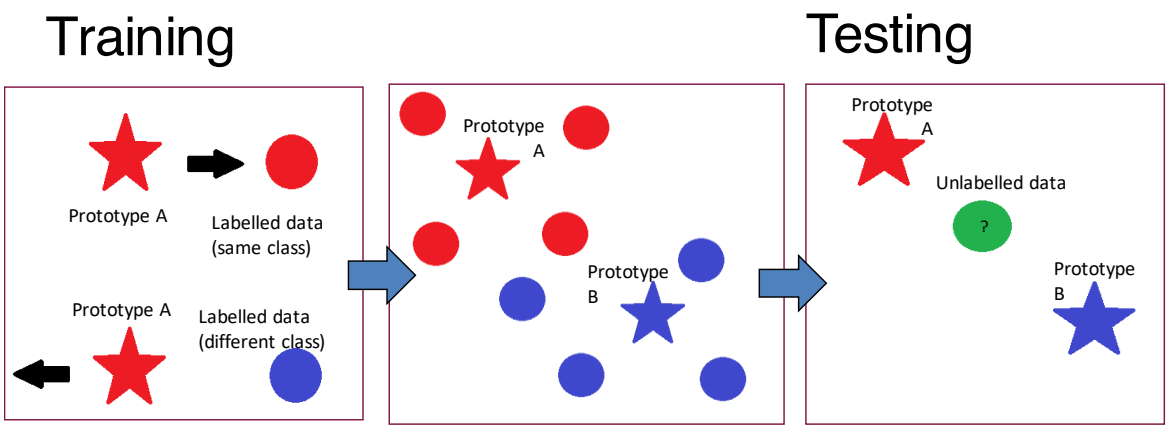


Figure 6. A simplified illustration of the training and testing stages in a winner-takes-all prototype classifier such as Angle-GMLVQ. The prototypes are positioned so as to be representative during the training phase, and unlabelled data in the test set are classified as belonging to the class of the nearest prototype

[2] Winner-takes-all can also refer to the training rule, in which only one same-class prototype (the 'winner') is moved closer to the instance, while the positions of the other same-class prototypes are not updated.

Generalised matrix LVQ (GMLVQ) is an adaptation of LVQ that introduces a matrix tensor to define the distance measure. The matrix represents individual features as well as pairwise dependencies or 'interplays' between each pair of features. This implicitly scales and rotates the data, providing a distance measure that is suited to the particular classification task, rather than using the standard Euclidean distance of the LVQ classifiers (Sato & Yamada, 1996; Schneider, Biehl, & Hammer, 2009; Alahmadi, Shen, Fouad, Luft, Bentham, Kourtzi, & Tino, 2016). Angle-GMLVQ is an extension of GMLVQ, which uses angles between feature vectors rather than distance between feature vectors as the measure of similarity/dis-similarity between instances. As a result, Angle-GMLVQ is sensitive to relative differences in magnitude between features, rather than the magnitudes of individual feature scores. For example, a participant who scores '4, 4, 4' on a set of questionnaire items will not be distinguished from a participant who scores '5, 5, 5' on the same set of items, because the differences between individual features (*i.e.*, zero) is the same for both participants[3]. Angle-GMLVQ is thus appropriate for situations where differences between features are key to classification, and when absolute values might be confounded by irrelevant factors *e.g.*, when classifying based on regional differences in grey matter volume or Likert-scale questionnaire data (which might be affected by individuals' overall response biases).

2.2.5 Performance Measures

Measures of classifier performance typically assess both overall performance (*e.g.*, classification accuracy or error rate) and performance for each class. In this thesis, the TD group is defined as the negative class (or, where classifying CD/HCU and CD/LCU,

[3] The classifier is insensitive to differences that are scale invariant.

the latter is the negative class). 'True' refers to correct classifications and 'false' refers to incorrect classifications. Thus, true positives are correctly classified members of the positive class and false positives are incorrectly classified members of the negative class, *etc*.

Model performance is assessed using the following performance metrics, where TP = true positives, FP = false positives, TN = true negatives and FN = false negatives:

- **Accuracy:** proportion of all participants classified correctly.
    - *Accuracy = (TN + TP) / (TN + TP + FN + FP)*
- **Positive predictive value (PPV):** true positives as a proportion of all positive classifications. This is a measure of confidence in positive classifications. PPV is sensitive to the prevalence of positive class members in the sample.
    - *PPV = TP / (TP + FP)*
- **Negative predictive value (NPV):** true negatives as a proportion of all negative classifications. This is a measure of confidence in negative classifications. NPV is sensitive to the prevalence of negative class members in the sample.
    - *NPV = TN / (TN + FN)*
- **True positive rate (TPR):** true positives as a proportion of all genuine positives, *i.e*. the proportion of CD (or CD/HCU) participants who are classified correctly. This is a measure of the ability of the classifier to detect members of the positive class. TPR is also known as sensitivity.
    - *TPR = TP / (TP + FN)*
- **True negative rate (TNR):** true negatives as a proportion of all genuine negatives, *i.e*., the proportion of TD (or CD/LCU) participants who are classified

correctly. This is a measure of the ability of the classifier to detect members of the negative class. TNR is also known as specificity.

  - *TNR = TN / (TN + FP)*

- **Macro-averaged classification error rate (MCER):** MCER is the mean of the error rates for the positive and negative classes, adapted from Baccianella, Esuli and Sebastiani (2009) and Fouad and Tino (2012). When class sizes are imbalanced, a classifier can perform superficially well by predicting that most instances belong to the larger class. This results in good performance for the larger class and poor performance for the smaller class, even when overall accuracy is good. MCER corrects for this problem by taking the mean across the two classes.

  - $MCER = {}^{1}/_{k} \sum_{k=1}^{k} \frac{\sum_{y_i=k} |y_i - \hat{y}_i|}{v_k}$

    where k is the number of classes, v_k is the number of instances whose true class is 1, y is the actual class label and $\hat{y}$ is the estimated (predicted) class label[4].

Another common measure of classifier performance is Area Under the Curve (AUC). This is a measure of the trade-off between true positives and false positives at different thresholds for sensitivity to the positive class. As the threshold changes so that a greater proportion of positives are correctly identified (*i.e.*, true positives), the number of false positives also increases. AUC thus measures the ability of a classifier to distinguish

[4] Note that this equation is designed for ordinal regression, where there are multiple classes, and misclassifying an instance as belonging to a neighbouring class is 'less incorrect' than misclassifications into a distant class. Since all the models presented in this thesis contain only two classes, this weighting of errors is irrelevant here.

between classes – the greater the AUC, the better the ability to discriminate. AUC is not used as a performance measure in this thesis. First, it requires the use of different sensitivity thresholds, and these are not easily generated in Angle-GMLVQ. Second, AUC is most useful when there is an obvious real-world trade-off between true and false positives. For example, a cancer-screening test should identify all positive cases, while minimising false positives that could lead to unnecessary, invasive and costly medical procedures. AUC is thus a useful measure for deciding which of a set of classifiers best achieves this aim. In this thesis, by contrast, the ability to distinguish between the classes is a question of scientific interest rather than practical necessity, and AUC is of less interest[5].

2.2.6 Limitations and Ethical Considerations

Despite its successes, the increasing (and sometimes uncritical) use of machine learning has generated controversy in recent years. From a scientific perspective, the field has been criticised for treating algorithm development as challenges to be 'won' based on high performance, rather than an exercise in empirical understanding (Sculley, Snoek, Wiltschko, & Rahimi, 2018). This results in a culture where empirical rigour is lacking, and in-depth understanding of algorithms is neglected and undervalued (Sculley *et al.*, 2018). In a widely discussed speech, Rahimi (Rahimi & Recht, 2017) compared machine learning to alchemy. He argues that while scientific understanding is not always the goal, it is vital for algorithms that have important real-world consequences. In fields such as healthcare or targeted online news, it is essential that algorithms be

[5] In addition, the incidence of CD in the data presented here is obviously vastly inflated relative to the incidence in the population. The classification models presented in this thesis are not intended to be a practical diagnostic tool.

based on rigorous and verifiable empirical knowledge (Rahimi & Recht, 2017). Voicing similar concerns, O'Neil (2016) discusses the serious ethical implications of allowing 'black box' algorithms to determine important outcomes for individuals. One such example is the disproportionate allocation of policing resources to potentially high-crime neighbourhoods in the USA, using algorithms that inadvertently placed a heavy weight on the ethnicity of the residents. Such covert (and illegal) discrimination was not questioned, due to an uncritical acceptance of the technology combined with a lack of transparency and understanding about its workings (O'Neil, 2016).

Given these ethical concerns, it is important to stress that the machine learning models presented in this thesis are intended to complement, rather than replace, traditional statistical techniques. The rationale is that group level characteristics should be reliable at the individual level if they are practically meaningful. This is not equivalent to a clinical diagnostic test. Indeed, while the diagnosis of CD is effectively defined as the fulfilment of a set of behavioural criteria, there is little sense in developing expensive and time-consuming alternatives that approximate the results of the diagnostic checklist. It is thus hoped that the use of machine learning will contribute to the neurobiological and psychological understanding of CD, rather than being taken as the basis for a diagnostic test.

2.3 MRI

2.3.1 Overview

MRI utilises the principles of electromagnetism to produce detailed anatomical images of the human body. Essentially, the electromagnetic 'signals' from hydrogen nuclei (protons) in the body tissue are brought into alignment with a powerful static magnetic

field (B_0), disturbed by a radiofrequency (RF) pulse (B_1), and then allowed to relax to their original alignment. This produces a detectable electromagnetic signal that contains information about the tissue type and its location in the body. In this chapter, the basic physics of structural MRI are explained, drawing on the work of Elster (2018).

2.3.2 Spin

Protons (and all other atomic and subatomic particles) possess a quantum mechanical property known as spin angular momentum, or spin. Spin is conceptually similar to the angular momentum of a rotating object in classical physics, although unlike angular momentum, it is a fundamental natural property. A 'spinning' particle can thus be visualised as a spinning top or gyroscope, rotating around the axis of its associated magnetic moment (see **Figure 7**). In nuclei with even numbers of both protons and neutrons, the spins 'cancel out', resulting in zero net spin. These nuclei cannot be detected with MRI. However, nuclei with odd numbers of protons always have a non-zero net spin.[6] All clinical MRI is based on the hydrogen nucleus, which consists of a single proton, has a spin value of ½ and is abundant in all body tissues due to its presence in water and organic compounds.

Protons can exist in two energy states: with their magnetic dipoles aligned parallel to B_0 (a low energy state) or antiparallel to B_0 (a high energy state).[7] The low energy state is more commonly observed, and the effect of this alignment creates a net magnetisation,

[6] If the number of neutrons is also odd, then the nucleus has an integer spin.
[7] In fact, protons exist in a combination of both states simultaneously, but are only ever observed in one state; the probability of being observed in the higher energy state is smaller than the probability of being observed in the lower energy state.

M_z, in the direction of B_0. M_z forms the basis of the T1 signal, as described in more detail below.

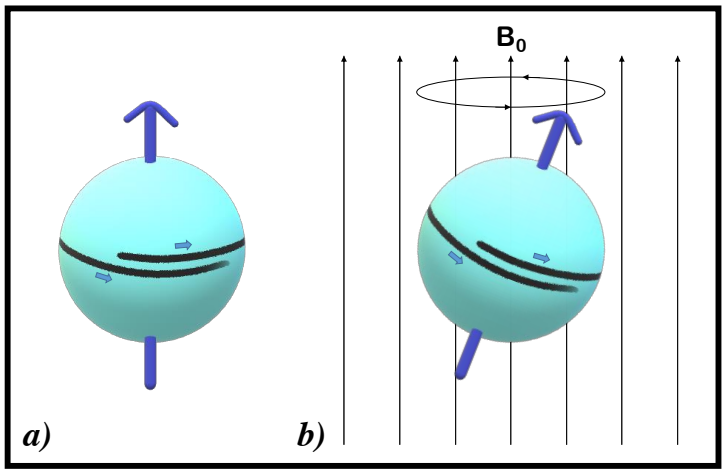


Figure 7. ***a)*** A proton spins around the axis of its magnetic dipole. The spin points in any direction when not exposed to a strong external magnetic field. ***b)*** A proton aligns with an external magnetic field, B_0, and precesses around the direction of B_0

2.3.3 Precession

Protons also precess ('wobble') around the axis of B_0. This results from a torque (twisting force) that is perpendicular to the direction of spin and B_0. In the spinning top or gyroscope analogy, this can be visualised as the spinning top 'tipping' off-centre as it spins, so that the point at the top of the gyroscope does not remain stationary but rather moves in a circular motion as viewed from above (see **Figure 7** ***b)***).

Precession frequency varies by particle type. The precession frequency for a given particle can be calculated using the Larmor equation:

$$\omega_0 = \gamma B_0$$

Where ω_0 is the angular precession (or Larmor) frequency, γ is a particle-specific constant known as the gyromagnetic ratio,[8] and B_0 is the external magnetic field strength. The precession frequency of a proton is thus proportional to the strength of the external magnetic field and the gyromagnetic ratio. Importantly, the precession frequency is influenced by both B_0 and the local magnetic field of the proton, which in turn differs according to the molecule in which the proton is located. This allows sensitivity to different tissue types in MRI.

Although protons precess around the axis of B_0, their precession is not in phase, *i.e.*, the axes of their magnetic dipoles do not all share the same orientation at the same time. This means that although each individual proton is precessing, M_z is not. The precession phase of the protons forms the basis of the T2 signal, as described below.

2.3.4 The Radiofrequency (RF) Pulse

During scanning, the rotating RF pulse, B_1, is applied perpendicular to B_0 and oscillating at or close to the protons' Larmor frequency (ω_0). Energy from the RF pulse is thus transmitted to the protons. This has two effects: first, the individual protons begin to precess in phase with each other, creating a net precession of M. Second, some protons move from the low (parallel) to the high (antiparallel) energy states. In consequence, the longitudinal net magnetisation is lost, and a new transverse net magnetisation (M_{xy}) appears due to the in-phase precession. (Alternatively, this is sometimes described as M rotating from the longitudinal to the transverse plane). M_{xy} 'sweeps' back and forth in the transverse plane due to the net precession, and this phenomenon is known as resonance (see **Figure 8**). The resonance signal is detected by

[8] The gyromagnetic ratio $\gamma = \mu I$, where μ is the magnetic moment and I is the spin of the nucleus.

receiver coils mounted (in the case of brain imaging) around the head. Once the RF pulse is turned off, the protons gradually lose phase with each other and simultaneously return from the high to the low energy states, and the resonance signal is lost.

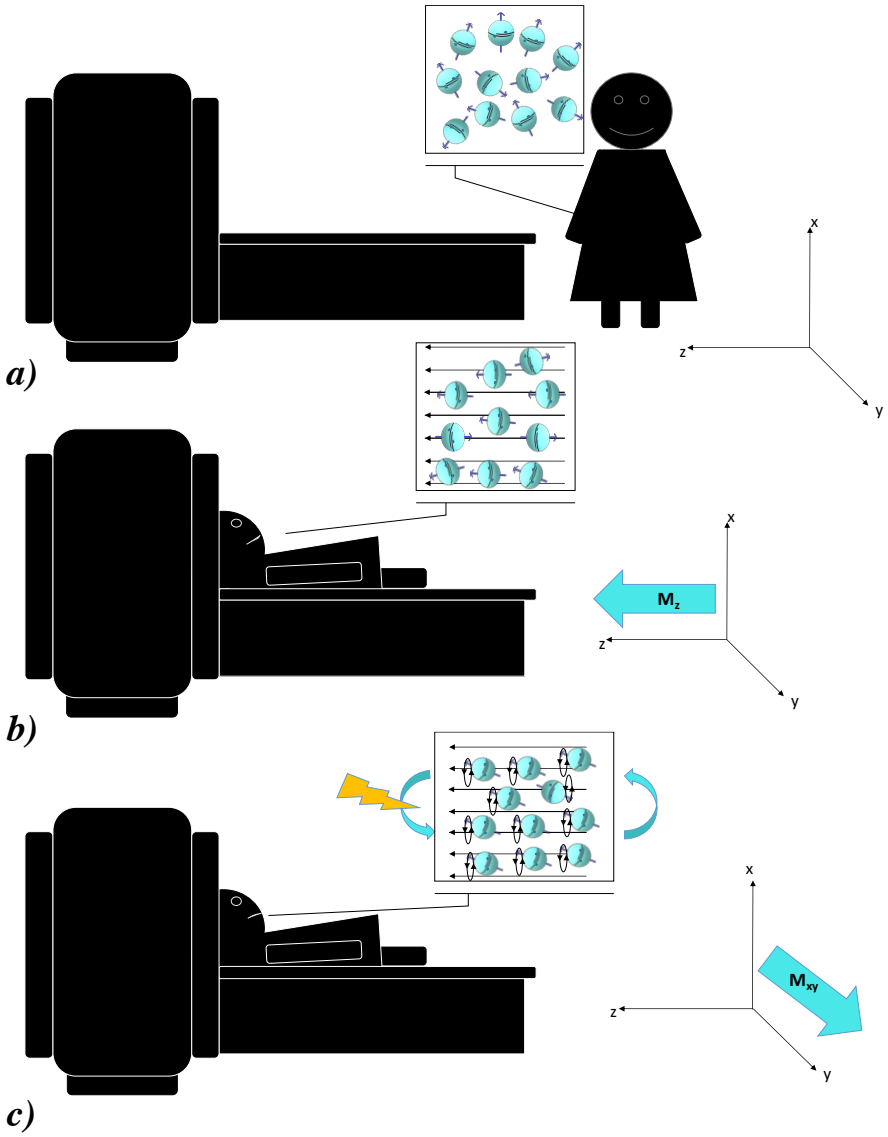


Figure 8. The external magnetic field and net magnetisation. ***a)*** Protons are aligned randomly in the absence of an external magnetic field. ***b)*** Inside the scanner, protons align with B_0 and precess out of phase, creating a longitudinal net magnetisation M_z. ***c)*** The RF pulse, B_1, is applied, causing more protons to transition to the higher energy state (thus eliminating M_z) and precess in phase (thus establishing M_{xy}, the resonance signal)

2.3.5 T1 and T2 Relaxation

As the protons return from the high to the low energy states, M_z regrows longitudinally in the direction of B_0. M_z does not re-appear instantly; instead, it grows exponentially with a time constant referred to as T1. The time constant T1 represents the time taken for M to reach 63% of its initial maximum value. Because protons lose thermal energy to surrounding molecules (the 'lattice') during the transition to the low energy state, T1 relaxation time is also known as spin-lattice relaxation time or thermal relaxation time. Longitudinal relaxation time is another synonym.

As the protons lose phase, M_{xy} decays exponentially with a time constant T2.[9] T2 is the time taken for M_{xy} to reduce to 37% of its initial maximum value. T2 is also known as transverse relaxation time, because the signal that decays is transverse to B_0. T2 relaxation can occur in the absence of T1 relaxation, for example because spins lose phase due to the presence of magnetic field inhomogeneities. However, T1 relaxation is always accompanied by T2 relaxation. This is because the change of energy state of a proton will naturally affect its spin orientation, resulting in a change to both M_z and M_{xy}.

2.3.6 Gradient Fields

Differences in T1 and T2 relaxation times enable different tissue types to be distinguished. However, they do not provide information about the location of the tissue in the body. This is achieved using three gradient coils (for the x, y and z directions).

[9] In practice, M_{xy} decays faster than predicted based on atomic/molecular mechanisms. The actual observed T2 is denoted as T2*.

The gradient coils slightly skew B_0 in a predictable way, meaning that the local Larmor frequencies are altered, and the location of the signal can be established.

2.3.7 Repetition Time (TR) and Echo Time (TE)

A basic spin-echo MRI sequence consists of two RF pulses, typically at a 90° flip angle followed by a 180° flip angle (*i.e.*, rotating M first by 90° and then by 180°). The second pulse re-phases some of the components of the decaying signal from the first pulse, increasing the transverse magnetisation M_{xy} (resonance signal) again. Tissues that have recovered more of their longitudinal magnetisation M_z by the second RF pulse take longer to flip their magnetisation back into the transverse plane. This creates a symmetrical 'echo' signal in the transverse plane that first increases and then decreases as it again begins to decay. The echo time (TE) is the time from the middle of the first RF pulse to the middle (peak) of the echo signal. The repetition time (TR) is the time between two corresponding RF pulses (or between any two corresponding points in the full cycle). The length of the TR determines how much M_z recovers between RF pulses.

2.3.8 T1-Weighted Images

T1-weighted images are normally used in structural MRI. T1-weighted images capture the longitudinal signal M_z as it recovers after the RF pulse. At any given time point before complete recovery of M_z, tissues with a shorter T1 relaxation time will therefore emit more signal and appear brighter than tissues with a longer T1 relaxation time. Fatty substances (*e.g.*, white matter) appear bright due to a short T1, while watery substances (*e.g.*, cerebrospinal fluid) appear dark due to a longer T1. Grey matter has a T1 lower than white matter but higher than cerebrospinal fluid (the opposite order is observed for

T2 relaxation times). A T1-weighted image has a short TR and TE, meaning that M_z does not fully recover between RF pulses, and tissue contrast is clear in the resulting image. These differences are illustrated in **Figure 9**.

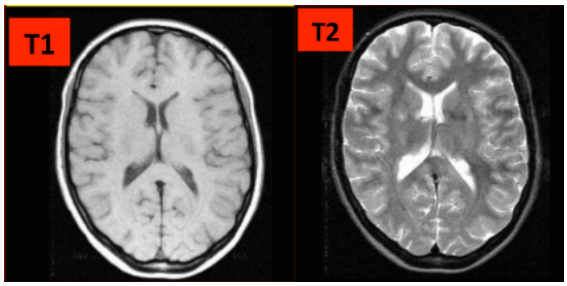


Figure 9. T1-weighted and T2-weighted images. Image reproduced from Elster (2018)

2.3.9 Voxel-Based Morphometry and Image Pre-Processing

Voxel-based morphometry (VBM) is a technique for studying the size and shape of the brain and its different regions, particularly differences in grey matter volume and concentration (Good, Johnsrude, Ashburner, Henson, Friston, & Frackowiak, 2001; Mechelli, Price, Friston, & Ashburner, 2005; Whitwell, 2009; Hutton, Draganski, Ashburner, & Weiskopf, 2009). Image pre-processing consists of several stages. First, the T1-weighted images for each individual are spatially normalised to a group template. This transforms all the individual images into the same space, so that the value (*i.e.*, signal intensity) at a given voxel (3D pixel) can be directly compared across images. Since each brain will differ in shape and size, this process necessitates stretching and compressing of the images to allow a one-to-one correspondence between voxels across individuals. A deformation field is thus created, mapping these changes in size and position from each 'raw' image to its spatially normalised version. Following spatial normalisation, each image is segmented into separate images

(segments) for grey matter, white matter and cerebrospinal fluid. These segments are constructed using the signal intensity at each voxel, combined with tissue probability maps containing prior information about the likelihood of each voxel containing each tissue type. For paediatric samples, customised sample-specific tissue probability maps should be used, which account for age-related differences in brain tissue composition compared to 'standard' adult populations; details of this procedure are provided in Chapter 4, section 4.2.3. The value at each voxel of the segmented images indicates the likelihood of that voxel belonging to the specified tissue type. Following segmentation, the grey matter segments are smoothed using a Gaussian kernel. In this process, the value at each voxel is converted to a weighted average of its neighbouring voxels, thus reducing noise. Finally, smoothed images are registered into a standard space prior to second-level analyses. This standard procedure yields voxel values that are usually interpreted as the concentration of grey matter at that voxel (Good *et al*., 2001; Mechelli *et al*., 2005). As an additional step, the spatially normalised images can be modulated (*i.e*., their values are multiplied by the deformation field) to account for the amount of stretching or compressing at each voxel during normalisation. This step preserves information about the amount of grey matter (or signal intensity) in each voxel prior to spatial normalisation. The resulting modulated images are usually interpreted as representing the volume of grey matter at each voxel (Good *et al*., 2001; Mechelli *et al*., 2005; Eriksson, Free, Thom, Symms, Martinian, Duncan, & Sisodiya, 2009; Whitwell, 2009). In Chapter 4 we base our analyses on modulated images, and interpret this as grey matter volume in line with the standard terminology described here. However, it should be noted that the signal in VBM is a qualitative measure without units, and 'volume' is an interpretation of what the signal represents, rather than a strictly accurate

description of the signal itself. For this reason, the terminology is not completely consistent across fields. For example, in medical research, the term 'grey matter volume' is sometimes used to refer to macroscopic changes in size and shape of brain structures, while the MRI signal used in VBM is referred to as 'grey matter probability' (*e.g.*, Lockwood-Estrin *et al.*, 2012). For clarity and consistency with previous research in the field, we use only the former terminology in this thesis.

2.4 The FemNAT-CD Project

2.4.1 Overview and Aims

All data presented in this thesis were collected as part of the European Commission's 7th Framework Health program (FP7) project, 'The Neurobiology and Treatment of Adolescent Female Conduct Disorder' (FemNAT-CD; Freitag *et al.*, 2018; see www.femnat-cd.eu). The FemNAT-CD project is a collaboration between 17 partner sites across eight European countries (Germany, Greece, Hungary, Ireland, the Netherlands, Spain, Switzerland and the United Kingdom). The overarching aim of the consortium is to elucidate the neurobiological, neurocognitive and environmental mechanisms underlying CD, with a special focus on females, given their sparsity in most CD samples (Freitag *et al.*, 2018). Recruitment took place between January 2014 and February 2018. In total, 1827 youths participated, including 880 with CD (61% females) and 947 TD youths (65% females). Data collection included questionnaires, behavioural, neurophysiological and genetic measures, and structural and functional MRI data. This chapter provides an overview of recruitment and testing procedures, with a focus on the key measures that are used throughout this thesis.

2.4.2 Recruitment

Participants were recruited at 12 sites across Europe. Sources included local clinics, mental health services, youth offending services, mainstream and special schools, youth clubs, other community services, outreach events and word of mouth.

2.4.3 Eligibility and Exclusion Criteria

Male and female youths aged 9-18 years were eligible to participate if they had an IQ of 70 or above, and no current or past indication of autism, schizophrenia, genetic syndromes, neurological disorders or traumatic brain injury. In addition, TD participants (those without CD or ODD) were required to have no history of externalising disorders, bipolar disorder or mania, and no current psychiatric disorders of any kind. Participants were accepted as 'cases' if they *a)* met the DSM-5 diagnostic criteria for CD, *b)* were aged 9-12 years and met the diagnostic criteria for ODD with at least one symptom of CD also present, or *c)* were over 12 years and met the criteria for ODD with at least two symptoms of CD present[10]. Siblings of previous participants were also excluded. All participants were requested to bring a parent, guardian or other responsible adult to the first session as an additional informant.

2.4.4 Informed Consent

A parent, guardian or other responsible adult gave written informed consent on behalf of minors (*i.e.*, those below 16 or 18 years depending on site of data collection). Minors gave written assent. Parents and guardians, and participants above the required age,

[10] The FemNAT-CD criteria were relaxed early during recruitment to include youths with ODD and some CD symptoms, due to the difficulty of recruiting younger participants with CD. Youths with ODD (n = 100) are included in the CD group, in line with FemNAT-CD standard procedures.

gave written informed consent for their own participation. All data were collected within six months of the date of informed consent.

2.4.5 Data Collection Procedure

Referrals to the project were initially screened by telephone to assess suitability. Participants judged likely to be eligible, and their parents or guardians, were then invited to a more comprehensive screening session, either at the local site (*i.e.*, university or associated clinic), or at the participant's home. The first session generally lasted 1-3 hours. Informed consent was provided at the beginning of the session. Participants and their parents/guardians were interviewed separately by trained researchers to assess for the presence of any mental disorders. An IQ test was also administered. Eligible participants then returned for a second session of approximately two hours, in which they completed questionnaire measures, provided saliva samples for genotyping, and completed a set of neurophysiological and neurocognitive assessments. Data on parenting and facial emotion recognition were usually collected during this session. At participating sites, MRI data were collected during a third testing session of approximately two hours. Participants received a small financial reimbursement, or equivalent vouchers, after each session, as approved by the local ethics committees[11]. Measures that are reported

[11] Aachen: Ethik Kommission Medizinische Fakultät der Rheinisch Westfälischen Technischen Hochschule Aachen (EK027/14). Amsterdam: Medisch Etische Toetsingscommissie (2014.188). Athens: Election Committee of the First Department of Psychiatry, Eginition University Hospital (641/9.11.2015). Barcelona: Child and Adolescent Mental Health - University Hospital Mutua Terrassa (acta 12/13). Basel: Ethik Kommission Nordwest- und Zentralschweiz (EKNZ 336/13). Bilbao: Hospital del Basurto. Birmingham and Southampton: University Ethics Committee and National Health Service Research Ethics Committee (NRES Committee West Midlands, Edgbaston; REC reference 3/WM/0483). Dublin: SJH/AMNCH Research Ethics Committee (2014/04/Chairman (3)). Frankfurt: Ethik Kommission Medizinische Fakultät Goethe Universität Frankfurt am Main (445/13). Szeged (Hungary): Egészségügyi Tudományos Tanács Humán Reprodukciós Bizottság (CSR/039/00392-3/2014).

throughout this thesis are described in detail below. Parenting, emotion recognition and MRI data collection is described separately in the relevant experimental chapters.

2.4.6 Key Measures

2.4.6.1 Schedule for Affective Disorders and Schizophrenia for School-Age Children: Present and Lifetime Version (K-SADS-PL)

The K-SADS-PL (Kaufman *et al*., 1997) is a semi-structured diagnostic interview used to assess current and past psychopathology in children and adolescents. The interview was administered separately to participants and parents (or another responsible adult informant) by trained researchers, and combined parent and child summary ratings of all symptoms (past, present and lifetime) were then generated. Where assessors gave discrepant ratings for a symptom, they discussed all available information until an agreement was reached for the summary rating. Except for CD, ODD, and ADHD, where DSM-5 criteria were used, all diagnoses were generated based on the DSM-IV-TR diagnostic criteria, which were current at the outset of the project (APA, 2000). Inter-rater reliability for current CD diagnoses was high (94.7% agreement across raters, Cohen's kappa=0.91).

2.4.6.2 Wechsler Intelligence Scale (WASI, WAIS, WISC)

In English-speaking sites, IQ was estimated with the vocabulary and matrix reasoning subscales of the WASI-I (Wechsler, 1999). Other sites used the vocabulary, block design and matrix reasoning tests of the WISC (for participants aged 16 or under) or WAIS (for participants aged 17-18 years; Wechsler, 2008).

2.4.6.3 Inventory of Callous-Unemotional Traits (ICU), Parent-Report Version

The ICU (Essau *et al*., 2006b) is a 24-item questionnaire measure of callous-unemotional traits in children and adolescents. There are three subscales (callous, uncaring and unemotional) and a total score. Items assessing callousness include, "Does not seem to know 'right' from 'wrong'", "Is concerned about the feelings of others" (reverse coded) and, "Shows no remorse when he/she has done something wrong". Uncaring items include, "Is concerned about schoolwork" (reverse coded), "Feels bad or guilty when he/she has done something wrong" and "Tries not to hurt others' feelings". Unemotional items include, "Expresses his/her feelings openly" (reverse coded), "Does not show emotions" and, "It is easy to tell how he/she is feeling". Items are rated on a four-point scale from 0 (*'not at all true'*) to 3 (*'definitely true'*).

2.4.6.4 Pubertal Development Scale (PDS)

The PDS (Petersen, Crockett, Richards, Maryse, & Boxer, 1988) is a self-report measure of pubertal development. Items related to growth of body and facial hair, height, change of voice and menstruation are rated on a scale from 1 (*'not yet started'*) to 4 (*'seems complete'*), and the participant is assigned to an overall category (pre-pubertal, early pubertal, mid-pubertal, late pubertal, or post-pubertal). The response option 0 (*'I don't know'*) was removed shortly after data collection began, as its inclusion resulted in excessive amounts of missing data.

2.4.7 Additional Measures

Although not forming part of the main analyses, additional measures are reported in each experimental chapter as a further illustration of sample characteristics. These measures are the Reactive-Proactive Aggression Questionnaire (RPQ; Raine *et al*.,

2006), the Child Behaviour Checklist (CBCL; Achenbach, 1991) and the Griffiths Empathy Measure (GEM; Dadds *et al*., 2008). The RPQ is a 23-item self-report measure of reactive, proactive and total aggression. The CBCL is a 120-item parent-report measure of emotional, social and behavioural problems. Composite subscales (reported in this thesis) measure internalising, externalising and total problems. As CBCL subscale totals are calculated automatically using software supplied with the CBCL, it was not possible to compute reliability measures for the CBCL in Chapters 3-5. The GEM is a 23-item parent-report questionnaire measure of cognitive and affective empathy, with two subscales and a total empathy score.

2.4.8 Imputation of Missing Data

Missing data were imputed by statisticians at the Institute of Medical Biometry and Statistics (IMBI), a member of the FemNAT-CD consortium. Missing data for the PDS were imputed separately, before the decision was made to impute missing values for other measures. The procedure for the PDS imputation is thus described separately from the other measures. The following description is a standard text provided by IMBI, for use in all FemNAT-CD consortium publications.

Missing values of the PDS score were imputed based on the whole FemNAT-CD sample. It has been shown that missing data in a multi-item instrument is best handled by imputation at the item level (Eekhout, de Vet, Twisk, Brand, de Boer, & Heymans, 2014). Thus, missing values of the single items were imputed first, and the scores were calculated based on the imputed items. The imputation was done in SAS® version 9.4 using the procedure PROC MI. Imputation by fully conditional specification (FCS) is used, which offers a flexible method to specify the multivariate imputation model for arbitrary missing patterns

including both categorical and continuous variables (Liu & De, 2015). As the items are measured at an ordinal level, the logistic regression method is specified in the FCS statement. For imputation diagnostics, distribution of the observed and imputed items and scores were checked. The imputation of the PDS items was done separately in males and in females because of sex specific items: item 2 (females and males) and items 4, 5a of the form for females or items 4, 5 of the form for males were imputed respectively. The following variables were included in the imputation model: sex specific items of the PDS as mentioned above and the two remaining PDS items (items 1 and 3), age at PDS and age at informed consent, to impute age at PDS if missing, weight, case/control status, site, and migration status.

Imputation for the remaining measures was conducted separately, following the same procedure as above. The following variables were included in the imputation model: all items of the respective questionnaire, age, IQ, group (case/control), sex (male/female), site, comorbidities (PTSD, ADHD, ODD, depression, anxiety), and items of other questionnaires if correlated with at least one of the items with ≥0.4. For imputation diagnostics, distribution of the observed and imputed items and scores were checked. For the parent-report Alabama Parenting Questionnaire (APQ; Essau, Sasagawa & Frick, 2006a) missing items were only imputed if at least one item was present.

2.4.9 Threshold for Dichotomisation into CD/HCU and CD/LCU

As discussed in Chapter 1 (section 1.3.2), there is no standard method for defining the cut-off point(s) between CD/HCU and CD/LCU. For pragmatic reasons, we alternate between a tertile split of the total ICU score for the CD group (Chapters 3 and 5) and a median split (Chapter 4). Where sample sizes allow, the tertile split ensures a good separation of the CD/HCU and CD/LCU groups. The median split results in groups that

are more similar to each other, but preserves the overall sample size. As previously discussed, there is evidence that psychopathic traits are dimensional rather than taxonic (Edens *et al*., 2006; Clark, 2007; Murrie *et al*., 2007). Thus, within reasonable bounds, differences in the cut-off criterion are unlikely to have drastic effects on results. A previous paper took a similar approach (Bergstrøm & Farrington, 2018), and there is evidence that differences in dichotomisation approach have minimal effect on results (Farrington & Loeber, 2000; Frogner, Andershed, & Andershed, 2018). Second, the prevalence of CD/HCU within CD samples has been estimated at 9% (Christian *et al*., 1997; based on a cluster analysis) up to 46% in the community (Rowe *et al*., 2010; based on criteria designed to approximate the LPE specifier), or 10% to 50% across clinic-referred and community samples (Kahn *et al*., 2012; also based on the LPE criteria). In this thesis, the percentages of the CD group with CD/HCU are 32% (Chapters 3 and 5, tertile split, threshold ICU score =40) and 52% (Chapter 4, median split, threshold ICU score = 34). This puts the CD/HCU prevalence in our sample within, or just slightly above, the range of previous estimates. Finally, despite alternating between tertile and median splits, there was considerable overlap of the samples in Chapters 3-5. Of all the participants who were included in any analyses in this thesis, 58% were included in more than one chapter. In Chapter 4, where a median split was used, 71% of the sample were included in at least one other chapter.

CHAPTER 3: POSITIVE AND NEGATIVE PARENTING IN CONDUCT DISORDER WITH HIGH VERSUS LOW LEVELS OF CALLOUS-UNEMOTIONAL TRAITS

3.1 Introduction

Researchers have posited different developmental pathways to CD/HCU and CD/LCU, with CD/HCU having a stronger genetic component and CD/LCU a stronger environmental component (*e.g.*, Viding *et al.*, 2005). Nonetheless, it is likely that parenting plays a role in the development of both subtypes (Waller *et al.*, 2013). We investigated whether parenting is differentially associated with CD/HCU and CD/LCU, with a particular focus on positive versus negative parenting.

Positive parenting includes involvement with the child, emotional warmth and positive reinforcement. These characteristics are thought to promote empathy development and moral conscience (Kochanska, Forman, Aksan, & Dunbar, 2005). Adoption studies provide compelling evidence that positive parenting reduces CU-type behaviour in high-risk toddlers (Hyde *et al.*, 2016; Waller *et al.*, 2016; Waller *et al.*, 2017). Likewise, a positive parenting intervention has demonstrated some success with at-risk pre-schoolers, promoting techniques such as effective discipline, positive reinforcement, and promotion of good behaviour through storytelling (Elizur & Somech, 2018). Positive parenting has also been associated with a reduction in CU traits in older children (Pardini, Lochman, & Powell, 2007; Muratori *et al.*, 2016). More recently, a genetically-informed, population-representative study demonstrated that while parental harshness was associated with both aggression and CU traits, low parental warmth was uniquely associated with CU traits (Waller, Hyde, Klump, & Burt, 2018; but see also

Flom, White, Ganiban, & Saudino, 2019). Evidently, there is a relationship between low levels of positive parenting and CU traits, especially in young children with subclinical conduct problems.

While not as strong as the evidence for a direct connection between positive parenting and CU traits, there is also evidence that positive parenting is associated with fewer conduct problems, at least at lower levels of CU traits. For example, in children aged four to 12 years, Pasalich *et al*. (2011) found that maternal warmth was associated with fewer behavioural problems in both CD/HCU and CD/LCU. Indeed, the association was stronger in CD/HCU. Conversely, Falk and Lee (2012) reported an association between positive parenting and CD severity only at low levels of CU traits, and no associations with oppositional-defiant behaviour. Finally, Ray *et al*. (2017) demonstrated that an association between CU traits and delinquent peer association, which led to offending, was weaker in adolescents with high levels of parental warmth and supervision.

Negative parenting includes harsh and inconsistent discipline, and is a common feature of the parenting environment in both CD/HCU and CD/LCU (*e.g*., Fontaine, McCrory, Boivin, Moffitt, & Viding, 2011; but *c.f*. Enebrink, Andershed, & Långström, 2005). Despite its prevalence, youths with CD/HCU are hypothesized to be insensitive to negative parenting because they are insensitive to punishment more generally (Dadds & Salmon, 2003; Blair *et al*., 2001). There is some evidence that negative parenting is less strongly associated with conduct problems in youths with CD/HCU than in those with CD/LCU (Wootton *et al*., 1997). Wootton and colleagues' findings have been partially replicated (Oxford, Cavell, & Hughes, 2003; Pasalich *et al*., 2011) but others have reported an association between negative parenting and CD severity in both CD/HCU and CD/LCU (Pardini *et al*., 2007). One study reported a stronger association between

negative parenting and severity at high levels of CU traits (Crum *et al*., 2015). By contrast, others have found no association between negative parenting and CD regardless of CU trait levels (Falk & Lee, 2012). The evidence for a direct effect of negative parenting on CU traits is also mixed (Pardini *et al*., 2007; Muratori *et al*., 2016; Waller *et al*., 2018).

In summary, low positive parenting is more strongly associated with CU traits while negative parenting is more strongly associated with conduct problems. However, it is less clear whether CU traits and parenting interact, such that negative parenting in particular is less predictive of CD/HCU than of CD/LCU. We addressed two questions regarding the associations between parenting, CD/HCU and CD/LCU. First, do youths with CD/HCU, youths with CD/LCU and typically developing (TD) youths differ, on average, in their exposure to positive and negative parenting practices? Second, how informative are differences in positive and negative parenting in determining the diagnosis of individual youths (*i.e*., CD/HCU, CD/LCU or TD)? Angle-GMLVQ (Bunte *et al*., 2016) is sensitive to relative differences between features (*i.e*., scores on different types of parenting behaviour) rather than absolute magnitude of feature scores. It should thus perform well when individuals are characterised by different patterns of parenting, while minimising the effect of subjective, idiosyncratic tendencies to give uniformly high or low ratings across items.

Given previous findings (*e.g*., Waller *et al*., 2018), we hypothesised that youths with CD/HCU would be characterised by low levels of positive parenting as well as high levels of negative parenting, while youths with CD/LCU would be characterised mainly by high levels of negative parenting. Similarly, we predicted that parenting behaviours would distinguish both youths with CD/HCU and youths with CD/LCU from TD youths

at above-chance levels in Angle-GMLVQ analyses. Next, we predicted that if youths with CD/HCU do indeed experience lower levels of positive parenting as well as (at least) similar levels of negative parenting compared to those with CD/LCU, then these groups too would be distinguished at above-chance levels in classification analyses. As a further test of the same hypothesis, we constructed a CD-against-TD classifier (without distinguishing between CD/HCU and CD/LCU, *i.e.*, Mixed-TD model) and compared its performance with the CD/HCU-against-TD ('HCU-TD') and CD/LCU-against-TD ('LCU-TD') classifiers. If youths with CD/HCU and youths with CD/LCU experience qualitatively distinct patterns of parenting as outlined above, then both of these classifiers should outperform the Mixed-TD model. Finally, in line with expected group differences, we predicted that both positive and negative parenting behaviours would be relevant for the HCU-TD model, negative parenting would be more relevant for the LCU-TD model, and positive parenting would be more relevant for the HCU-LCU model.

3.2 Methods

3.2.1 Participants

Participants were selected from the FemNAT-CD sample in November 2017 (Freitag *et al.*, 2018). Eligibility criteria, recruitment and testing procedures are described in Chapter 2 (see also Kersten *et al.*, 2017). The full FemNAT-CD sample consisted of 1743 participants in November 2017. Of these, one was excluded for not meeting FemNAT-CD eligibility criteria, 67 were excluded due to missing data on CU traits, 14 for missing child-report data on parenting, 413 for missing parent-report data on parenting, and 37 because they did not live with a parent or guardian. Participants with

CD and scores in the 2nd tertile (ICU scores of 30-40; n = 166) were excluded. TD participants with scores in the 1st tertile (n = 6) or 2nd tertile (n = 36) were also excluded, on the grounds that youths with elevated CU traits are unlikely to be truly TD, even in the absence of a CD diagnosis (Rowe *et al*., *2010*). Next, 247 participants (TD n = 226, CD/LCU n = 21) were excluded so that groups were matched for site of data collection, number of males and females, and mean age and pubertal status (Match software; van Casteren & Davis, 2007). This left a final sample of 756 participants (females: n = 436; CD/HCU: n = 164, CD/LCU: n = 164, TD: n = 428). Finally, a 'mixed' CD group (CD/mixed) was formed by combining the CD/HCU and CD/LCU groups (n = 328).

The final sample differed significantly from the excluded participants on age and IQ; excluded participants were older ($t_{(1741)} = -5.64$, $p < .001$, 2-tailed, partial $\eta^2 = .02$) and had lower total IQ scores ($t_{(1682)} = 4.66$, $p < .001$, 2-tailed, partial $\eta^2 = .01$). There was also a greater proportion of females in the excluded participants, reflecting the deliberate oversampling of females in FemNAT-CD ($\chi^2 = 14.14$, $P < .001$, $\varphi = .09$). Comparing only youths with CD, the final sample did not differ significantly from the excluded participants on CU scores or CD symptoms (CU scores: $t_{(566.63)} = 1.65$, $p = .10$, partial $\eta^2 = .004$. Current CD symptoms: $t_{(745.23)} = -1.22$, $p = .22$, partial $\eta^2 = .002$. Lifetime CD symptoms: $t_{(741.18)} = -1.71$, $p = .09$, partial $\eta^2 = .003$).

3.2.2 Questionnaire and Interview Measures

The K-SADS-PL (Kaufman *et al*., 1997) was used to assess for CD and other disorders. The ICU was used as the measure of CU traits; reliability was good in the current sample (callous $\alpha = 0.88$, uncaring $\alpha = 0.88$, unemotional $\alpha = 0.78$, total $\alpha = 0.93$). The PDS (Petersen *et al*., 1988) was used to measure pubertal development, and the

Wechsler Intelligence Scales (Wechsler, 1999; Wechsler, 2008) were used to estimate IQ. As a further illustration of sample characteristics, we also report scores from the Reactive-Proactive Aggression Questionnaire (RPQ; Raine *et al*., 2006), Griffiths Empathy Measure (GEM; Dadds *et al*., 2008) and the Child Behaviour Checklist internalising, externalising and total problems scales (CBCL; Achenbach, 1991). RPQ and GEM subscale reliabilities were good; RPQ proactive $\alpha = 0.86$, RPQ reactive $\alpha = 0.88$, RPQ total $\alpha = 0.91$, GEM affective $\alpha = 0.82$, GEM cognitive $\alpha = 0.73$, GEM total $\alpha = 0.87$). All of these measures are described in Chapter 2.

Parenting was assessed with the child-report and parent-report versions of the Alabama Parenting Questionnaire (APQ; Essau *et al*., 2006a). The APQ is a 42-item measure of parenting. There are separate subscales for maternal and paternal involvement, positive parenting, poor supervision, inconsistent discipline, and corporal punishment. In order to avoid confusion with the broader concept of positive parenting, the APQ positive parenting subscale is referred to as positive reinforcement hereafter. Example items include, "You have a friendly talk with your mom" (maternal involvement), and "Your parents reward or give something extra to you for behaving well" (positive reinforcement). For the negatively-worded subscales, example items include, "You stay out in the evening past the time you are supposed to be home" (poor supervision), "Your parents threaten to punish you and then do not do it" (inconsistent discipline), and, "Your parents slap you when you have done something wrong" (corporal punishment).

Items are rated on a five-point scale ranging from 1 (*'never'*) to 5 (*'always'*). The parent-report APQ consists of the same subscales as the child-report version, except that maternal and paternal involvement are replaced by a single parental involvement

subscale. For the current analyses, the more negative of the child and parent ratings for each item was taken as the 'summary' item score, *i.e.*, the lower score on the positive parenting items and the higher score on the poor supervision and negative parenting items. For the parental involvement items, the higher score from the child-report maternal involvement and corresponding paternal involvement items was first taken, as this was assumed to reflect the involvement of the primary carer. The lower score from this and the parent-rated parental involvement item was then taken as the summary score. In line with previous studies, parental involvement and positive reinforcement subscales were used as measures of positive parenting. Inconsistent discipline and corporal punishment were used as measures of negative parenting, and poor supervision was treated as a distinct component (Muratori *et al*., 2016; Molinuevo, Pardo, & Torrubia, 2011). Reliability was good for all subscales (Cronbach's alphas: involvement $\alpha = 0.78$, positive reinforcement $\alpha = 0.80$, poor supervision $\alpha = 0.82$, inconsistent discipline $\alpha = 0.65$, corporal punishment $\alpha = 0.77$).

3.2.3 Analysis

3.2.3.1 Univariate Analyses

Group differences on parenting measures were assessed with one-way ANOVAs, after regressing out variance associated with IQ, sex, pubertal status and site of data collection. We also report results after additionally regressing out variance associated with family structure. Differences on other measures (*e.g.*, CD symptoms, age) were assessed with one-way ANOVAs and chi square tests as appropriate. We additionally explored sex differences in the relationship between parenting and group status (*i.e.*, CD/HCU, CD/LCU and TD) using two (sex) by three (group) way ANOVAs for each

measure of parenting, after regressing out variance associated with IQ, pubertal status and site of data collection.

3.2.3.2 Classification Models

Angle-GMLVQ was used for classification analyses. Since the aim was to generate feature relevance scores that distinguished between specific groups, models were created for each pair of groups of interest:

1. CD/mixed against TD (referred to hereafter as 'Mixed-TD')
2. CD/HCU against TD ('HCU-TD')
3. CD/LCU against TD ('LCU-TD')
4. CD/HCU against CD/LCU ('HCU-LCU')

3.2.3.3 Training and Testing Procedure

The classifier was trained and tested for each model, with one prototype per class. Features were parenting scores on each of the five APQ subscales, after regressing out variance associated with IQ, sex, pubertal status and site of data collection (information on family structure was not available for the full sample, and thus was not controlled for in these analyses). Performance was assessed using a holdout design with an 80/20 training/testing split, repeated for 1000 random sub-samplings in order to ensure stability of the model. Where classes were initially balanced, 80% of each class was selected at random in each re-sampling. Where classes were imbalanced, the larger class was instead randomly down-sampled to the size of the smaller class in each re-sampling (see Japkowicz, 2000). Mean performance metrics across all re-samplings were then compared between models.

3.2.3.4 Assessment of Model Performance

In each model, the CD group was defined as the positive class and TD as the negative class. In the HCU-LCU model, the CD/HCU group was the positive class. Overall model performance was judged by MCER and classification accuracy. Confidence in positive and negative classifications was assessed with PPV and NPV respectively. Finally, the ability of the classifier to detect members of the positive and negative classes was assessed using TPR (*i.e.*, sensitivity) and TNR (specificity) respectively. These measures are described fully in Chapter 2 (section 2.2.5).

3.2.3.5 Assessment of Feature Relevance

Relevance scores were considered 'high' if they were in the top 20% of scores across all re-samplings with a corresponding MCER of 0.40 or below (*i.e.*, the most poorly performing models were excluded from the relevance analysis). Features were then ranked by number of high scores. Relevance scores were normalized for each re-sampling, so that direct comparisons could be made across re-samplings. Finally, to verify that feature relevance was not dependent on the chosen threshold for 'high' scores (*i.e.*, top 20%), relevance was compared at two additional thresholds; top 15% (*i.e.*, a stricter threshold) and top 25% (a more lenient threshold).

3.2.3.6 SVM Classifiers

Analyses were repeated with SVM classifiers to ensure that classifier performance was broadly similar across different methods. SVM classifiers were trained and tested using the standard MATLAB function 'fitcsvm'. We trained and tested six classifiers for each model: linear SVM and SVMs with second, third, fourth, fifth and sixth-order non-linear polynomial transformation kernels. We then selected the SVM with the lowest MCER for comparison with the corresponding Angle-GMLVQ classifier.

3.3 Results

3.3.1 Sample Characteristics

In line with the group matching process, there were no differences between CD/HCU, CD/LCU and TD groups in the proportion of participants included from each site (χ^2 = 24.68, p =.21, φ = .21), nor in the proportion of females, mean age or pubertal status (see **Table 1**). The CD/HCU and CD/LCU groups did not differ from each other on performance, verbal or total IQ scores, but both had significantly lower IQ scores than the TD group (**Table 1**). As expected, the CD/HCU group had significantly more CD symptoms, aggression and externalising problems than the CD/LCU group (see **Table 1** and **Table 2**). Information on family composition is displayed in **Table 3**. Although youths living independently were excluded, information on caregivers living in the household was only available for a subset of the full sample (information available for female carer: CD/HCU: 119, CD/LCU: 127, TD: 414. Male carer: CD/HCU: 91, CD/LCU: 91, TD: 336).

Table 1. Demographic and clinical characteristics (mean (95% confidence intervals of the mean) unless stated otherwise)

Measures	CD/HCU (n = 164)	CD/LCU (n = 164)	TD (n = 428)	Test statistic (p), effect size
Age	13.84 (13.50, 14.18) [a]	13.97 (13.60, 14.34) [a]	13.73 (13.50, 13.97) [a]	F = 0.59 (*.55*), partial η^2 = .00
Females (%)	54	52	61	χ^2 = 4.54 (*.10*), φ = .08
PDS pubertal stage	3.40 (3.23, 3.57) [a]	3.49 (3.31, 3.66) [a]	3.43 (3.32, 3.53) [a]	F = 0.29 (*.75*), partial η^2 = .00
Performance IQ	98.28 (95.95, 101.62) [a]	97.36 (95.12, 99.59) [a]	103.86 (102.44, 105.27) [b]	F = 15.44 (<*.001*), partial η^2 = .04
Verbal IQ	94.33 (92.02, 96.65) [a]	95.03 (92.60, 97.45) [a]	104.83 (103.36, 106.31) [b]	F = 39.81 (<*.001*), partial η^2 = .10
Total IQ	96.54 (94.63, 98.45) [a]	96.48 (94.52, 98.44) [a]	104.60 (103.39, 105.81) [b]	F = 37.57 (<*.001*), partial η^2 = .09
ICU callous	19.12 (18.41, 19.83) [a]	7.45 (6.92, 7.97) [b]	3.88 (3.65, 4.10) [c]	F = 1336.31 (<*.001*), partial η^2 = .78
ICU uncaring	18.90 (18.48, 19.33) [a]	10.99 (10.37, 11.61) [b]	7.24 (6.89, 7.58) [c]	F = 637.35 (<*.001*), partial η^2 = .63
ICU unemotional	9.86 (9.40, 10.31) [a]	5.49 (5.07, 5.91) [b]	4.59 (4.33, 4.84) [c]	F = 221.74 (<*.001*), partial η^2 = .37
ICU total	47.88 (46.91, 48.86) [a]	23.92 (22.99, 24.85) [b]	15.70 (15.09, 16.31) [c]	F = 1542.61 (<*.001*), partial η^2 = .80
K-SADS CD symptoms	5.65 (5.27, 6.02) [a]	4.76 (4.40, 5.13) [b]	0.06 (0.04, 0.09) [c]	F = 976.67 (<*.001*), partial η^2 = .72
K-SADS ODD symptoms	6.87 (6.55, 7.21) [a]	5.50 (5.08, 5.92) [b]	0.07 (0.03, 0.12) [c]	F = 1332.05 (<*.001*), partial η^2 = .78
K-SADS ADHD symptoms	8.78 (7.69, 9.88) [a]	6.71 (5.70, 7.71) [b]	0.06 (0.02, 0.10) [c]	F = 277.72 (<*.001*), partial η^2 = .43
K-SADS GAD diagnosis (%)	15	20	2	χ^2 = 60.99 (<*.001*), φ = .28
K-SADS MDD diagnosis (%)	25	21	1	χ^2 = 94.66 (<*.001*), φ = .35
K-SADS SUD diagnosis (%)	19	21	0	χ^2 = 92.73(<*.001*), φ = .35

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. K-SADS = Schedule for Affective Disorders and Schizophrenia for School-age Children: Present and Lifetime Version (lifetime maximum symptoms/diagnosis), CD = conduct disorder, ODD = oppositional defiant disorder, ADHD = attention deficit/hyperactivity disorder, GAD = generalised anxiety disorder, MDD = major depressive disorder, SUD = substance use disorder. PDS = Self-rating Scale for Pubertal Development, ICU = Inventory of Callous-Unemotional Traits. Groups with different superscript indices differ significantly in post-hoc comparisons (p < .05, Bonferroni corrected)

Table 2. Additional clinical characteristics (mean (95% confidence intervals of the mean))

Measures	CD/HCU (n = 164)	CD/LCU (n = 164)	TD (n = 428)	Test statistic (p), effect size
RPQ reactive aggression	12.30 (11.56, 13.05) [a]	11.41 (10.70, 12.13) [a]	5.36 (5.01, 5.71) [b]	F = 225.74 (<.001), partial η^2 = .38
RPQ proactive aggression	5.12 (4.45, 5.78) [a]	4.08 (3.38, 4.78) [b]	0.84 (0.70, 0.98) [c]	F = 139.86 (<.001), partial η^2 = .27
RPQ total aggression	17.42 (16.17, 18.67) [a]	15.49 (14.23, 16.76) [b]	6.20 (5.77, 6.63) [c]	F = 244.79 (<.001), partial η^2 = .40
GEM affective empathy	-0.74 (-3.06, 1.59) [a]	6.53 (4.27, 8.78) [b]	7.81 (6.53, 9.10) [b]	F = 21.98 (<.001), partial η^2 = .07
GEM cognitive empathy	-1.21 (-2.77, 0.35) [a]	6.76 (5.05, 8.47) [b]	11.73 (10.93, 12.53) [c]	F = 117.24 (<.001), partial η^2 = .28
GEM total empathy	2.11 (-2.55, 6.77) [a]	28.09 (24.67, 31.51) [b]	36.75 (34.54, 38.97) [c]	F = 114.96 (<.001), partial η^2 = .28
CBCL internalizing	65.04 (63.42, 66.67) [a]	61.67 (59.75, 63.59) [b]	49.76 (48.76, 50.75) [c]	F = 150.50 (<.001), partial η^2 = .31
CBCL externalizing	75.99 (74.93, 77.05) [a]	65.67 (63.78, 67.56) [b]	47.18 (46.26, 48.11) [c]	F = 606.89 (<.001), partial η^2 = .64
CBCL total problems	72.68 (71.17, 74.19) [a]	65.22 (63.26, 67.17) [b]	47.59 (46.58, 48.59) [c]	F = 384.73 (<.001), partial η^2 = .53

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. RPQ = Reactive-Proactive Aggression Questionnaire, GEM = Griffiths Empathy Measure, CBCL = Child Behaviour Checklist. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

Table 3. Caregivers living in family home (% of group)

Caregiver	CD/HCU (*n* = 164)	CD/LCU (*n* = 164)	TD (*n* = 428)	χ^2 (*p*), φ
Biological mother	63	70	96	120.05 (*<.001*), .40
Biological father	32	40	71	57.48 (*<.001*), .33
Adoptive mother	4	2	0	16.93 (*<.001*), .15
Adoptive father	4	1	0	20.78 (*<.001*), .20
Stepmother	3	2	0	9.51 (*.01*), .11
Stepfather	16	12	6	29.73 (*<.001*), .24
Foster mother	2	1	0	7.23 (*.03*), .10
Foster father	2	1	0	9.90 (*.007*), .14
Other female carer	1	1	0	4.83 (*.09*), .08
Other male carer	2	1	0	2.00 (*.37*), .06

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. Participants who were not living with parents or caregivers were excluded, but information about caregivers was not available for the full sample. Significance tests were conducted using only those with data concerning the relevant caregiver

3.3.2 Group Differences in Parenting

Mean group differences for each APQ subscale are displayed in **Table 4**. Raw scores are displayed for ease of interpretation, but we indicate where the pattern of significant differences changed after regressing out variance associated with IQ, sex, pubertal status, site of data collection and family structure. All three groups differed significantly on positive parenting, with the TD group scoring highest and the CD/HCU group scoring lowest for both parental involvement and positive reinforcement. Each group also differed significantly on poor supervision. Interestingly, the CD/LCU and TD groups no longer differed significantly on positive parenting after regressing out variance associated with differences in family structure (although it should be noted that

information regarding family structure was not available for the full sample). The CD/HCU and CD/LCU groups did not differ significantly on negative parenting, although both groups experienced significantly more negative parenting than the TD group. These group differences support the hypothesis that youths with CD/HCU are characterised by high negative and low positive parenting. Contrary to our hypothesis, however, the CD/LCU was also characterised by low positive parenting in addition to the predicted high negative parenting. Notably, only positive parenting differed significantly between the CD/HCU and CD/LCU groups.

Table 4. Group differences in parenting (mean (95% confidence intervals of the mean))

APQ subscales	**CD/HCU (*n* = 164)**	**CD/LCU (*n* = 164)**	**TD (*n* = 428)**	**F (*p*), partial η^2**
Positive parenting				
Parental involvement	24.48 (23.49, 25.47) [a]	27.60 (26.50, 28.70) [b *]	30.89 (30.42, 31.36) [c *]	76.50 (<.001), .17
Positive reinforcement	17.52 (16.78, 18.26) [a]	18.94 (18.16, 19.72) [b *]	20.82 (20.45, 21.19) [c *]	36.63 (<.001), .09
Poor supervision	30.41 (29.26, 31.56) [a]	28.01 (26.74, 29.27) [b]	22.44 (21.82, 23.05) [c]	89.57 (<.001), .19
Negative parenting				
Inconsistent discipline	19.52 (18.94, 20.11) [a]	18.85 (18.24, 19.47) [a]	15.81 (15.49, 16.14) [b]	82.28 (<.001), .18
Corporal punishment	5.42 (5.01, 5.83) [a]	5.18 (4.80, 5.55) [a]	3.96 (3.81, 4.11) [b]	40.19 (<.001), .10

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. APQ = Alabama Parenting Questionnaire. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected). Regressing out variance associated with IQ, sex, pubertal status and site of data collection did not change the pattern of significant group differences

***** These groups no longer differed significantly after regressing out variance associated with family structure in addition to IQ, sex, pubertal status and site of data collection

3.3.3 Sex Differences

Parenting scores per group are reported separately for females (**Table 5**) and males (**Table 6**). Main effects of sex and sex by group interactions were investigated for each parenting measure separately, using two (sex) by three (group) way ANOVAs for each measure of parenting (after regressing out variance associated with IQ, pubertal status and site of data collection). There were no main effects of sex for parental involvement ($F_{(1, 742)} = 2.88, p = .09$), positive reinforcement ($F_{(1, 742)} = 1.77, p = .18$), inconsistent discipline ($F_{(1, 742)} = 3.59, p = .06$), or corporal punishment ($F_{(1, 742)} = 3.21, p = .07$). However, there was a main effect of sex for poor supervision ($F_{(1, 742)} = 4.86, p = .03$), with males experiencing higher levels of poor supervision than females.

Table 5. Group differences in parenting for females only (mean (95% confidence intervals of the mean))

APQ subscales	CD/HCU (*n* = 89)	CD/LCU (*n* = 86)	TD (*n* = 261)
Positive parenting			
Parental involvement	24.10 (22.93, 25.27)	27.16 (25.63, 28.70)	31.28 (30.68, 31.87)
Positive reinforcement	17.27 (16.28, 18.26)	18.34 (17.21, 19.46)	21.07 (20.61, 21.52)
Poor supervision	31.37 (29.85, 32.89)	28.90 (27.11, 30.68)	21.93 (21.15, 22.71)
Negative parenting			
Inconsistent discipline	19.91 (19.09, 20.73)	18.97 (18.03, 19.90)	15.46 (15.04, 15.87)
Corporal punishment	5.58 (4.98, 6.19)	5.65 (5.07, 6.23)	3.89 (3.70, 4.07)

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. APQ = Alabama Parenting Questionnaire

Table 6. Group differences in parenting for males only (mean (95% confidence intervals of the mean))

APQ subscales	CD/HCU (*n* = 89)	CD/LCU (*n* = 86)	TD (*n* = 261)
Positive parenting			
Parental involvement	24.93 (23.24, 26.63)	28.08 (26.47, 29.69)	24.93 (23.24, 26.63)
Positive reinforcement	17.81 (16.67, 18.95)	19.60 (18.51, 20.69)	20.44 (19.80, 21.07)
Poor supervision	29.27 (27.51, 31.02)	27.03 (25.23, 28.82)	23.23 (22.25, 24.21)
Negative parenting			
Inconsistent discipline	19.07 (18.23, 19.91)	18.73 (17.94, 19.52)	16.37 (15.87, 16.87)
Corporal punishment	5.23 (4.66, 5.79)	4.65 (4.21, 5.09)	4.07 (3.82, 4.32)

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. APQ = Alabama Parenting Questionnaire

There was a significant sex by group interaction for parental involvement ($F_{(2, 742)} = 3.19$, $p = .04$), poor supervision ($F_{(2, 742)} = 5.91$, $p = .003$), inconsistent discipline ($F_{(2, 742)} = 3.94$, $p = .02$), and corporal punishment ($F_{(2, 742)} = 6.56$, $p = .002$), but no significant sex by group interaction for positive reinforcement ($F_{(2, 742)} = 2.49$, $p = .08$). Each of the significant interactions resulted from the same pattern; in TD youths, females experienced 'better' parenting than males, but in both CD groups, females experienced 'poorer' parenting than males. The pattern of group differences within each sex was nonetheless in the same direction as for the main effects (*i.e.* CD/HCU < CD/LCU < TD).

3.3.4 Angle-GMLVQ Classifier Performance

Angle-GMLVQ model performance is shown in **Table 7**. All models performed significantly better than chance, as hypothesised (binomial tests, $p < .001$). The HCU-

TD model demonstrated the lowest error rate (MCER of 0.26), followed by the Mixed-TD model (MCER 0.29) and then the LCU-TD model (MCER 0.33). Although the HCU-LCU model was significantly above chance, it did not perform well (MCER 0.42). This pattern of performance indicates considerable overlap in the parenting experiences of youths with CD/HCU and CD/LCU. Furthermore, performance for the LCU-TD model was significantly worse than for the Mixed-TD model, indicating that splitting the CD/mixed group into CD/HCU and CD/LCU groups was beneficial only to the CD/HCU group in terms of classifier performance. This reflects greater overlap between CD/LCU and TD groups than between CD/HCU and TD groups. Our final hypothesis – that both HCU-TD and LCU-TD classifiers would outperform the Mixed-TD classifier – was thus not supported.

Table 7. Angle-GMLVQ model performance (mean (95% confidence intervals of the mean))

	Mixed-TD	**HCU-TD**	**LCU-TD**	**F (*p*), partial η^2**	**HCU-LCU**
Accuracy	0.71 (0.71, 0.71) [a]	0.75 (0.75, 0.76) [b]	0.69 (0.68, 0.69) [c]	724.84 (<*.001*), .33	0.58 (0.57, 0.58)
PPV	0.73 (0.73, 0.74) [a]	0.62 (0.62, 0.62) [b]	0.53 (0.52, 0.53) [c]	3675.35 (<*.001*), .71	0.58 (0.58, 0.58)
NPV	0.69 (0.69, 0.70) [a]	0.84 (0.83, 0.84) [b]	0.79 (0.79, 0.80) [c]	3820.59 (<*.001*), .72	0.58 (0.58, 0.58)
TPR	0.66 (0.66, 0.67) [a]	0.69 (0.68, 0.69) [b]	0.63 (0.62, 0.63) [c]	176.23 (<*.001*), .11	0.58 (0.57, 0.58)
TNR	0.76 (0.75, 0.76) [a]	0.79 (0.78, 0.79) [b]	0.72 (0.71, 0.72) [c]	422.08 (<*.001*), .22	0.58 (0.57, 0.59)
MCER	0.29 (0.28, 0.29) [a]	0.26 (0.26, 0.27) [b]	0.33 (0.33, 0.33) [c]	609.73 (<*.001*), .29	0.42 (0.42, 0.43)

Notes: Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. PPV = positive predictive value, NPV = negative predictive value, TPR = true positive rate, TNR = true negative rate, MCER = macro-averaged classification error rate. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected). Note that the HCU-LCU model (column 6) was not included in statistical tests as comparisons between this and other models were not relevant to hypotheses

3.3.5 Feature Relevance

Feature relevance scores for the HCU-TD, LCU-TD and HCU-LCU models are displayed in **Figure 10**. The pattern of relevance scores generally supported our hypotheses, *i.e.*, that a combination of positive and negative parenting would be relevant to the HCU-TD model, negative parenting would be more relevant to the LCU-TD model and positive parenting would be more relevant to the HCU-LCU model. The exception to this was positive reinforcement, which was consistently low in relevance across all models. Despite this, youths with CD/LCU were distinguished from TD youths almost entirely by negative parenting, and from youths with CD/HCU almost entirely by positive parenting and poor supervision.

3.3.6 Effects of Changing the Relevance Threshold

Figure 11 displays the relevance scores for the HCU-TD, LCU-TD and HCU-LCU models at the stricter (top 15%), standard (top 20%) and more lenient (top 25%) thresholds for 'high' scores. Altering the threshold did not change the pattern of feature relevance in any model. This confirms that differences in relevance were unlikely to be an artefact of the chosen threshold.

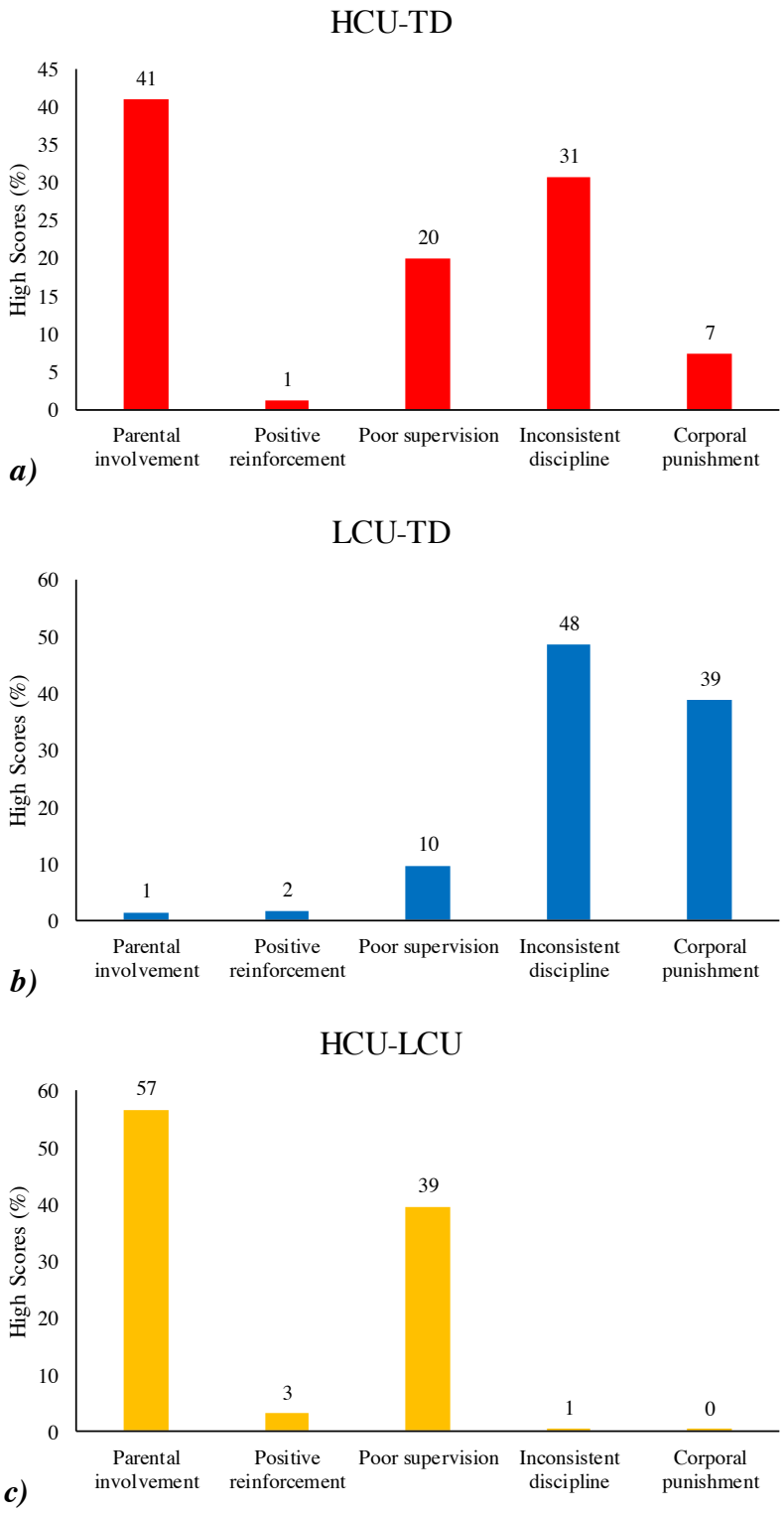


Figure 10. Feature relevance for ***a)*** HCU-TD model, ***b)*** LCU-TD model and ***c)*** HCU-LCU models. Bars show percentage of re-samplings in which feature relevance was in the top 20% of relevance scores across all re-samplings with MCER ≤ 0.40

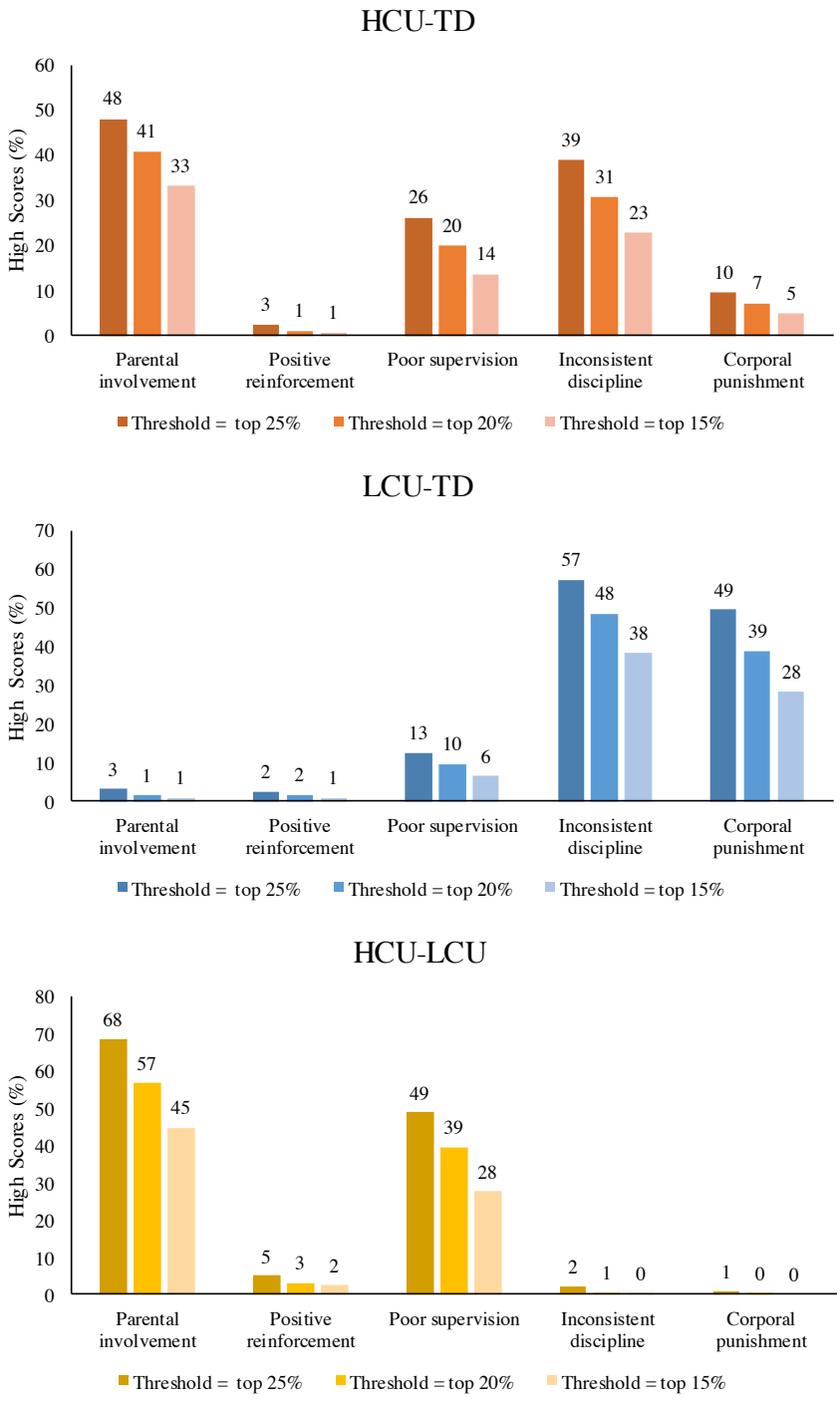


Figure 11. Effect on feature relevance scores of changing the 'high score' threshold. Bars show percentage of re-samplings in which feature relevance was above the threshold, across all runs with MCER ≤ 0.40

3.3.7 SVM Classifier Performance

Of all the SVM classifiers tested, the linear SVM exhibited the best performance for each model (see **Table 8**). MCER values for the linear SVMs were 0.28 for the Mixed-TD model, 0.25 for the HCU-TD model, 0.32 for the LCU-TD model and 0.42 for the HCU-LCU model. For the Mixed-TD, HCU-TD and LCU-TD models the linear SVM classifier outperformed the corresponding Angle-GMLVQ model by a very small but statistically significant margin (Mixed-TD: $t_{(1998)} = 2.90, p = .004$, partial $\eta^2 = .004$. HCU-TD: $t_{(1998)} = 5.04, p < .001$, partial $\eta^2 = .01$. LCU-TD: $t_{(1998)} = 3.95, p < .001$, partial $\eta^2 = 0.01$). For the HCU-LCU model, the two classifiers did not differ significantly ($t_{(1998)} = 1.30, p = .19$, partial $\eta^2 = .001$). These differences in performance indicate that the absolute magnitudes of differences in parenting scores between participants were generally very slightly more informative than the pattern of differences alone.

Table 8. Performance (MCER) for linear and non-linear SVM models (mean (95% confidence intervals of the mean))

	Mixed-TD	HCU-TD	LCU-TD	HCU-LCU
Linear	0.28 (0.28, 0.29) [a]	0.25 (0.25, 0.26) [a]	0.32 (0.32, 0.32) [a]	0.42 (0.41, 0.42) [a]
2nd order polynomial	0.29 (0.29, 0.29) [a]	0.28 (0.28, 0.28) [b]	0.32 (0.32, 0.33) [a]	0.42 (0.42, 0.42) [a]
3rd order polynomial	0.31 (0.31, 0.31) [b]	0.32 (0.32, 0.32) [c]	0.37 (0.36, 0.37) [b]	0.47 (0.46, 0.47) [b]
4th order polynomial	0.36 (0.35, 0.36) [b]	0.37 (0.36, 0.37) [d]	0.41 (0.40, 0.41) [c]	0.49 (0.49, 0.50) [c]
5th order polynomial	0.40 (0.39, 0.40) [c]	0.37 (0.37, 0.37) [d]	0.42 (0.41, 0.42) [d]	0.49 (0.49, 0.50) [c]
6th order polynomial	0.40 (0.40, 0.40) [c]	0.40 (0.39, 0.40) [e]	0.43 (0.43, 0.43) [e]	0.50 (0.49, 0.50) [c]
F (*p*)	1772.20 (*<.001*)	1219.54 (*<.001*)	891.61 (*<.001*)	318.26 (*<.001*)

Notes: MCER = macro-averaged classification error rate. Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. Within columns, models with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

3.4 Discussion

This study addressed differences in exposure to positive and negative parenting in youths with CD/HCU, youths with CD/LCU and TD youths. We first investigated differences at the group level, hypothesising that youths with CD/HCU would experience high levels of negative parenting and low levels of positive parenting relative to TD youths, while youths with CD/LCU would be characterised mainly by high levels of negative parenting. This hypothesis was partially supported, since youths with CD/LCU as well as those with CD/HCU experienced more negative and less positive parenting than TD youths (although only positive parenting differed significantly between the CD/HCU and CD/LCU groups). Next, we used Angle-GMLVQ classifiers to assess the extent to which positive and negative parenting was

predictive of the diagnostic status of individual youths. As predicted, all models performed at above chance levels. Contrary to our next hypothesis, splitting the CD/mixed group into CD/HCU and CD/LCU subtypes resulted in improved classification only for youths with CD/HCU. Finally, as hypothesised, both positive and negative parenting behaviours were highly relevant when distinguishing youths with CD/HCU from TD youths, while youths with CD/LCU were distinguished from TD youths almost exclusively by negative parenting and from youths with CD/HCU almost exclusively by higher positive parenting and lower levels of poor supervision.

3.4.1 Group Level Differences in Positive Parenting

At the group level, youths with CD/HCU experienced significantly lower levels of parental involvement and positive reinforcement than youths with CD/LCU, who in turn experienced lower levels than TD youths. Thus, in terms of group level differences, low positive parenting was more strongly associated with CD/HCU than with CD/LCU. It is also interesting to note that poor supervision varied in line with positive parenting rather than negative parenting in the current sample. Notably, differences between the CD/LCU and TD groups were reduced to non-significance after controlling for family structure, suggesting that low positive parenting in the CD/LCU group in particular might partially reflect caregiver absence rather than poor parenting *per se*.

Numerous studies have reported similar associations between low positive parenting and CD/HCU or CU traits (*e.g.*, Pasalich *et al.*, 2011; Ray *et al.*, 2017; Elizur & Somech, 2018; Hyde *et al.*, 2016; Pardini *et al.*, 2007; Muratori *et al.*, 2016; Waller *et al.*, 2016; Waller *et al.*, 2017; Waller *et al.*, 2018; but *c.f.* Falk & Lee, 2012). In early childhood, positive parenting has been associated with better empathy and pro-sociality

via more enjoyable parent-child interactions and desire to comply with parental demands (Kochanska *et al*., 2005). It is plausible that a similar mechanism occurs in adolescence (*e.g*., Ray *et al*., 2017). Alternatively, adolescents might continue to benefit from positive parenting experienced earlier in life. Since punishment seems to be less effective at high levels of CU traits (Blair *et al*., 2001), reciprocally warm and committed relationships are likely to be an especially important protective factor throughout childhood, even if positive parenting is indeed more effective in infancy.

There are other, non-causative factors that likely contribute to the observed associations between low positive parenting and CD/HCU. Genetic similarity between parent and child is an obvious candidate, given the high heritability of CU traits (Viding *et al*., 2005). It seems likely that CU traits in parents would manifest as lack of warmth and interest towards the child. Nonetheless, it is clear from adoption studies that positive parenting is associated with reductions in CU traits and behaviour even in the absence of a genetic relationship between parent and child (Hyde *et al*., 2016; Waller *et al*., 2016; Waller *et al*., 2017). Early warm-positive parenting also reduces the heritability of CU traits (Henry, Dionne, Viding, Vitaro, Brendgen, Tremblay, & Boivin, 2018). Furthermore, positive parenting interventions have proved efficacious for young children (Elizur & Somech, 2018), suggesting a causal relationship in addition to shared genetic effects. Finally, associations with parenting are bidirectional; child temperament influences parenting as well as vice versa (Flom *et al*., 2019; Muñoz, Pakalniskiene & Frick, 2011; Salihovic, Kerr, Özdemir, & Pakalniskiene, 2012). Furthermore, genetic predispositions in the child might indirectly elicit certain parenting behaviours (Flom *et al*., 2019). Thus, while low positive parenting is potentially causally linked to CD/HCU,

there are additional factors that plausibly contribute to the observed associations between low positive parenting and CD/HCU.

3.4.2 Group Level Differences in Negative Parenting

Youths with both CD/HCU and CD/LCU reported higher levels of negative parenting than TD youths, but unlike positive parenting, exposure to negative parenting did not differ significantly in CD/HCU versus CD/LCU. The literature is divided on the relative importance of negative parenting in these subtypes. There are reports of potential insensitivity to negative parenting in CD/HCU (Wootton *et al*., 1997) as well as high levels of negative parenting (Fontaine *et al*., 2011; Barker *et al*., 2011) and a role for negative parenting (harshness) in both aggression and CU traits (Waller *et al*., 2018). The associations between negative parenting, CD/HCU and CD/LCU observed here do not warrant conclusions about causation. However, in the current sample at least, negative parenting did not appear to be a specific risk factor for CD/HCU over and above the risk for CD generally.

3.4.3 The Relative Importance of Positive and Negative Parenting Behaviours for Classifying Individuals

There were very clear differences between classifier models in the relative importance of positive and negative parenting behaviours. Overall, positive parenting was highly relevant when distinguishing individuals who differed substantially on CU traits (HCU-TD and HCU-LCU models), while negative parenting was highly relevant when distinguishing individuals who differed substantially on CD symptoms (HCU-TD and LCU-TD models). It is especially interesting to note that positive parenting was very

low in relevance in the LCU-TD model, despite significant differences in both positive and negative parenting at the group level. This implies that even when exposure to positive parenting is low, negative parenting is a much more reliable indicator of CD/LCU. This pattern of relevance aligns with previous research indicating that negative parenting is linked to conduct problems and CU traits while low positive parenting is linked primarily to CU traits (*e.g.*, Waller, Gardner, Viding, Shaw, Dishion, Wilson, & Hyde, 2014). Notably, however, positive reinforcement was low in relevance across all models. This was unexpected, and suggests that when the broader parenting context is taken into account, low positive reinforcement is not strongly indicative of CD/HCU or CD/LCU. In this context, it is interesting to note that the APQ positive reinforcement subscale includes items assessing material rewards ('your parents reward or give something extra to you for behaving well') as well as warmth (*e.g.*, 'your parents hug or kiss you when you have done something well'). In previous studies linking CU traits to low positive parenting, warmth, rather than material rewards, has often been a key measure (*e.g.*, Waller *et al.*, 2014). The effects of warmth and praise versus provision of material incentives will be an interesting topic for future research on parenting.

Finally, it should be noted again here that relevance scores were derived only from the higher performing models. This ensured that relevance scores were not influenced by models that failed to distinguish between groups. However, in the case of the poorly performing HCU-LCU model, it also means that relevance scores are not necessarily reflective of the full CD sample. Thus, while positive parenting was key to classifying those youths who could be distinguished, many youths with CD/HCU could not be distinguished from those with CD/LCU based on parenting. Thus, although positive

parenting clearly differs between CD/HCU and CD/LCU at the group level, and between many individuals in these groups, positive parenting is certainly not a universally distinguishing factor between CD/HCU and CD/LCU.

3.4.4 Strengths and Limitations

This study has several strengths, including the large and well-characterised sample, the combination of parent and child-report parenting measures and the use of both univariate and multivariate techniques. However, it also has several limitations, particularly in terms of mechanistic understanding. For example, age effects were controlled for – rather than investigated – due to practical constraints with the sample size. Cultural differences between sites were also not investigated for the same reason. In addition, the cross-sectional nature of the data meant that the direction of effects could not be explored. Genetically informed longitudinal studies will be essential for understanding the mechanisms underlying the observed relationships between parenting, CD and CU traits. Finally, we relied on parent and child reports of parenting, which though meaningful, are not necessarily objective. These reports would ideally have been complemented with observational data.

3.4.5 Summary and Conclusions

In summary, CD/HCU and CD/LCU were distinguished from TD by both positive and negative parenting , although only negative parenting was highly relevant for distinguishing CD/LCU from TD at the individual level. CD/HCU was distinguished from CD/LCU primarily by positive parenting, at both the group and individual levels. This adds to a growing body of literature suggesting that parenting is associated with

both CD/HCU and CD/LCU, but that the specific parenting practices associated with CD and CU traits are different. We suggest that future research should further distinguish between different parenting behaviours (*e.g.*, verbal praise and affection versus provision of material incentives), as well as multiple types of externalising behaviour (*e.g.*, CU-type/proactive versus reactive aggression, hyperactive-impulsive *etc*.). Making such distinctions will hopefully contribute to the development of more targeted parenting interventions in the future.

CHAPTER 4: GREY MATTER VOLUME DIFFERENCES IN CONDUCT DISORDER WITH HIGH VERSUS LOW LEVELS OF CALLOUS-UNEMOTIONAL TRAITS

4.1 Introduction

The affective and behavioural characteristics of CD/HCU and CD/LCU are reflected in neurobiological differences in the brain. According to Blair's neurocognitive theory of psychopathy, these differences relate to levels of proactive and reactive aggression (*e.g.*, Blair, 2013; see also Chapter 1, section 1.4.3). Proactive aggression is particularly prominent in CD/HCU, and is hypothesised to reflect dysfunction in a network of regions including the amygdala, caudate (part of the dorsal striatum), orbitofrontal and ventromedial prefrontal cortex and anterior insula. These regions are implicated in stimulus-reinforcement and response-outcome learning, and disruptions to these processes mean that social cues are not associated with appropriate behavioural responses. For example, a frightened facial expression from a victim is not associated with a cessation of violence. Perhaps as a secondary consequence of this failure, the anterior insula – which is implicated in interoception (*e.g.*, Craig, 2009) – is also dysfunctional in youths with CD/HCU (Blair, 2013). By contrast, reactive aggression is elevated in both CD/HCU and CD/LCU, and is hypothesised to arise from hyper-reactivity of the basic neural threat circuit, which includes the amygdala, hypothalamus and periaqueductal gray. Dysfunction in this circuit is associated with increased startle response and difficulty recognising anger, as well as reactive aggression (Blair *et al.*, 2005).

Neurobiological differences between the subtypes are evidenced by divergent blood-oxygen-level dependent (BOLD) signal to emotional stimuli in fMRI studies (*e.g.*, Viding *et al.*, 2012), and there is some evidence for grey matter volume differences as well (*e.g.*, Sebastian *et al.*, 2016). However, this evidence is currently limited. It is unclear to what extent, and in what brain regions, the subtypes differ from each other and from TD youths. We aimed to address these questions using a combination of traditional univariate methods and machine learning classification.

4.1.1 Grey Matter Volume Reductions in CD

Three large meta-analyses have demonstrated reductions in grey matter volume in youths with conduct problems, including CD, relative to TD youths (Raschle, Menks, Fehlbaum, Tshomba, & Stadler, 2015; Noordermeer, Luman, & Oosterlaan, 2016; Rogers & De Brito, 2016). In a meta-analysis of eight VBM studies, Raschle *et al.* (2015) found 19 clusters of reduced grey matter volume. The largest of these clusters was in the region of the right inferior frontal lobe, precentral gyrus and insula. Other large clusters were found across the right subcallosal gyrus, putamen, amygdala and lateral globus pallidus, the right inferior frontal gyrus and the left insula. Similarly, Noordermeer *et al.* (2016) reported reductions in four clusters, located in the left amygdala, left insula, left medial/superior frontal gyrus and right insula. Finally, in a meta-analysis of 13 studies, Rogers and De Brito (2016) demonstrated reductions in the left amygdala, left fusiform gyrus, right insula, superior frontal gyrus and (in youths with childhood-onset conduct problems only) the left insula. Interestingly, within their CD group, CU traits were negatively correlated with the magnitude of reduction in grey

matter volumes in the left lentiform nucleus/putamen and right amygdala (although this analysis included only five studies).

4.1.2 Grey Matter Differences in Youths with CD/HCU, CD/LCU and TD Youths

Only one VBM study has directly compared group differences in grey matter volume in youths with CD/HCU, youths with CD/LCU and TD youths (Sebastian *et al*., 2016). Out of four regions of interest (amygdala, orbitofrontal cortex, anterior insula and anterior cingulate cortex), the CD group as a whole exhibited lower grey matter volume in the bilateral orbitofrontal cortex relative to TD youths. The CD/HCU group exhibited reductions in the left orbitofrontal cortex relative to CD/LCU and TD groups (who did not differ significantly from each other), and in the right anterior cingulate cortex relative to the TD group only. Furthermore, there was a significant negative correlation between CU traits and grey matter volume in the left orbitofrontal cortex within the CD group. In whole brain analyses, the CD/HCU group also had reduced grey matter volume in the left middle frontal gyrus compared to the TD group. Interestingly, a study that compared youths with CD/HCU to TD youths (De Brito *et al*., 2009) reported increased grey matter concentration in the medial orbitofrontal and anterior cingulate cortices of boys with CD/HCU. Whole brain analyses also revealed increased grey matter volume and concentration in the bilateral temporal lobes. Based on post-hoc analyses, these increases were attributed to delayed cortical maturation in the CD/HCU group (who, with a mean age of 11 years, were younger than in most studies).

4.1.3 Associations between Grey Matter and Dimensional Measures of CU Traits

Several studies comparing CD and TD groups have also included dimensional measures of CU traits. These studies commonly report negative correlations between CU traits and grey matter volume in the anterior insula (Fairchild *et al*., 2013; Sterzer *et al*., 2007). A negative correlation between CU traits and grey matter concentration in the right insula has also been reported (Cohn *et al*., 2016). However, the same study reported a positive correlation between CU traits and grey matter volume in the left anterior insula in youths with below-median CD scores. Similarly, Fairchild *et al*. (2015) reported a positive correlation between CU traits and insula cortical folding. Positive correlations between CU traits and anterior insula grey matter volume have also been reported in a TD-only sample (Raschle *et al*., 2018). Thus, while the direction of association in CD is more commonly negative, it is not consistent across studies.

Besides the insula, significant correlations in other regions are reported more scarcely. In addition to the association they observed in the anterior insula, Fairchild *et al*. (2013) found that CU traits in females were positively correlated with bilateral middle/superior orbitofrontal cortex grey matter volume and negatively correlated with left striatal volume (although these correlations were reduced to non-significance when controlling for CD symptoms). Finally, Wallace, White, Robustelli, Sinclair, Hwang, Martin and Blair (2014) demonstrated a negative correlation between cortical thickness and CU traits in the superior temporal lobe.

4.1.4 Machine Learning Classification in CD and Psychopathy

In recent years, a small number of researchers have applied machine learning classifiers to neuroimaging data in the field of CD and psychopathy. Cope *et al*. (2014) used

SVMs to distinguish youth homicide offenders (n = 20) from non-homicide offenders (n = 135) using a combination of grey matter volumes and other measures. Regions of interest for grey matter volumes were selected because they discriminated between groups in a separate sample. After feature selection, the features used in the final model were mean grey matter volumes in the left orbitofrontal cortex, medial orbitofrontal cortex, anterior and posterior cingulate, right and left temporal poles, and psychopathy scores (affective dimension), total number of convictions, and socioeconomic status. Overall classification accuracy was 81%. Notably, however, psychopathy scores were a predictor rather than an outcome in this study. In a separate study, youths were classified into psychopathic (n = 71) and non-psychopathic offenders (n = 72) using grey matter volumes alone (Steele *et al*., 2017). Features used in this model were mean grey matter volumes in the anterior cingulate cortex, bilateral amygdala, bilateral hippocampus, medial orbitofrontal cortex, bilateral parahippocampus, posterior cingulate cortex and bilateral temporal poles. Overall accuracy was 69%. Additional classifiers, using the same features but distinguishing non-offenders (n = 21) from *a)* psychopathic and *b)* non-psychopathic offenders, achieved 83% and 81% accuracy respectively. In a similar study with a smaller sample (15 psychopaths and 15 healthy controls), Sato *et al*. (2011) achieved an accuracy of 80% when distinguishing psychopaths from healthy controls. In contrast to Cope *et al*. (2014) and Steele *et al*. (2017), individual voxels were used as features rather than mean volumes from larger anatomical areas. A feature selection process was applied, so that only the most discriminative voxels were used in the final model. The most relevant voxels for classification were located in the bilateral superior temporal sulcus.

Finally, two studies recently used SVMs to distinguish youths with CD from TD youths based on grey matter volumes. First, Zhang *et al*. (2018a) used SVMs with a searchlight method to distinguish youths with CD (n = 60) from TD youths (n = 60). With a searchlight, multiple classifiers are trained and tested throughout the brain. Each classifier takes all the voxels within a local sphere (the searchlight) as features. The classification accuracy is then assigned to the central voxel. In this way, a map of regional classification accuracies is generated across the whole brain. Zhang *et al*. (2018a) achieved a maximum classification accuracy of 83%, with significantly above-chance classification in the posterior lobe of the cerebellum, temporal lobe, parahippocampal gyrus, lingual gyrus, insula, parietal lobe and medial frontal gyrus. Second, Zhang *et al*. (2018b) used three classifier types – SVM, logistic regression and random forest – to classify youths with CD (n = 60) from TD youths (n = 60). Regions where grey matter volumes differed significantly between groups (including the frontal and parietal lobes, anterior cingulate, posterior lobe of the cerebellum, lingual gyrus and insula) were used as features (importantly, the subset of the sample on whom the classifier was tested was not included in the univariate analyses used to select the features). Classification accuracies ranged from 78% to 80%. In summary, regional differences in grey matter volume show good promise for distinguishing between psychopathic and non-psychopathic individuals, and between youths with CD and TD youths. However, it is notable that sample sizes have generally been small, which might inflate accuracy due to sample homogeneity (see Janssen *et al*., 2018). The utility of grey matter volumes for distinguishing between CD/HCU and CD/LCU remains to be seen.

4.1.5 Hypotheses

In the present study, we first used VBM to establish whether there were group level differences in grey matter volume between CD and TD groups, and between CD/HCU, CD/LCU and TD groups. We then used Angle-GMLVQ to investigate whether grey matter volumes in key regions of interest differentiated between these youths at the individual level. In doing so, we aimed to clarify which regions (if any) differed between CD subtypes, and whether these differences could reliably distinguish between individual youths.

Guided by meta-analytic evidence (Raschle *et al*., 2015; Noordermeer *et al*., 2016; Rogers & De Brito, 2016), we anticipated reductions in grey matter volume in youths with CD relative to TD, particularly in frontal regions, insula, amygdala and fusiform gyrus. Our hypotheses regarding CD/HCU and CD/LCU differences were more exploratory, given the paucity of literature and previous inconsistencies in the direction of association between CU traits and grey matter volume (Fairchild *et al*., 2013; Cohn *et al*., 2016). Likewise, we did not make predictions about classifier performance when comparing CD/HCU, CD/LCU and TD groups. We did, however, anticipate that if the subtypes are indeed distinctive in terms of regional grey matter volumes, then performance of a CD-against-TD classifier (Mixed-TD) would be worse than for CD/HCU-against-TD (HCU-TD) and CD/LCU-against-TD (LCU-TD) classifiers.

4.2 Methods

4.2.1 Participants

Participants were drawn from the FemNAT-CD sample. Eligibility criteria, recruitment and testing procedures are described in Chapter 2 (see also Kersten *et al*., 2017). The FemNAT-CD sample consisted of 1743 participants in November 2017, 612 of whom had structural MRI data. Of these, eight participants were excluded because they did not appear to meet the inclusion criteria for FemNAT-CD, 35 were excluded due to poor quality MRI data or pre-processing failure (described in more detail below) and 13 did not complete the measure of CU traits.

Participants with CD were divided into CD/HCU and CD/LCU groups based on a median split (median = 34) of the total parent-report CU scores for the CD group, in line with previous research (*e.g*., Viding *et al*., 2012)[12]. Ten TD participants were excluded for having CU scores above the CD median, and 94 TD participants were excluded during the group matching process using Match software (van Casteren & Davis, 2007). Groups were matched on age, pubertal status, proportion of females, site of data collection and structural scan quality (see below). This left a final sample of 452 participants (females n = 190; CD/HCU n = 113, CD/LCU n = 113, TD n = 226). Finally, a 'mixed' CD group consisting of both high and low CU participants ('CD/mixed') was formed by combining all the CD/HCU and CD/LCU participants into one group (n = 226).

[12] Note that while a tertile split would have resulted in better separation of groups, the resultant loss of data could not be justified here given the smaller overall sample size than in Chapter 3. The cut-off points for a tertile split were <30 (CD/LCU) and >39 (CD/HCU). Only 85 participants qualified as CD/HCU and 81 as CD/LCU, before group matching.

The final sample (n = 452) did not differ from the excluded participants (n = 1291) on age, sex, IQ, CU trait scores or number of current CD symptoms (t-tests, 2-tailed, all *p*s > .05). The CD group in the final sample also did not differ from the excluded participants with CD (n = 629) on number of current CD symptoms. However, the CD group in the final sample had lower CU trait scores than the excluded participants with CD (mean for excluded participants = 35.63, mean for CD sample = 33.54; $t_{(862)} = 2.31$, $p = .02$).

4.2.2 Questionnaire and Interview Measures

The K-SADS-PL (Kaufman *et al*., 1997) was used to assess participants for CD and other mental disorders. Wechsler Intelligence Scales (Wechsler, 1999; Wechsler, 2008) were used to estimate IQ, and the PDS (Petersen *et al*., 1988) was used to measure pubertal development. The ICU (Essau *et al*., 2006b) was used as the measure of CU traits. Reliability was good in the current sample (Cronbach's alphas: callous $\alpha = 0.83$, uncaring $\alpha = 0.86$, unemotional $\alpha = 0.77$, total CU $\alpha = 0.91$). As an illustration of sample characteristics, the RPQ (Raine *et al*., 2006), GEM (Dadds *et al*., 2008) and the CBCL internalising, externalising and total problems scales (Achenbach, 1991) are also reported. RPQ and GEM subscale reliabilities were good; RPQ proactive $\alpha = 0.87$, reactive = 0.88, total aggression $\alpha = 0.92$; GEM affective $\alpha = 0.77$, cognitive $\alpha = 0.76$, total $\alpha = 0.84$). These measures are described more fully in Chapter 2 (sections 2.4.6 – 2.4.7).

4.2.3 MRI Data Acquisition

T1-MPRAGE scans were collected at five sites: Frankfurt, Aachen, Southampton, Basel and Birmingham. Details of scanners and acquisition parameters at each site are shown in **Table 9**. Site qualification procedures were conducted to ensure comparability of procedures across sites. These procedures involved scanning of phantoms and a human volunteer at each site, followed by reviews and adjustments as necessary by an MRI physicist at the University of Birmingham.

Table 9. Site-specific scanner information and acquisition parameters

Site:	**Frankfurt (n = 101)**	**Aachen (n = 91)**	**Southampton (n = 110)**	**Basel (n = 55)**	**Birmingham (n = 95)**
Scanner model	Siemens Trio	Siemens Prisma	Siemens Trio	Siemens Prisma	Phillips Achieva
No. slices	192	192	192	192	192
TR	1900ms	1900ms	1900ms	1900ms	1900ms
TE	2.74ms	3.42ms	4.10ms	3.42ms	3.70ms
TI	900ms	900ms	900ms	900ms	900ms
Flip angle (°)	9	9	9	9	9
Field of view	256mm	256mm	256mm	256mm	256mm
Voxel size	1×1×1mm	1×1×1mm	1×1×1mm	1×1×1mm	1×1×1mm

Notes: TR = repetition time, TE = echo time, TI = inversion time

MRI data Pre-processing

The data were pre-processed using Statistical Parametric Mapping 12 (SPM12) and two SPM toolboxes; the Computational Anatomy Toolbox (CAT12; Gaser & Dahnke, 2016) and the Template O'Matic (TOM) toolbox (Wilke, Holland, Altaye & Gaser, 2008). Given the paediatric nature of the sample, customised tissue probability maps were created using the matched-pair approach of the TOM toolbox with each participant's

age and sex as defining variables. The pre-processing included the following steps: first, an affine transformation was used to align the origin (0, 0, 0) of each scan to the anterior commissure, with its axis parallel between the anterior and posterior commissures. Next, scans were corrected for bias-field inhomogeneities, segmented with reference to the customised tissue probability maps, spatially normalised (affine-only transformation), and modulated with respect to the deformation fields produced during normalisation. Segmentation accuracy was visually inspected for each participant, and scan quality was checked using CAT12 quality reports. These reports are generated automatically by CAT12 during segmentation, and provide a percentage score and corresponding letter grade indicating the scan quality. Quality ratings are based on combined ratings of noise, bias or inhomogeneities and image resolution[13]. Following this initial segmentation, participants with a structural scan quality rating below B- were excluded from further analyses (n = 34, as described above). One additional participant was excluded due to repeated segmentation failure. Scans for the remaining participants (n = 560)[14] were then re-segmented with a second set of customised tissue probability maps generated as above. All scans had a quality rating of B- or higher after the second segmentation. Scans were then smoothed with a 6mm full width at half maximum (FWHM) kernel and normalised to MNI (ICBM) space using a study-specific custom template generated with the Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL; Ashburner, 2007) toolbox. Finally, total intracranial volume was estimated for each participant.

[13] Ratings of A+ to A- are considered excellent, ratings of B+ to B- are considered good, C+ to C- are satisfactory, D+ to D- are sufficient, E+ to E- are critical, and scans below this level are rated as failed.
[14] Four participants were discovered to be ineligible only after pre-processing; hence the total number of participants excluded after pre-processing was 108 (ineligibility/missing data = 4, excluded during matching = 104). Matching was conducted after pre-processing so that participants could be matched on scan quality.

4.2.4 VBM Mass Univariate Analyses

Analyses were performed on a voxel-by-voxel basis using the General Linear Model (GLM) in SPM12, comparing first the CD/mixed and TD groups and then the CD/HCU, CD/LCU and TD groups individually (*i.e.*, CD/HCU with TD, CD/LCU with TD, and CD/HCU with CD/LCU). For all analyses, we used a height threshold of $p < .05$, Family-Wise Error (FWE) corrected for multiple comparisons across the brain (for whole brain analyses) and within each region (for region of interest analyses).

In the main VBM models, we included full-scale IQ, sex, pubertal status, site of data collection and total intracranial volume as covariates of no interest. We also created a further set of VBM models using ADHD diagnoses as an additional covariate of no interest. These models were used to control for potential confounds due to comorbid ADHD.

We explored correlations between grey matter volume and CU traits in the CD group only, using the same covariates as above (including the additional models with ADHD diagnoses as a covariate). Finally, group by sex interactions were explored, using a full factorial 3 (group) × 2 (sex) design. Covariates of no interest in this model were IQ, pubertal status, site of data collection and total intracranial volume. As above, this analysis was then repeated with ADHD diagnoses as an additional covariate of no interest.

4.2.5 Region of Interest Selection

We defined regions of interest using 10mm-radius spheres centred on MNI peak coordinates from Rogers and De Brito (2016) and Sebastian *et al*. (2016). These studies

focused specifically on youths with CD, and previously distinguished CD from TD groups (Rogers & De Brito, 2016) and CD/HCU from CD/LCU groups (Sebastian *et al*., 2016). The size of the sphere is consistent with a previous study using a similar approach (Cope *et al*., 2014). The regions distinguishing CD from TD participants were the left amygdala (centred on MNI coordinates x = -32, y = 2, z = -20), right insula (x = 36, y = 20, z = -16), left superior frontal gyrus (x = -6, y = 54, z = 28), left fusiform gyrus (x = -34, y = -78, z = -16) and left insula (x = -40, y =12, z = -12). Finally, the left orbitofrontal cortex (x = -39, y = 44, z = -6) previously distinguished CD/HCU from CD/LCU participants in the only study to compare these groups directly (Sebastian *et al*., 2016). Note that these regions are very similar to those identified in the two other meta-analyses (Raschle *et al*., 2015; Noordermeer *et al*., 2016), hence the decision to define regions based on only the largest meta-analysis. The number of grey matter voxels in each region was 1166 (left fusiform gyrus), 1149 (left amygdala), 998 (left insula), 1094 (left orbitofrontal cortex) 1148 (left superior frontal gyrus) and 1169 (right insula)[15]. These regions are shown in **Figure 12**. For comparison, the significant clusters from Rogers and De Brito (2016) and Sebastian *et al*. (2016), on which these regions of interest are centred, are reproduced below (**Figure 13**).

[15] We repeated the VBM and Angle-GMLVQ analyses using anatomically defined bilateral regions instead of the spherical regions listed here. These analyses are reported in Appendix A. Classifier performance was similar to the main results.

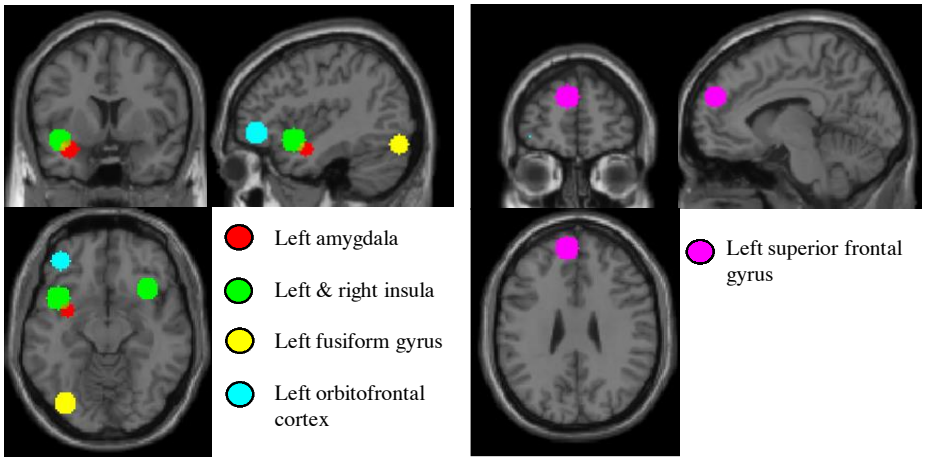
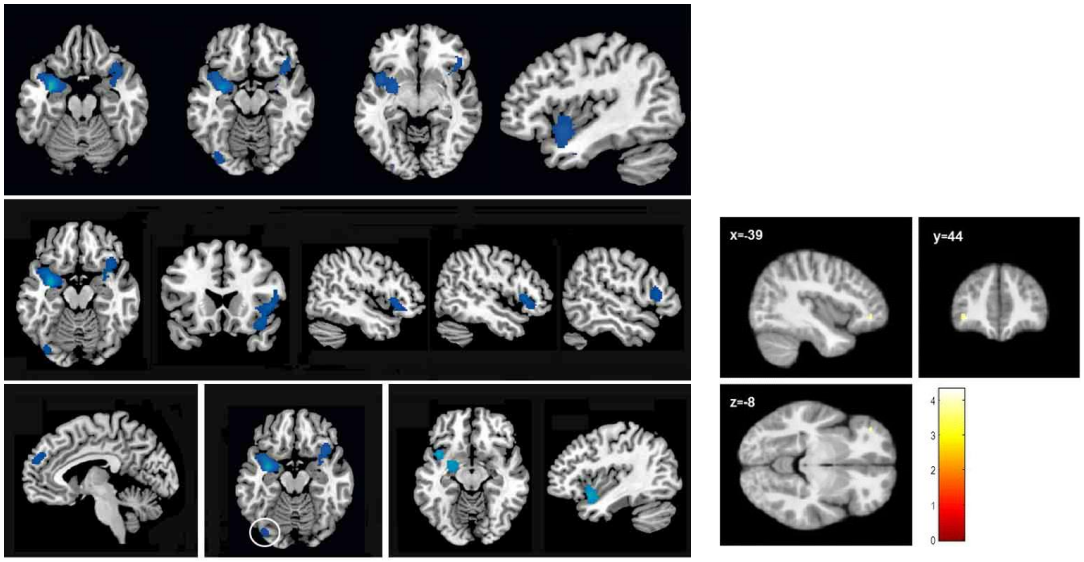


Figure 12. Regions of interest



a) ***b)***

Figure 13. Clusters of significant reductions in grey matter volume for ***a)*** youths with conduct problems relative to TD, in the meta-analysis by Rogers and De Brito (2016), and ***b)*** youths with CD/HCU relative to CD/LCU, from Sebastian *et al*. (2016). Figures reproduced from Rogers and De Brito (2016) and Sebastian *et al*. (2016). Regions of interest in the present study were based on these clusters

4.2.6 Classification Analyses

4.2.6.1 Features

The mean grey matter volume in each region of interest was extracted for each participant. Standardised residuals were then calculated by regressing out the variance associated with IQ, sex, pubertal status, site of data collection and total intracranial volume. These standardised residuals were used as features[16]. Since there were only six features (regions), additional feature selection during model training was not necessary.

4.2.6.2 Classification Models

Data were classified using Angle-GMLVQ classifiers (Bunte *et al*., 2016). For comparison, additional analyses were conducted with SVM classifiers (see below). The classifiers were run separately for each pair of groups of interest:

1. CD/mixed against TD ('Mixed-TD')
2. CD/HCU against TD ('HCU-TD')
3. CD/LCU against TD ('LCU-TD')
4. CD/HCU against CD/LCU ('HCU-LCU')

4.2.6.3 Training and Testing Procedure

The Angle-GMLVQ classifier was trained and tested for each model, with one prototype per class. For each model, a holdout design, with 80% of the data used for training and 20% held out for testing, was repeated for 1000 re-samplings. For imbalanced classes (*e.g*., in the HCU-TD model), in each re-sampling the larger class

[16] We also repeated the classification analyses after having regressed out variance associated with ADHD diagnoses, in addition to IQ, sex, pubertal status, site of data collection and total intracranial volume. These analyses are reported in Appendix B. Classifier performance was generally similar, but poorer for the HCU-TD and HCU-LCU models, suggesting that ADHD diagnoses did contribute somewhat to classification of the CD/HCU group.

was down-sampled to match the size of the smaller class, by removing participants at random from the larger class. For balanced classes (Mixed-TD model), 80% of the participants from each class were used at random in each re-sampling. Mean performance measures were then calculated for each model.

4.2.6.3 Assessment of Model Performance

Models were assessed using overall classification accuracy, PPV, NPV, TPR, TNR and MCER. These measures are described in Chapter 2 (section 2.2.5).

4.2.6.4 Assessment of Feature Relevance

Relevance scores were considered 'high' if they were in the top 20% of scores across all re-samplings with a corresponding MCER of 0.40 or below (with relevance scores normalized for each re-sampling). Features were ranked by number of high scores.

4.2.6.4 SVM Classifiers

The same analyses were repeated with SVM classifiers for comparison. SVM classifiers were trained and tested using the standard MATLAB function 'fitcsvm'. To optimise parameters, we trained and tested six SVM classifiers for each model; linear SVM and SVMs with second, third, fourth, fifth and sixth-order non-linear polynomial transformation kernels. We then selected the best performing SVM (*i.e.*, lowest MCER) for comparison with the corresponding Angle-GMLVQ classifier.

4.3 Results

4.3.1 Sample Characteristics

In line with the group matching criteria, there were no significant differences between groups in number of participants from each site ($\chi^2 = 03.89$ $p = 0.87$, $\varphi = .09$), or

structural scan CAT12 quality ratings (means: CD/HCU = 85.00, CD/LCU = 85.22, TD = 85.08, F = 1.25, $p = 0.29$, partial $\eta^2 = .01$). As seen in **Table 10** and **Table 11**, the CD/HCU and CD/LCU groups did not differ significantly on measures of IQ, aggression or number of CD symptoms. However, all groups differed significantly in number of ODD and ADHD symptoms and CBCL externalising scores, with the CD/HCU group having the more severe presentation. As expected, there were also significant differences between all groups in levels of CU traits (ICU total score), as well as callous and uncaring traits.

Table 10. Demographic and clinical characteristics (mean (95% confidence intervals of the mean) unless stated otherwise)

Measures	CD/HCU (*n* = 113)	CD/LCU (*n* = 113)	TD (*n* = 226)	Test statistic (*p*), effect size
Age	14.23 (13.80, 14.76) [a]	14.30 (13.84, 14.76) [a]	14.19 (13.87, 14.51) [a]	$F = 0.08$ (*.93*), partial $\eta^2 = .00$
Females (%)	43	41	42	$\chi^2 = 0.16$ (*.92*), $\varphi = .02$
PDS pubertal stage	3.48 (3.29, 3.67) [a]	3.37 (3.15, 3.59) [a]	3.42 (3.27, 3.57) [a]	$F = 0.26$ (*.77*), partial $\eta^2 = .00$
Performance IQ	95.00 (92.96, 98.83) [a]	98.37 (95.74, 101.00) [a]	103.51 (101.73, 105.30) [b]	$F = 12.04$ (< *.001*), partial $\eta^2 = .05$
Verbal IQ	91.39 (88.64, 94.15) [a]	94.16 (91.21, 97.12) [a]	103.86 (101.88, 105.85) [b]	$F = 30.78$ (< *.001*), partial $\eta^2 = .12$
Total IQ	94.12 (91.80, 96.43) [a]	96.53 (94.14, 98.92) [a]	103.93 (102.36, 105.51) [b]	$F = 28.71$ (< *.001*), partial $\eta^2 = .11$
ICU callous	15.92 (15.06, 16.78) [a]	7.59 (6.87, 8.31) [b]	4.04 (3.72, 4.37) [c]	$F = 440.44$ (< *.001*), partial $\eta^2 = .66$
ICU uncaring	17.87 (17.30, 18.43) [a]	11.08 (10.29, 11.87) [b]	7.95 (7.43, 8.46) [c]	$F = 254.95$ (< *.001*), partial $\eta^2 = .53$
ICU unemotional	8.99 (8.37, 9.61) [a]	5.62 (5.07, 6.17) [a]	4.99 (4.63, 5.34) [b]	$F = 73.36$ (< *.001*), partial $\eta^2 = .25$
ICU total	42.78 (41.49, 44.06) [a]	24.29 (22.93, 25.65) [b]	16.98 (16.11, 17.83) [c]	$F = 532.19$ (< *.001*), partial $\eta^2 = .70$
K-SADS CD symptoms	5.84 (5.34, 6.34) [a]	5.50 (5.02, 5.20) [a]	0.17 (0.11, 0.23) [b]	$F = 474.69$ (< *.001*), partial $\eta^2 = .68$
K-SADS ODD symptoms	6.80 (6.33, 7.27) [a]	5.72 (5.20, 6.25) [b]	0.16 (0.08, 0.24) [c]	$F = 573.23$ (< *.001*), partial $\eta^2 = .72$
K-SADS ADHD symptoms	9.45 (8.14, 10.76) [a]	7.16 (6.00, 8.31) [b]	0.09 (0.03, 0.15) [c]	$F = 180.71$ (< *.001*), partial $\eta^2 = .45$
K-SADS GAD diagnosis (%)	22	21	2	$\chi^2 = 40.76$ (< *.001*), $\varphi = .30$
K-SADS MDD diagnosis (%)	30	20	1	$\chi^2 = 60.69$ (< *.001*), $\varphi = .37$
K-SADS SUD diagnosis (%)	13	12	0	$\chi^2 = 33.35$ (< *.001*), $\varphi = .27$

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. K-SADS = Schedule for Affective Disorders and Schizophrenia for School-age Children: Present and Lifetime Version (lifetime maximum symptoms/diagnosis), CD = conduct disorder, ODD = oppositional defiant disorder, ADHD = attention deficit/hyperactivity disorder, GAD = generalised anxiety disorder, MDD = major depressive disorder, SUD = substance use disorder. PDS = Self-rating Scale for Pubertal Development, ICU = Inventory of Callous-Unemotional Traits. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

Table 11. Additional clinical characteristics (mean (95% confidence intervals of the mean) unless stated otherwise)

Measures	CD/HCU (*n* = 113)	CD/LCU (*n* = 113)	TD (*n* = 226)	Test statistic (*p*), effect size
RPQ reactive aggression	11.74 (10.84, 12.64) [a]	11.29 (10.32, 12.26) [a]	5.48 (5.02, 5.93) [b]	F = 111.84 (< *.001*), partial η^2 = .33
RPQ proactive aggression	4.04 (3.27, 4.82) [a]	4.42 (3.52, 5.33) [a]	0.90 (0.68, 1.12) [b]	F = 154.40 (< *.001*), partial η^2 = .20
RPQ total aggression	15.79 (14.29, 17.29) [a]	15.71 (14.00, 17.42) [a]	6.37 (5.78, 6.97) [b]	F = 105.30 (< *.001*), partial η^2 = .32
GEM affective empathy	-2.80 (-6.06, 0.45) [a]	4.97 (1.76, 8.18) [b]	2.43 (0.74, 4.12) [b]	F = 7.31 (*.001*), partial η^2 = .05
GEM cognitive empathy	-1.73 (-4.10, 0.64) [a]	7.22 (4.82, 9.62) [b]	10.62 (9.30, 11.93) [c]	F = 44.95 (< *.001*), partial η^2 = .24
GEM total empathy	-1.70 (-7.82, 4.43) [a]	28.55 (23.09, 34.01) [b]	28.33 (25.28, 31.38) [c]	F = 50.69 (< *.001*), partial η^2 = .26
CBCL internalizing	66.50 (64.37, 68.63) [a]	61.79 (59.18, 64.41) [b]	50.35 (49.02, 51.68) [c]	F = 89.91 (< *.001*), partial η^2 = .34
CBCL externalizing	74.38 (73.05, 75.72) [a]	66.22 (64.27, 68.18) [b]	48.21 (47.02, 49.40) [c]	F = 367.21 (< *.001*), partial η^2 = .67
CBCL total problems	74.02 (72.23, 75.82) [a]	66.17 (64.02, 68.32) [b]	48.70 (47.39, 50.00) [c]	F = 266.36 (< *.001*), partial η^2 = .60

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. RPQ = Reactive-Proactive Aggression Questionnaire, GEM = Griffiths Empathy Measure, CBCL = Child Behaviour Checklist. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

4.3.2 VBM Whole Brain Analyses

No significant clusters were identified for any group contrasts in the main VBM models (*i.e.*, those covarying for IQ, sex, pubertal stage, site of data collection, and total intracranial volume). In the models with ADHD as a covariate, however, youths with CD exhibited reduced grey matter volume in the right rolandic operculum, extending into the insula, compared to the TD youths (x = 53, y = -3, z = 8, $Z = 4.86$, $k = 11$; $p_{(FWE\text{-}corrected)} = .02$). This difference appeared to be driven by the youths with CD/HCU, since

a much larger cluster in the same region remained significant for the contrast between the CD/HCU and TD groups (x = 53, y = -3, z = 6, $Z = 5.45$, $k = 90$; $p_{(FWE\text{-}corrected)} = .001$; see **Figure 14**).

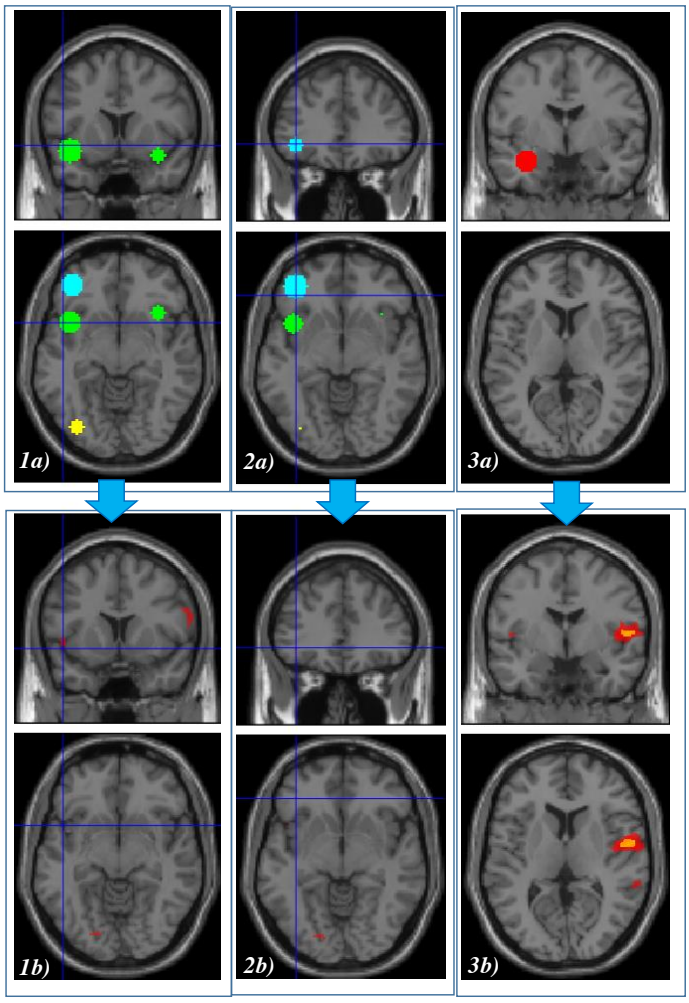


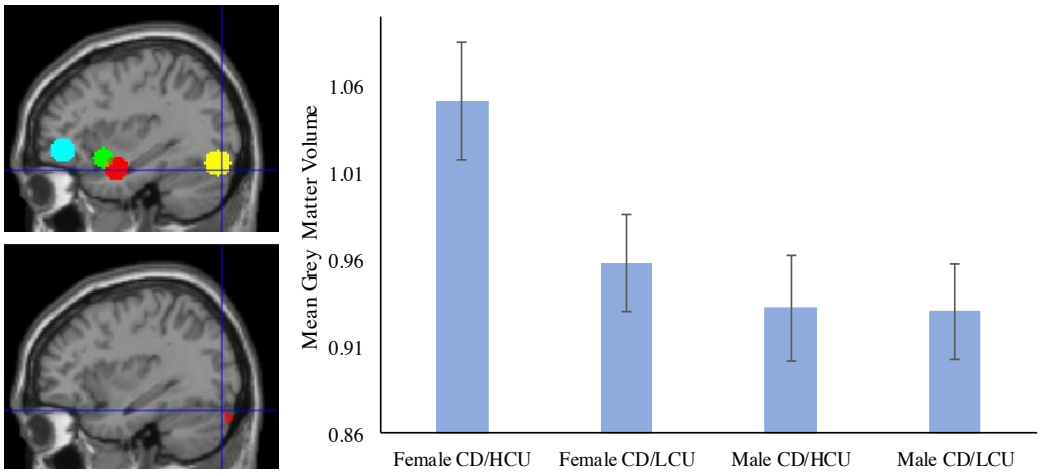
Figure 14. Regions of significantly reduced grey matter volume in the CD/HCU group relative to the TD group, when controlling for ADHD. *1a)* shows the left insula region of interest, and *1b)* shows the cluster in this region. Panels *2a)* and *2b)* show the left orbitofrontal cortex region of interest and associated cluster. Panel *3b)* shows the cluster in the right insula/operculum in the whole brain analyses (there is no associated region of interest for this cluster – *3a)*. Panels *1b) – 3b)*: yellow indicates a threshold of $p_{(FWE\text{-}corrected)} < .05$. Red indicates a threshold of $p < .001$, uncorrected

4.3.3 VBM Region of Interest Analyses

In the model controlling for ADHD, youths with CD exhibited decreased grey matter volume in the left insula relative to the TD group (x = -45, y = 12, z = -3, $Z = 3.57$, $k = 4$; $p_{(FWE\text{-}corrected)} = .02$) and a similar effect was observed for the CD/HCU group (x = -47, y = 12, z = -5, $Z = 3.64$, $k = 5$; $p_{(FWE\text{-}corrected)} = .01$). In addition, the CD/HCU group had decreased grey matter volume in the left orbitofrontal cortex relative to the TD group (x = -39, y = 36, z = -8, $Z = 3.27$, $k = 3$; $p_{(FWE\text{-}corrected)} = .04$; see **Figure 14,** panels *2a – 2b*).

4.3.4 Sex by Group Interactions

There were no significant sex by group interactions at the whole brain level, and this did not change when additionally covarying for ADHD diagnoses. However, in the region of interest analyses, there was a significant interaction in the cerebellum, falling within the left fusiform gyrus region of interest. Here, female youths with CD/HCU had greater grey matter volumes than female youths with CD/LCU, whereas in males, those with CD/LCU had greater volumes than those with CD/HCU (see **Figure 15**). This interaction was significant when covarying for IQ, pubertal stage, site of data collection, and total intracranial volume (x = -30, y = -72, z = -23, Z = 3.32, k =4; $p_{(FWE\text{-}corrected)} = .04$) and also when additionally covarying for ADHD diagnoses (x = -30, y = -72, z = -23, Z = 3.33, k =4; $p_{(FWE\text{-}corrected)} = .04$).



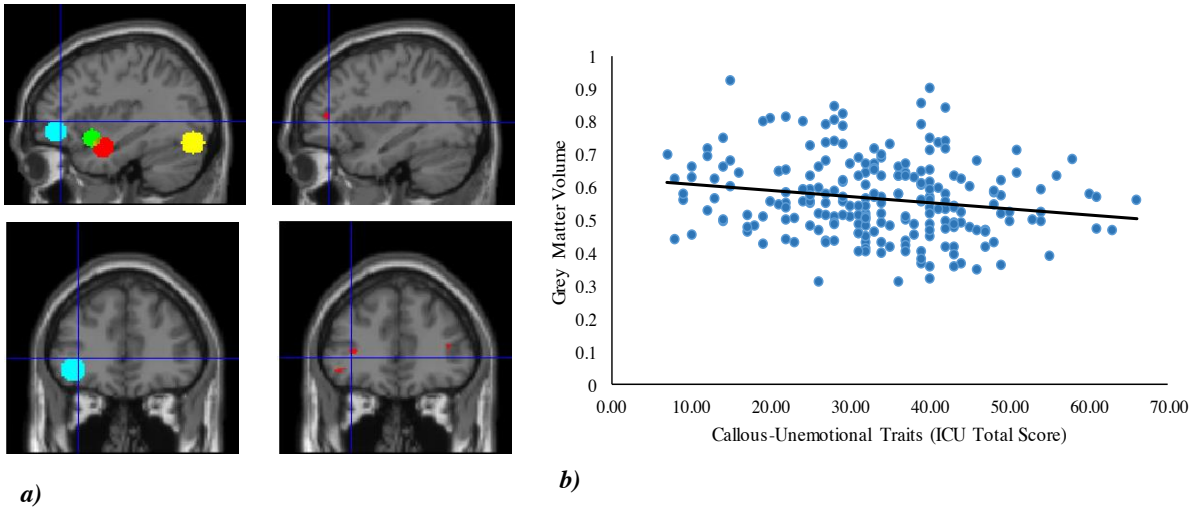
a) ***b)***

Figure 15. ***a)*** Cluster in the cerebellum (lower panel) within the left fusiform gyrus region of interest (top panel) exhibiting a significant interaction between sex and group status. Results are shown with a threshold of $p < .001$, uncorrected. ***b)*** Mean grey matter volumes per sex/group at the peak coordinate for the cluster shown in ***a)***

4.3.5 Correlations between Grey Matter Volume and CU Traits

As exploratory analyses, relationships between grey matter volume and CU traits were tested in youths with CD[17], using regression models in SPM12. Two regression models were created. The first controlled for IQ, sex, pubertal stage, site of data collection, and total intracranial volume. There were no significant relationships between grey matter volume and CU traits in the whole brain or regions of interest in this model. The second model additionally controlled for ADHD diagnoses. Here, there was a significant negative relationship in the left orbitofrontal cortex region of interest (x = -44, y = 38, z = -6, $Z = 3.32$, $k = 7$; $p_{(FWE\text{-}corrected)} = .042$; see **Figure 16**).

[17] In TD youths, there were no significant correlations between grey matter volume and CU traits. However, youths with high levels of CU traits were removed from this group, thus reducing heterogeneity and biasing the analysis.



a) ***b)***

Figure 16. ***a)*** Cluster in the left orbitofrontal cortex in which callous-unemotional (CU) traits and grey matter volume exhibited a significant negative correlation, in youths with conduct disorder (when including ADHD as a covariate). The region of interest is shown in the upper panel and the cluster in the lower panel, $p < .001$, uncorrected. ***b)*** Correlation between callous-unemotional (CU) traits and grey matter volume at the peak coordinates of the cluster shown in ***a)***

4.3.6 Mean Grey Matter Volume in Regions of Interest

Mean grey matter volumes for each region of interest (*i.e.*, the features for the Angle-GMLVQ analyses) are displayed in **Table 12**. Across all regions, volumes were highest for the CD/HCU group and lowest for the CD/LCU group, although these differences were not always statistically significant. The effect sizes of these group differences were small.

Table 12. Mean grey matter volumes in regions of interest per group (mean (95% confidence intervals of the mean))

Region of interest	CD/HCU (*n* = 113)	CD/LCU (*n* = 113)	TD (*n* = 226)	F (*p*), partial η^2
Left amygdala	0.68 (0.67, 0.69) [a *]	0.65 (0.64, 0.66) [b *]	0.66 (0.65, 0.67) [b *]	F = 8.03 (< *.001*), partial η^2 = .04
Right insula	0.76 (0.75, 0.78) [a]	0.73 (0.71, 0.74) [b]	0.75 (0.74, 0.76) [a]	F = 5.50 (*.00*), partial η^2 = .02
Left superior frontal gyrus	0.65 (0.64, 0.66) [a]	0.62 (0.61, 0.63) [b]	0.64 (0.63, 0.65) [a]	F = 5.45 (*.01*), partial η^2 = .02
Left fusiform gyrus	0.73 (0.71, 0.74) [a]	0.69 (0.68, 0.70) [b]	0.71 (0.70, 0.72) [a]	F = 7.62 (*.00*), partial η^2 = .03
Left insula	0.73 (0.71, 0.74) [a]	0.69 (0.68, 0.71) [b]	0.71 (0.70, 0.72) [a]	F = 6.89 (*.00*), partial η^2 = .03
Left orbitofrontal cortex	0.61 (0.60, 0.62) [a *]	0.59 (0.58, 0.60) [b *]	0.62 (0.61, 0.63) [a, b *]	F = 5.49 (*.00*), partial η^2 = .02

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)
* These groups no longer differed significantly after regressing out variance associated with IQ, sex, pubertal status, site of data collection and total intracranial volume. However, in the left superior frontal gyrus, the pattern changed so that the TD group had the largest mean volume after removing this variance

4.3.7 Exploratory Analysis with Whole Brain Mean Grey Matter Volume

Since this pattern of mean differences (**Table 12**) was not expected based on results from Rogers and De Brito (2016), we investigated whether the pattern was specific to our regions of interest. We compared mean group differences across the whole brain (after regressing out variance as above) by combining in one mask all anatomical regions from the Wake Forest University (WFU) PickAtlas toolbox's Automated Anatomical Labelling (AAL) atlas (Maldjian, Laurienti, Kraft, & Burdette, 2003). In line with many of our regions of interest, the CD/HCU and TD group did not differ significantly in mean grey matter volume across the whole brain (although the CD/HCU group had a slightly larger mean volume). However, the CD/LCU group had a significantly reduced volume relative to the other groups (CD/HCU mean = 0.66,

CD/LCU = 0.63, TD = 0.65. F (2, 249) = 13.76, $p < .001$. Pairwise comparisons: CD/LCU significantly smaller than other groups (ps < .05, Bonferroni corrected). The pattern of reductions in youths with CD/LCU was thus not limited to areas previously associated with CD or CU traits.

4.3.8 Angle-GMLVQ Classifier Performance

Angle-GMLVQ classifier performance is shown in **Table 13**. The HCU-LCU model achieved the lowest error rate, misclassifying on average 0.35 of each class (MCER). The Mixed-TD, HCU-TD and LCU-TD models differed significantly across all performance metrics, with the LCU-TD model achieving the second lowest MCER and the Mixed-TD the highest. In each model, TPR and TNR values were similar, indicating that the classifier correctly classified approximately the same proportion of observations in each class (the large differences in PPV and NPV values for the HCU-TD and LCU-TD models reflect the imbalanced class sizes).

4.3.9 Feature Relevance

Feature relevance scores are shown in **Figure 15**. The left amygdala was the most relevant feature in the HCU-TD model (58% high scores), where the CD/HCU group showed increased grey matter volume relative to TD youths. In the LCU-TD model, the left fusiform gyrus was most relevant (47%), where the CD/LCU group showed reduced grey matter volume relative to TD youths. In the HCU-LCU model, the left superior frontal gyrus (49%) was most relevant, where the CD/HCU group showed increased grey matter volume relative to CD/LCU. Interestingly, the left insula scored poorly in all models, despite being the only region containing a significant cluster in the VBM

analyses. However, poor model performances mean that the practical significance of these relevance scores for distinguishing between groups should be treated with caution.

Table 13. Angle-GMLVQ model performance (mean (95% confidence intervals of the mean))

	Mixed-TD	HCU-TD	LCU-TD	F (*p*), partial η^2	HCU-LCU
Accuracy	0.48 (0.48, 0.49) [a]	0.55 (0.55, 0.57) [b]	0.58 (0.57, 0.58) [c]	669.20 (<.001), .31	0.65 (0.64, 0.65)
PPV	0.48 (0.48, 0.49) [a]	0.39 (0.39, 0.39) [b]	0.40 (0.40, 0.41) [c]	718. 57 (<.001), .32	0.64 (0.64, 0.65)
NPV	0.48 (0.48, 0.49) [a]	0.72 (0.72, 0.72) [b]	0.72 (0.72, 0.72) [b]	5771.68 (<.001), .79	0.66 (0.66, 0.67)
TPR	0.47 (0.47, 0.48) [a]	0.58 (0.58, 0.59) [b]	0.54 (0.53, 0.55) [c]	286.80 (<.001), .16	0.68 (0.68, 0.69)
TNR	0.50 (0.49, 0.50) [a]	0.54 (0.53, 0.54) [b]	0.59 (0.59, 0.60) [c]	379.06 (<.001), .20	0.61 (0.61, 0.62)
MCER	0.52 (0.51, 0.52) [a]	0.44 (0.44, 0.44) [b]	0.43 (0.43, 0.44) [c]	570.91 (<.001), .28	0.35 (0.35, 0.36)

Notes: Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. PPV = positive predictive value, NPV = negative predictive value, TPR = true positive rate, TNR = true negative rate, MCER = macro-averaged classification error rate. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < 0.05$, Bonferroni corrected). Note that the HCU-LCU model (column 6) was not included in statistical tests as comparisons between this and other models were not relevant to hypotheses

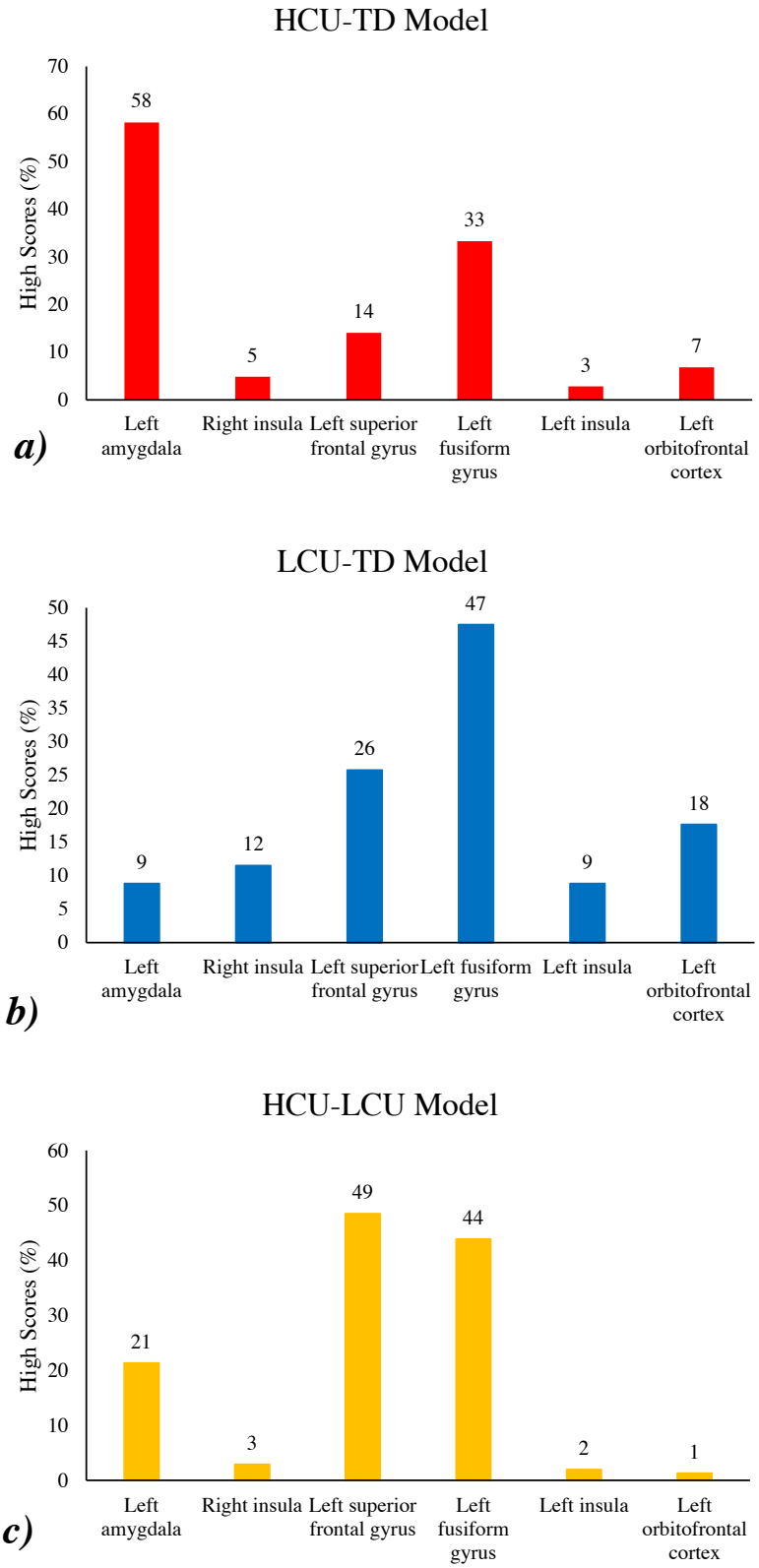


Figure 17. Feature relevance for ***a)*** HCU-TD model, ***b)*** LCU-TD model and ***c)*** HCU-LCU models. Bars show percentage of re-samplings in which feature relevance was in the top 20% of relevance scores across all re-samplings with MCER ≤ 0.40

4.3.10 SVM Classifiers

SVM classifier performance is shown in **Table 14**. For the Mixed-TD model, the best SVM classifier was a non-linear model using a fifth-order polynomial transformation kernel. This classifier achieved an MCER of 0.49, which was significantly better than for the corresponding Angle-GMLVQ model ($t_{(1998)} = 9.35, p < .001$). For the other models, the linear SVM was the highest performing SVM classifier. MCERs from linear SVMs were 0.45 for the HCU-TD model, 0.44 for the LCU-TD model and 0.36 for the HCU-LCU model, which were all significantly worse than the corresponding Angle-GMLVQ models (independent samples t-test, all $p < .05$).

Table 14. Performance (MCER) for linear and non-linear SVM models (mean (95% confidence intervals of the mean))

	Mixed-TD	**HCU-TD**	**LCU-TD**	**HCU-LCU**
Linear	0.52 (0.52, 0.52) [a]	0.45 (0.45, 0.46) [a]	0.44 (0.44, 0.44) [a]	0.36 (0.36, 0.37) [a]
2nd order polynomial	0.53 (0.52, 0.53) [a]	0.50 (0.49, 0.50) [b, d]	0.46 (0.45, 0.46) [b]	0.39 (0.39, 0.40) [b]
3rd order polynomial	0.51 (0.50, 0.51) [b]	0.49 (0.49, 0.50) [b]	0.50 (0.49, 0.50) [c]	0.41 (0.41, 0.42) [c]
4th order polynomial	0.51 (0.50, 0.51) [b]	0.51 (0.51, 0.51) [c]	0.49 (0.49, 0.50) [c]	0.43 (0.42, 0.43) [d]
5th order polynomial	0.49 (0.49, 0.50) [c]	0.50 (0.49, 0.50) [b, d]	0.47 (0.46, 0.47) [d]	0.40 (0.40, 0.41) [c]
6th order polynomial	0.51 (0.50, 0.51) [b]	0.50 (0.50, 0.51) [c, d]	0.48 (0.48, 0.49) [e]	0.43 (0.42, 0.43) [d]
F (*p*)	47.50 (*<.001*)	106.44 (*<.001*)	133.67 (*<.001*)	103.70 (*<.001*)

Notes: MCER = macro-averaged classification error rate. Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. Within each column, models with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

4.4 Discussion

Our first question was whether grey matter volume differs in youths with CD compared to TD youths. We hypothesised reductions in the CD/mixed group relative to the TD group. This hypothesis was partially supported; when controlling for ADHD diagnoses, the CD/mixed group exhibited reductions in grey matter volume relative to the TD youths in a region extending from the right operculum into the insula (whole brain analysis), and in the left insula (region of interest analysis). Our second question was whether there were differences between CD/HCU, CD/LCU and TD groups. The CD/HCU group exhibited reductions in the left insula and orbitofrontal cortex relative to TD youths when controlling for ADHD. The CD/LCU youths, by contrast, did not exhibit significant clusters of reductions in the VBM analyses. However, in a comparison of mean grey matter volumes in each region of interest, youths with CD/LCU were characterised by a fairly consistent pattern of reduced volume across multiple areas. Counterintuitively, given the VBM results, youths with CD/HCU were characterised by slight increases in regional mean grey matter volumes relative to other groups. An exploratory analysis using mean grey matter volume across the whole brain confirmed that these differences were not specific to our regions of interest.

Our final question concerned the extent to which differences in grey matter volume were informative when distinguishing between individual youths. The groups that differed most in mean regional volumes – CD/HCU and CD/LCU – were distinguished with an MCER of 0.35 (corresponding to an accuracy of 0.65). Indeed, when these two groups were combined in the CD/mixed group, they were distinguished from TD youths at slightly below chance level. These performances supported our second hypothesis

that youths with CD/HCU and CD/LCU are distinctive in terms of grey matter volume patterns.

4.4.1 Reductions in Insula Grey Matter Volume

The anterior insula is a key region for interoception and emotional self-awareness. It has been implicated in a range of processes related to the integration of bodily states (related to emotion) and cognition, including motivated behaviour and the prioritisation of salient stimuli for cognitive processing (Namkung *et al*., 2017; Craig, 2009). Impairments in these processes have been linked to psychopathic traits, via a diminished ability to use others' emotional responses to guide future behaviour (Blair, 2013). As previously noted, reductions in anterior insula grey matter volume are commonly observed in CD (Raschle *et al*., 2015; Noordermeer *et al*., 2016; Rogers & De Brito, 2016). We observed significant reductions in our region of interest in the left central/anterior insula. Whole brain analyses also revealed a larger area of reduction in the right hemisphere with a peak in the rolandic operculum, extending into the posterior insula. A recent machine learning study reported grey matter volume reductions in the posterior insula in youths with CD (Zhang *et al*., 2018b). According to Craig (2009), the posterior insula responds to interoceptive signals (*e.g*., thirst, warmth, *etc*.), which are then integrated with signals from other cortical and sub-cortical regions in the anterior insula when rating these stimuli subjectively. Furthermore, Craig (2009) notes that activation in the anterior insula often extends to the frontal operculum, and there is some evidence that the rolandic operculum is also involved in the processing of combined interoceptive and exteroceptive signals (Blefari, Martuzzi, Salomon, Bello-Ruiz, Herbelin, Serino, & Blanke, 2017). Our findings, together with these previous

studies, suggest that the operculum and posterior insula might be implicated in CD as well as the anterior insula. They are thus candidate regions of interest for future studies.

4.4.2 Reductions in Orbitofrontal Cortex Grey Matter Volume in CD/HCU

Alongside the insula, a smaller region of reduced grey matter volume was also observed in the left orbitofrontal cortex in youths with CD/HCU relative to TD youths. This finding is consistent with the only study to date to compare CD/HCU, CD/LCU and TD groups directly (Sebastian *et al*., 2016). However, the one other study to compare youths with CD/HCU to TD youths reported increases in grey matter concentration in CD/HCU in a more medial and inferior right-hemisphere orbitofrontal cortex region (De Brito *et al*., 2009). These discrepancies highlight that neighbouring sub-regions might be differently implicated in CD/HCU. In psychopathy, the orbitofrontal cortex is specifically implicated in risk for reactive aggression due to dysfunctional regulation of threat responsivity (*e.g*., Blair, 2004; Blair *et al*., 2005). However, there is some evidence for differing functions of the lateral and medial portions of the orbitofrontal cortex. While the medial portion has been linked to the monitoring of reward values of different stimuli, the lateral portion has been linked to the evaluation of punishing stimuli and contingent behaviour changes (Kringelbach & Rolls, 2004; Elliott, Dolan, & Frith, 2000). It is interesting that the reductions observed in the current sample were in areas linked to behavioural responses to punishment, since punishment insensitivity is a risk factor for the development of CD/HCU (Barker *et al*., 2011).

4.4.3 Direction of Associations with CU Traits

In our sample, the reduction in insula grey matter volume appeared to be driven by the CD/HCU group. In previous studies, the direction of association between grey matter volume and CU traits was often negative, though inconsistent (*e.g.*, Fairchild *et al.*, 2013; Cohn *et al.*, 2016). Interestingly, two of the studies reporting positive associations between CU traits and grey matter volume used a TD sample (Raschle *et al.*, 2018) or found a positive association in youths with below-median CD scores only (Cohn *et al.*, 2016; but *c.f.* Fairchild *et al.*, 2013). Although there were no positive associations between CU traits and grey matter volume in the current sample, and no associations in the insula, a negative association in the left orbitofrontal cortex was observed in youths with CD[18]. It is difficult to account for these inconsistencies in the literature. One possibility is that the relationship between CU traits and grey matter volume might differ depending on conduct problem severity (see Cohn *et al.*, 2016). Unfortunately, the wide variety of measures used to assess CD and CU traits means that the severity of different samples cannot easily be compared. The relationship between conduct problems, CU traits and grey matter volume thus remains a question for future research.

4.4.4 Comorbid ADHD and Grey Matter Volume Reductions

Given the level of comorbidity in CD (see **Table 10**), it should be noted that reductions in grey matter volume in both the CD/mixed and CD/HCU groups were significant only when controlling for ADHD. Since group differences became stronger when controlling for ADHD, it appears that the observed reductions are more strongly related to CD and

[18] There was no significant association in the TD group. However, as previously noted, TD youths with high levels of CU traits were excluded from the sample, thus biasing the TD group towards reduced heterogeneity relative to the CD group. We therefore do not report the correlations for the TD group.

CD/HCU than ADHD. However, given the high rates of comorbid ADHD in CD, controlling for the disorder might lead to results that are less reflective of clinical reality. Indeed, controlling for ADHD did slightly reduce classifier performance for the CD/HCU group (see Appendix B).

4.4.5 Decreases in Mean Grey Matter Volume in CD/LCU

While localised reductions in grey matter volume were expected, the pattern of group differences in regional mean grey matter volumes (*i.e.*, those that were used as features in the classification analyses) was unexpected. Specifically, it was the CD/LCU group, rather than the CD/HCU group, who exhibited consistent decreases in mean grey matter volume relative to other groups. Although less often significant, the CD/HCU group even had slight increases in mean grey matter volume relative to the TD and CD/LCU groups. Importantly, these differences were not confined to regions previously associated with CD or CU traits. One potential explanation for these widespread differences is maturity or 'brain age'. Longitudinal MRI data reveal a region-specific inverted 'U' trajectory of grey matter volume across childhood and adolescence. Cortical grey matter volume peaks between the ages of 12-20 years and declines thereafter, with frontal and parietal lobes peaking earliest, followed by the temporal and then occipital lobes (Giedd *et al.*, 1999; Gogtay *et al.*, 2004). There is some preliminary evidence that the characteristic decline in early adolescence might be delayed or altered in boys with CD/HCU (De Brito *et al.*, 2009), although this has not been tested systematically across adolescence. Since our participants were all within the period of peaking and declining grey matter volume, a delay in cortical maturity could theoretically contribute to the slightly elevated mean grey matter volumes in CD/HCU.

Meanwhile, youths with CD/LCU exhibited mean reductions in grey matter volume that are generally more characteristic of CD. Longitudinal data would be needed for investigating these hypothesised explanations.

4.4.6 Distinguishing between Individuals using Grey Matter Volume Differences

First, it is important to note that our *a-priori* regions of interest did not capture much of the local reductions in grey matter volume that distinguished groups in the VBM analyses. This might explain why performance for the HCU-TD model in particular was quite poor. Nonetheless, in the HCU-LCU model, 65% of youths were correctly classified as CD/HCU or CD/LCU based on measures of grey matter volume (*i.e.*, 35% MCER and 65% overall accuracy). This highlights the opposing pattern of mean grey matter volumes in CD/HCU and CD/LCU, with the TD youths lying intermediate to the CD subtypes (albeit not always significantly different from the CD/HCU group). Interestingly, a similar pattern was reported in an fMRI study investigating amygdala reactivity to subliminally presented fearful faces; youths with CD/HCU were hypo-reactive relative to TD youths, while those with CD/LCU were hyper-reactive (Viding *et al.*, 2012). Our data provide additional support for the view that youths with CD/HCU and CD/LCU exhibit different, and in some respects opposite, neurological characteristics. In fact, this pattern indicates that a single 'CD' diagnosis in this context is not merely uninformative; it is actively unhelpful, increasing rather than decreasing the overlap between CD and TD.

Feature relevance scores indicated that the left superior frontal gyrus was the most relevant region in the HCU-LCU model. This was closely followed by the left fusiform gyrus, which was also the most relevant region in the LCU-TD model. Although these

regions are not classically associated with psychopathy or antisocial behaviour (*e.g.*, Poeppl *et al.*, 2019), structural differences have emerged in more than one meta-analysis of youths with CD (Rogers & De Brito, 2016; Noordermeer *et al.*, 2016). The left fusiform gyrus has also been implicated in 'hot' executive functioning in CD (Noordermeer *et al.*, 2016), and psychopathy in adults has been associated with greater grey matter volume in the right fusiform gyrus (Aoki, Inokuchi, Nakao, & Yamasue, 2013). Although less relevant than the left superior frontal and fusiform gyri, the left amygdala was also one of the more relevant features in the HCU-LCU model (with 21% high scores). Interestingly, in the HCU-TD model, the left amygdala was the most relevant region. Amygdala hypo-reactivity to distress cues has been repeatedly linked to CU traits, and amygdala abnormalities are central to neurocognitive models of psychopathy and CU traits (*e.g.*, Blair, 2013). However, it is noteworthy that mean amygdala grey matter volume was increased in CD/HCU relative to TD in our sample, reflecting the pattern of slight increases that was seen across the brain.

It is noticeable that none of the high-performing regions contained significant clusters in the VBM analyses. Indeed, the regions of interest that contained, or were situated close to, significant clusters in the VBM analyses were consistently low in relevance (*i.e.*, insulae and left orbitofrontal cortex). Presumably, this reflects a 'cancelling out' effect whereby the mean group differences in a region were attenuated by localised reversals in the pattern. This loss of spatial acuity highlights the need for a more data-driven approach to classification in future studies, *e.g.*, a searchlight analysis, using more fine-grained, spatially sensitive measures of grey matter volume (*e.g.*, Zhang *et al.*, 2018a). Equally, the use of the current regions of interest revealed a pattern of subtle,

widespread group differences that is in itself interesting and was not detected with traditional analysis methods.

4.4.7 Strengths and Limitations

Our study has several strengths. First, the sample was large and well characterised, with all participants having been assessed for CD and other disorders using both self- and parent-report semi-structured clinical interviews. The sample was also international, having been recruited from multiple European centres, and included males and females across a wide age range. Aside from the intentional oversampling of females, we can thus have reasonable confidence that the sample is representative of European youths with CD. However, this study is not without limitations. First, although we controlled for sex and pubertal status by matching groups and regressing out the variance associated with these variables, and explored group by sex interactions, ideally separate analyses should also have been conducted with males and females within different age groups. This would have improved our understanding of how CD/HCU and CD/LCU present in different youth demographics. Finally, as previously mentioned, the *a priori* regions of interest did not capture the regions with the largest group differences. Further analyses using a more data-driven approach would indicate whether the discriminative power of the classifier would improve with more, and smaller, regions of interest. Such an approach could incorporate measures such as surface area and cortical folding, which are associated with different developmental trajectories (Raznahan *et al*., 2011) and of which volume is a composite. This would provide a richer understanding of grey matter differences in youths with CD/HCU and CD/LCU.

4.4.8 Summary and Conclusions

In summary, our analyses yielded three main findings. First, youths with CD, and especially CD/HCU, exhibited grey matter volume reductions in the left anterior insula, right operculum extending into the insula, and left orbitofrontal cortex (CD/HCU only). Second, youths with CD/LCU were characterised by small but consistent decreases in mean grey matter volume, relative to TD youths, in multiple regions across the brain. Conversely, youths with CD/HCU exhibited very slight (usually non-significant) increases in mean grey matter volume. Third, classifier performance dropped below chance when grouping youths with CD/HCU and CD/LCU together (in the CD/mixed group), and was highest when distinguishing CD/HCU from CD/LCU directly. This pattern of performance indicates that in terms of mean grey matter volumes, youths with CD/HCU and CD/LCU are not only neurologically distinctive from each other, but in fact differ from each other more than from TD youths. Replicating these subtle, widespread group differences, and understanding how they relate to the commonly observed reductions in grey matter in CD and CD/HCU, will be important topics for future research.

CHAPTER 5: RECOGNITION OF FACIAL EXPRESSIONS OF EMOTION IN CONDUCT DISORDER WITH HIGH VERSUS LOW LEVELS OF CALLOUS-UNEMOTIONAL TRAITS

5.1 Introduction

CD has frequently been associated with a reduced ability to recognise facial expressions of emotion (*e.g.*, Bowen *et al.*, 2014; Sully, Sonuga-Barke, & Fairchild, 2015). However, it is not clear whether impoverished emotion recognition abilities are a consistent feature of CD, or whether only certain youths are affected (*e.g.*, Pajer, Leininger, & Gardner, 2010; Rehder *et al.*, 2017). Of particular interest is the extent to which youths with CD/HCU and youths with CD/LCU differ from each other and from TD youths on facial emotion recognition abilities.

5.1.1 Evidence for Selective Difficulties with Negative Emotions in CD/HCU

A number of studies point towards selective difficulties in identifying negative emotions in CD/HCU. Relative to children with behavioural problems alone (*i.e.*, CD/LCU), Blair, Colledge, Murray and Mitchell (2001) found that children with behavioural problems and high levels of psychopathic traits were slower to recognise sadness in a gradually morphing neutral-to-sad face. They also made more errors in recognising sadness and fear relative to those with behavioural problems alone. Likewise, in a small sample (N = 18), Stevens, Charman and Blair (2001) tested children with CD/HCU and CD/LCU on recognition of sadness, fear, happiness and anger. Children with CD/HCU underperformed relative to those with CD/LCU on sadness and fear only. Fairchild, van Goozen, Calder, Stollery and Goodyer (2009) found significantly poorer fear, sadness

and surprise recognition in youths high in psychopathic traits, relative to those with low levels of psychopathic traits (youths with early-onset CD were also less able to identify happiness, anger and disgust). Likewise, Fairchild, Stobbe, van Goozen, Calder and Goodyer (2010) found poorer recognition of sadness in female youths with CD and high psychopathic traits, relative to those with CD and low levels of psychopathic traits. Indeed, youths with CD as a whole were poor at recognising negative emotions (anger and disgust) compared to TD youths. However, Schwenck *et al*. (2012) reported no differences between CD/HCU and CD/LCU groups in their speed or accuracy in identifying any emotions in a morphing face task.

Other studies have used dimensional measures of psychopathy (*i.e*., CU and impulsive-antisocial traits). For example, in a community sample of adolescents, Blair and Coles (2000) demonstrated a negative correlation between psychopathic traits and identification of sadness, fear and anger. When the two sub-factors of psychopathy were analysed separately, negative correlations were observed between CU traits and sadness and fear recognition. Impulsive-antisocial behaviour, meanwhile, was negatively correlated with fear recognition only. There were no associations between psychopathic traits and happiness, disgust or surprise. In young offenders, Bowen *et al*. (2014) reported a reduced ability to recognise sadness, mild disgust and intense fear relative to TD controls. Furthermore, sadness and disgust recognition difficulties were related to CD symptoms and psychopathic trait severity, while anger recognition was related to offence severity.

The prominence of fear and sadness recognition difficulties in CD/HCU is notable. A neurocognitive model of psychopathy (to which CD/HCU is thought to be a developmental precursor) links such difficulties to amygdala dysfunction. According to

Blair (*e.g.*, Blair, 2003; Blair *et al.*, 2005), expressions of fear and sadness (and perhaps also happiness) act as reinforcers, decreasing (or increasing) the probability that an associated behaviour will be repeated. Blair (2003) argues that due to amygdala dysfunction, fearful and sad expressions are neither salient nor aversive to individuals with psychopathy. Consequently, these individuals do not learn to avoid behaviours that frighten or upset others, and indeed willingly cause harm for instrumental gain. Abnormal responses to angry expressions have also been hypothesized to occur in youths with CD, due to orbitofrontal cortex dysfunction (Blair, 2003; Blair *et al.*, 2005), but it is not clear how this relates to anger recognition *per se*.

5.1.2 Evidence for Non-Specific Difficulties or Normal Abilities in CD/HCU

However, emotion recognition difficulties in CD/HCU are neither universally reported nor, when they are reported, consistently limited to negative emotions such as fear or sadness. Rehder *et al.* (2017) found that European American second graders (*i.e.*, 7-8-year-olds) with conduct problems, regardless of CU trait levels, underperformed relative to their typically developing peers on overall emotion recognition and happiness recognition in particular. Furthermore, African American children's performance was similar for those with and without conduct problems or elevated CU traits. Likewise, in an all-female adolescent sample, Pajer *et al.* (2010) found no evidence for emotion recognition difficulties in CD. Other studies also report difficulties with both positive and negative emotions in CD. Relative to TD controls, Sully *et al.* (2015) reported reduced ability to identify anger, fear, happiness, sadness and surprise in youths with CD. In a meta-analysis that included several of the above studies, Dawel, O'Kearney, McKone and Palermo (2012) concluded that emotion recognition difficulties in

adolescent psychopathy are pervasive rather than limited to specific emotions. However, the strongest effects were nonetheless found for fear, then anger and sadness. More recently, in a study from the FemNAT-CD project, Kohls *et al*. (2019) reported emotion recognition difficulties for all six basic emotions, in both male and female adolescents with CD. There are several factors that ought to be considered when noting discrepant results between studies. These include the use of different measures of CU traits, including some self-report measures (*e.g*., Fairchild *et al*., 2009), the focus on CU traits versus broader psychopathic traits, different types of sample (*e.g*., community versus clinic-referred), and the use of CD/LCU versus TD as the control group for CD/HCU. These factors are all likely to contribute to the discrepancies in findings.

5.1.3 The Role of Comorbidities

Finally, it should be noted that the presence of comorbidities is a complicating factor. Some studies point to emotion recognition difficulties that are specific to 'pure' CD. For example, Short, Sonuga-Barke, Adams and Fairchild (2016) found that while CD was associated with poorer emotion recognition ability, youths with CD and comorbid anxiety performed similarly to TD controls. This suggests that anxiety might 'offset' the difficulties associated with CD. Likewise, Schepman, Taylor, Collishaw and Fombonne (2012) reported emotion recognition difficulties in youths with CD alone relative to TD youths, but not in youths with comorbid CD and depression. Conversely, severe mood disorders have been associated with poorer emotion recognition than in CD (Guyer *et al*., 2007).

5.1.4 Summary and Hypotheses

In summary, youths with CD often perform more poorly than TD youths when identifying facial expressions of emotion (Bowen *et al*., 2014; Sully *et al*., 2015; *c.f.* Rehder *et al*., 2017 (African Americans); Pajer *et al*., 2010). Furthermore, recognition of negative emotions has been reported to be especially poor in youths with CD/HCU relative to those with CD/LCU (*e.g*., Blair *et al*., 2001; Stevens *et al*., 2001), although again, this is not consistent, with some studies reporting more general difficulties (*e.g*., Rehder *et al*., 2017 (European Americans); Dawel *et al*., 2012). There is also a relative lack of studies comparing CD/HCU, CD/LCU and TD groups (see Schwenck *et al*., 2012; Rehder *et al*., 2017). Overall, then, youths with CD/HCU typically exhibit more severe emotion recognition difficulties than youths with CD/LCU. However, it is not clear whether these differences are exaggerated for negative emotions, or occur more generally across all emotions.

In the present study, we compared facial emotion recognition abilities in youths with CD/HCU, youths with CD/LCU and TD youths. First, we used traditional univariate statistical techniques to test for group level differences in recognition of anger, disgust, fear, happiness, sadness and surprise. Based on previous experimental evidence (especially Kohls *et al*. (2019), given the overlap with the current sample), we hypothesised that performance would be poorer in CD than for TD youths. More specifically, we hypothesised that performance would be poorest for youths with CD/HCU, followed by youths with CD/LCU and then TD youths. We also explored interactions between diagnostic group and emotion, in order to investigate whether youths with CD/HCU have specific difficulties with negative emotions (*i.e*., over and above any difficulties with recognising emotions generally). Finally, we explored sex

differences in these interactions, and group differences in error type (*i.e.*, close versus distant emotions).

Second, we used Angle-GMLVQ (Bunte *et al.*, 2016) to quantify the extent to which differences in emotion recognition abilities could determine the diagnosis of individual youths, *i.e.*, whether they belonged to the CD/HCU, CD/LCU or TD groups. Since Angle-GMLVQ classifies on the basis of relative differences between features for each individual – in this case relative differences in recognition for each emotion – it should perform well if group differences in emotion recognition are more exaggerated for negative emotions, and less well if differences are similar across emotions. As previously, we also constructed a classifier for a mixed-CU CD group against the TD group (*i.e.*, Mixed-TD model), to verify whether performance was improved by distinguishing between CD/HCU and CD/LCU. Finally, we compared Angle-GMLVQ performance to an SVM classifier, to check whether performance would improve or deteriorate with a non-angle-based classifier. We predicted that if youths with CD/HCU do indeed have specific difficulties with negative emotions, then classifier performance should be better for the CD/HCU against TD model (HCU-TD) and CD/LCU against TD model (LCU-TD) than for the mixed-CU against TD (Mixed-TD) model. The CD/HCU against CD/LCU model (HCU-LCU) should also perform above chance level. Finally, in line with our expected group differences, we predicted that negative emotions would be more relevant to accurate classification than other emotions for the HCU-TD classifier.

5.2 Methods

5.2.1 Participants

Participants were selected from a subset of the FemNAT-CD sample who were included in the sample from Kohls *et al*. (2019; N = 1252). These participants had data from the full set of FemNAT-CD behavioural tasks, including the emotion recognition task reported here. From this subset, we excluded five participants who appeared to be ineligible for FemNAT-CD[19], and 16 who did not complete the measure of CU traits. We then separated participants with CD into CD/HCU and CD/LCU groups using a tertile split of CU scores (*i.e*., ICU total score) for the CD group. We excluded participants with CD who had CU scores in the second tertile (scores between 32 and 39; n = 147), as well as TD participants with CU scores in the first tertile (n = 7) or second tertile (n = 33). Finally, we excluded 120 participants while matching CD/HCU, CD/LCU and TD groups for age, pubertal status, proportion of females, and site of data collection. Group matching was conducted with Match software (van Casteren & Davis, 2007). The final sample size was 924 (192 CD/HCU, 183 CD/LCU and 549 TD). A mixed CU group ('CD/mixed') was also formed by combining the CD/HCU and CD/LCU groups (n = 375). The final sample did not differ from excluded participants on age or IQ, and participants with CD who were included did not differ on CD symptoms or CU scores from excluded youths with CD (t-tests, two-tailed, all $p > .05$). The group of excluded participants did contain proportionately fewer males than the

[19] These participants were initially recruited as cases, but later appeared not to meet the diagnostic criteria for this group.

final sample (26% versus 40%; $\chi^2 = 20.66$, $p < .001$), reflecting the deliberate over-sampling of females in the full sample.

5.2.2 Questionnaire and Interview Measures

The K-SADS-PL (Kaufman *et al*., 1997) was used to assess for CD and other disorders, the Wechsler Intelligence Scales (Wechsler, 1999; Wechsler, 2008) were used to estimate IQ, and the PDS (Petersen *et al*., 1988) was used to measure pubertal development. The ICU (Essau *et al*., 2006b) was used as the measure of CU traits. Reliability was good in the current sample (Cronbach's alphas: callous $\alpha = 0.88$, uncaring $\alpha = 0.89$, unemotional $\alpha = 0.79$, total $\alpha = 0.93$). We also report scores from the RPQ (Raine *et al*., 2006), GEM (Dadds *et al*., 2008) and the CBCL internalising, externalising and total problems scales (Achenbach, 1991). RPQ and GEM subscale reliabilities were good; RPQ proactive $\alpha = 0.87$, RPQ reactive $\alpha = 0.88$, RPQ total $\alpha = 0.92$; GEM affective $\alpha = 0.81$, GEM cognitive $\alpha = 0.74$, GEM total $\alpha = 0.86$. These measures are described in full in Chapter 2 (sections 2.4.6 – 2.4.7).

5.2.3 Hexagon Facial Emotion Recognition Task

The Hexagon task was developed by Calder (1996) as a test of facial emotion recognition deficits in patients with amygdala damage. Stimuli consist of 'blended' faces, each displaying combinations of two of the six basic emotions (anger, disgust, fear, happiness, sadness and surprise; Ekman & Friesen, 1976). To create the stimuli, Calder placed the six expressions in sequence so that each expression had as neighbours the expressions with which it was most easily confused. Anger and happiness were then placed next to each other to create a hexagonal arrangement. Each expression was

morphed into the neighbouring expression, generating blended expressions in intensity ratios of 90:10, 70:30, 50:50, 30:70 and 10:90 (see **Figure 18**).

Faces were presented on a computer monitor in random order for five seconds each. After each face, participants selected from a list of the six emotions the label that best described the emotion presented. There was no time limit for responding and no feedback was provided. Participants completed a practice block followed by five experimental blocks. Each of the 30 blended expressions was presented once per block. Correct responses were coded as those where the dominant emotion was selected (50:50 expressions were not scored). Percentage recognition accuracy was then calculated for each emotion.

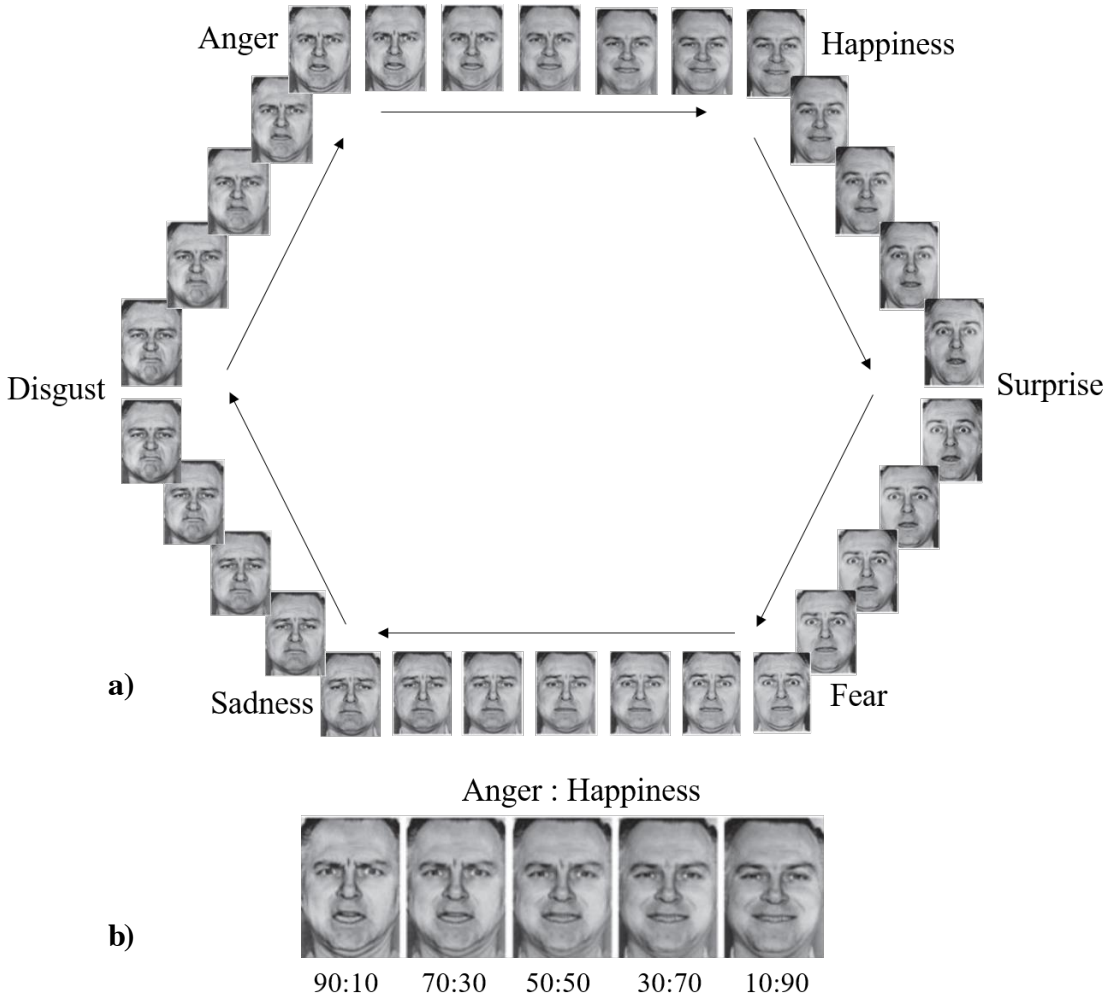


Figure 18. ***a)*** The complete set of blended expressions arranged in a hexagon. The six basic emotions (anger, happiness, surprise, fear, sadness and disgust) lie on the vertices adjacent to their most easily confused emotion. The faces on the edges of the hexagon are the blended expressions used as task stimuli. ***b)*** An example of the blended expressions for the anger-to-happiness continuum, with anger: happiness ratios labelled

5.2.4 Univariate Analyses

Interactions between group, emotion, intensity (*i.e.*, 70% versus 90% expression intensity) and sex were investigated using a repeated measures ANCOVA (using IBM SPSS Statistics 25), with group as a between-subjects factor and emotion and intensity as within-subjects factors. Site of data collection was included as an additional factor of no interest, and grand-mean-centred IQ and pubertal status were entered as covariates.

To investigate whether comorbid psychiatric disorders were contributing to any observed group differences, the same analysis was repeated on residuals after regressing out variance associated with comorbid ADHD, generalised anxiety disorder, depression and substance use disorder. As exploratory analyses, we investigated sex differences, and finally, we explored group differences in the type of error made (*i.e.*, confusion with neighbouring versus distant emotions).

5.2.5 Classification Analyses

5.2.5.1 Features

Mean recognition accuracies for each of the 12 emotion × intensity trial types were calculated for each participant (*i.e.*, 12 features). For the final features, standardised residuals were calculated for each trial type by regressing out variance associated with IQ, sex, pubertal status and site of data collection.

5.2.5.2 Classification Models

Data were classified using Angle-GMLVQ classifiers. Additional analyses were conducted with SVM classifiers. The classifiers were run separately for each pair of groups of interest:

1. CD/mixed against TD ('Mixed-TD')
2. CD/HCU against TD ('HCU-TD')
3. CD/LCU against TD ('LCU-TD')
4. CD/HCU against CD/LCU ('HCU-LCU')

5.2.5.3 Training and Testing Procedure

The Angle-GMLVQ classifier was trained and tested for each model using a holdout design, with 80% training data and 20% testing data. This process was repeated for

1000 re-samplings to ensure stability of performance. In each re-sampling, the larger class was down-sampled at random to the size of the smaller class. Mean performance measures were then calculated for each model.

5.2.5.4 Assessment of Model Performance

Models were assessed using overall classification accuracy, PPV, NPV, TPR, TNR and MCER. These measures are described in Chapter 2 (section 2.2.5).

5.2.5.5 Assessment of Feature Relevance

Normalised relevance scores were considered 'high' if they were in the top 20% of scores across all re-samplings with a corresponding MCER of 0.40 or below. Features were ranked by number of high scores.

5.2.5.6 SVM Classifiers

SVM classifiers were trained in MATLAB (R2016a) using the 'fitcsvm' function. Five SVM classifiers were trained and tested for each model; linear SVM and SVMs with second, third, fourth and fifth-order non-linear polynomial transformation kernels. The best performing SVM (*i.e.*, lowest MCER) was then selected for comparison with the corresponding Angle-GMLVQ classifier.

5.3 Results

5.3.1 Sample Characteristics

There were no significant differences between groups in site of data collection ($\chi^2 = 27.97$, $p = .06$, $\varphi = .17$, nor on other matching criteria (see **Table 15**). The CD/HCU and CD/LCU groups did not differ significantly on IQ. The CD/HCU group had the more severe presentation in terms of CD symptoms, CU traits and ODD and ADHD

symptoms (**Table 15**). They also exhibited lower cognitive and affective empathy and more externalising problems than the CD/LCU group, but no differences in aggression as measured by the RPQ (see **Table 16**). More surprisingly, youths with CD/HCU exhibited a greater number of internalising problems than youths with CD/LCU (although this finding is not unprecedented; see Sebastian *et al*., 2014).

Table 15. Demographic and clinical characteristics (mean (95% confidence intervals of the mean) unless stated otherwise)

Measures	CD/HCU (*n* = 192)	CD/LCU (*n* = 183)	TD (*n* = 549)	Test statistic (*p*), effect size
Age	14.24 (13.93, 14.55) [a]	14.49 (14.15, 14.83) [a]	14.03 (13.83, 14.24) [a]	F = 2.64 (*.07*), partial η^2 = .01
Females (%)	57	56	62	χ^2 = 2.98 (*.23*), φ= .06
PDS pubertal stage	3.62 (3.49, 3.75) [a]	3.67 (3.52, 3.81) [a]	3.52 (3.44, 3.61) [a]	F = 1.57 (*.21*), partial η^2 = .00
Performance IQ	97.24 (95.06, 99.42) [a]	96.69 (94.57, 98.81) [a]	103.18 (102.01, 104.36) [b]	F = 20.74 (< *.001*), partial η^2 = .04
Verbal IQ	92.94 (90.96, 94.91) [a]	91.92 (89.80, 94.03) [a]	103.17 (101.85, 104.48) [b]	F = 56.67 (< *.001*), partial η^2 = .11
Total IQ	95.47 (93.72, 97.23) [a]	94.61 (92.81, 96.41) [a]	103.49 (102.47, 104.51) [b]	F = 53.36 (< *.001*), partial η^2 = .10
ICU callous	18.86 (18.21, 19.52) [a]	7.30 (6.79, 7.81) [b]	3.89 (3.69, 4.09) [c]	F = 1560.72 (< *.001*), partial η^2 = .77
ICU uncaring	18.93 (18.54, 19.31) [a]	11.09 (10.52, 11.65) [b]	7.31 (7.00, 7.62) [c]	F = 759.48 (< *.001*), partial η^2 = .62
ICU unemotional	9.79 (9.36, 10.22) [a]	5.52 (5.14, 5.91) [b]	4.55 (4.33, 4.77) [c]	F = 262.21 (< *.001*), partial η^2 = .36
ICU total	47.58 (46.65, 48.51) [a]	23.91 (23.08, 24.75) [b]	15.75 (15.21, 16.30) [c]	F = 1796.21 (< *.001*), partial η^2 = .80
K-SADS CD symptoms	6.10 (5.77, 6.43) [a]	5.58 (5.22, 5.94) [b]	0.09 (0.06, 0.11) [c]	F = 1584.99 (< *.001*), partial η^2 = .78
K-SADS ODD symptoms	5.68 (5.42, 5.94) [a]	4.88 (4.57, 5.19) [b]	0.06 (0.03, 0.09) [c]	F = 1334.44 (< *.001*), partial η^2 = .80
K-SADS ADHD symptoms	7.91 (6.91, 8.91) [a]	6.13 (5.17, 7.08) [b]	0.08 (0.04, 0.12) [c]	F = 1838.53 (< *.001*), partial η^2 = .39
K-SADS GAD diagnosis (%)	19	18	2	χ^2 = 75.33 (< *.001*), φ = .29
K-SADS MDD diagnosis (%)	27	25	2	χ^2 = 128.96 (< *.001*), φ = .37
K-SADS SUD diagnosis (%)	21	23	0	χ^2 = 134.32 (< *.001*), φ = .38

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. K-SADS = Schedule for Affective Disorders and Schizophrenia for School-age Children: Present and Lifetime Version (lifetime maximum symptoms/diagnosis), CD = conduct disorder, ODD = oppositional defiant disorder, ADHD = attention deficit/hyperactivity disorder, GAD = generalised anxiety disorder, MDD = major depressive disorder, SUD = substance use disorder. PDS = Self-rating Scale for Pubertal Development, ICU = Inventory of Callous-Unemotional Traits. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

Table 16. Additional clinical characteristics (mean (95% confidence intervals of the mean))

Measures	CD/HCU (*n* = 192)	CD/LCU (*n* = 183)	TD (*n* = 549)	F (*p*), partial η^2
RPQ reactive aggression	12.49 (11.82, 13.16) [a]	11.56 (10.82, 12.31) [a]	5.53 (5.22, 5.84) [b]	269.05 (< *.001*), .37
RPQ proactive aggression	5.28 (4.61, 5.95) [a]	4.57 (3.84, 5.31) [a]	0.88 (0.75, 1.01) [b]	169.34 (< *.001*), .27
RPQ total aggression	17.77 (16.57, 18.97) [a]	16.14 (14.79, 17.48) [a]	6.41 (6.03, 6.79) [b]	287.65 (< *.001*), .39
GEM affective empathy	0.53 (-0.39, 2.46) [a]	5.77 (4.00, 7.54) [b]	7.37 (6.34, 8.40) [b]	21.37 (*.001*), .05
GEM cognitive empathy	1.22 (-0.20, 2.64) [a]	7.62 (6.27, 8.98) [b]	11.51 (10.81, 12.21) [c]	96.68 (< *.001*), .18
GEM total empathy	7.33 (3.42, 11.24) [a]	28.48 (25.52, 31.44) [b]	35.84 (34.00, 37.68) [c]	110.55 (< *.001*), .20
CBCL internalizing	63.74 (62.11, 65.36) [a]	60.57 (58.72, 62.42) [b]	50.88 (49.99, 51.77) [c]	117.36 (< *.001*), .22
CBCL externalizing	73.01 (71.52, 74.49) [a]	64.42 (62.62, 66.21) [b]	48.75 (47.91, 49.59) [c]	431.42 (< *.001*), .50
CBCL total problems	70.47 (68.83, 72.11) [a]	63.93 (62.06, 65.80) [b]	49.02 (48.12, 49.92) [c]	304.12 (< *.001*), .42

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. RPQ = Reactive-Proactive Aggression Questionnaire, GEM = Griffiths Empathy Measure, CBCL = Child Behaviour Checklist. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

5.3.2 Hexagon Task Performance

Raw mean group accuracies for each emotion × intensity trial type are shown in **Table 17**. Interactions between group, emotion and intensity were investigated using a repeated measures ANCOVA (see Methods, section 5.2.4)[20]. Degrees of freedom were

[20] The Hexagon performance data were not normally distributed and could not be normalised with log-transformations. Given the large size of the sample, however, a decision was made to use parametric tests in line with advice from the FemNAT-CD statisticians and previous work with the same sample (Kohls *et al*., 2019).

adjusted using the Greenhouse-Geisser method to account for non-sphericity. There was a significant main effect of group ($F_{(2, 867)} = 9.17, p < .001$, partial $\eta^2 = .02$); youths with both CD/HCU and CD/LCU significantly underperformed relative to TD youths (see **Figure 19**). There was also a significant main effect of emotion ($F_{(3.69, 3196.87)} = 84.09, p < .001$, partial $\eta^2 = .09$). Recognition accuracy was highest for happiness, followed by sadness and surprise, and lowest for anger, fear and disgust (see **Figure 20**). Finally, there was a significant main effect of intensity, with greater accuracy for 90% intensity trials than for 70% intensity trials ($F_{(1, 867)} = 113.54, p < .001$, partial $\eta^2 = .12$). However, there was no significant interaction between group and intensity ($F_{(2, 4016.57)} = .96, p = .38$, partial $\eta^2 = .002$) or, crucially, between group and emotion ($F_{(7.38, 867)} = .54, p = .82$, partial $\eta^2 = .001$; see **Figure 21**). This indicates that emotion recognition difficulties in CD/HCU (and indeed CD generally) are not specifically pronounced for negative emotions.

Interestingly, the relative performance of the CD/HCU, CD/LCU and TD groups differed between sites. There was a significant main effect of site ($F_{(9, 867)} = 4.00, p < .001$, partial $\eta^2 = .04$) and a significant site × group interaction ($F_{(18, 867)} = 2.85, p < .001$, partial $\eta^2 = .06$). Closer inspection of the data revealed that the CD/HCU group actually outperformed the CD/LCU group at some sites. However, when ANCOVAs were conducted for each site separately, there were no clear cultural patterns (see Appendix C). Since these site differences did not appear to have a theoretically meaningful explanation, and due to the small numbers of participants at many sites (*e.g.*, n = 25), we thus proceeded with the remaining analyses as planned, without further separating participants by site.

Table 17. Emotion recognition accuracy per trial type and group (mean (95% confidence intervals of the mean))

Trial type		CD/HCU (*n* = 192)	CD/LCU (*n* = 183)	TD (*n* = 549)
Anger:	***70%***	56.77 (52.33, 61.21)	59.34 (55.01, 63.68)	66.72 (64.31, 69.13)
	90%	64.38 (59.75, 69.00)	65.63 (61.17, 70.09)	75.76 (73.38, 78.14)
Disgust:	***70%***	55.47 (50.84, 60.09)	58.69 (53.94, 63.44)	66.23 (63.62, 68.84)
	90%	57.14 (51.95, 62.32)	59.73 (54.37, 65.09)	70.18 (67.32, 73.05)
Fear:	***70%***	57.34 (53.22, 61.46)	57.76 (53.82, 61.69)	69.54 (67.50, 71.59)
	90%	60.20 (56.16, 64.25)	64.59 (60.26, 68.92)	75.48 (73.46, 77.51)
Happiness:	***70%***	88.12 (85.71, 90.64)	86.83 (83.98, 89.69)	94.13 (93.21, 95.06)
	90%	89.84 (87.23, 92.45)	89.94 (87.46, 92.43)	95.76 (95.00, 96.51)
Sadness:	***70%***	75.26 (71.51, 79.01)	76.39 (72.88, 79.90)	85.90 (84.37, 87.43)
	90%	78.65 (75.06, 82.24)	81.86 (78.33, 85.38)	85.60 (88.06, 91.14)
Surprise:	***70%***	73.13 (69.37, 76.88)	77.16 (73.89, 80.43)	83.26 (81.69, 84.83)
	90%	77.92 (74.04, 81.79)	81.04 (77.80, 84.28)	88.54 (87.07, 90.02)

Notes: CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing

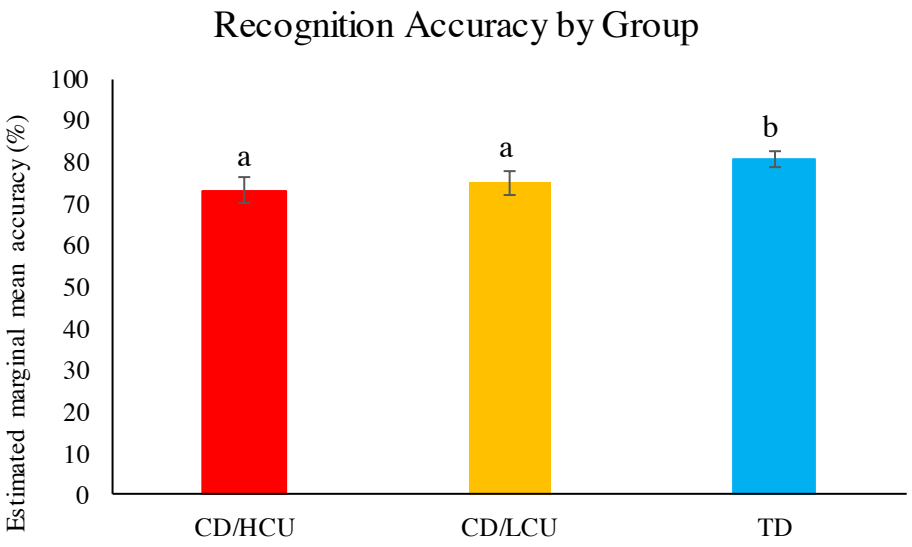


Figure 19. Estimated marginal mean emotion recognition accuracy per group. Error bars show 95% confidence intervals of the mean. Emotions marked with different letters differ significantly (post-hoc tests with Bonferroni corrections, $p < .05$)

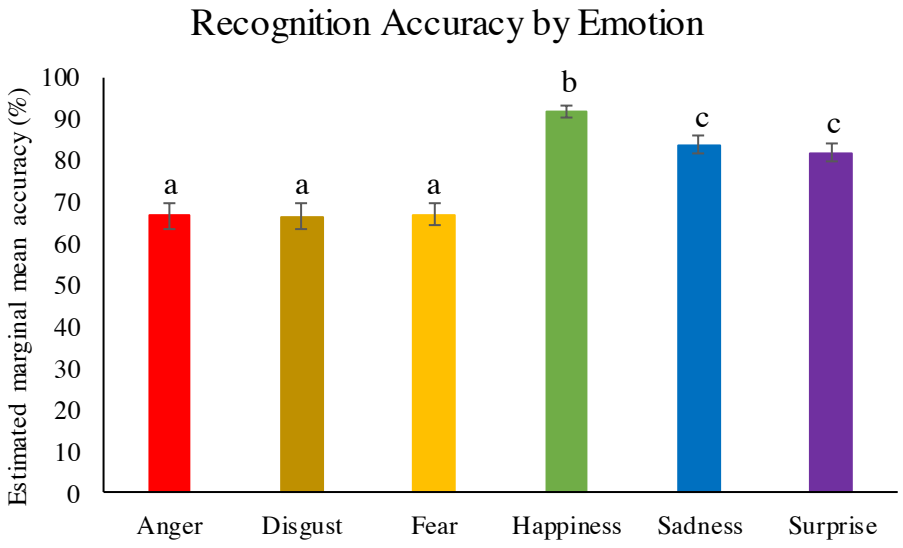


Figure 20. Estimated marginal mean emotion recognition accuracy per emotion. Error bars show 95% confidence intervals of the mean. Emotions marked with different letters differ significantly (post-hoc tests with Bonferroni corrections, $p < .05$)

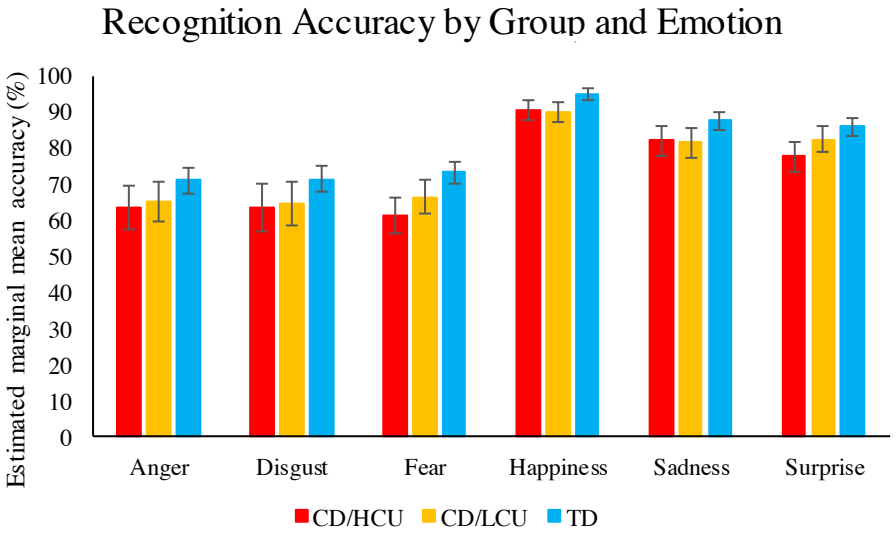


Figure 21. Estimated marginal mean emotion recognition accuracy per group, for each emotion. Error bars show 95% confidence intervals of the mean

5.3.3 Effect of Comorbid Diagnoses on Group Differences

The ANCOVA analysis was repeated after regressing out variance associated with lifetime diagnoses of ADHD, major depressive disorder, generalised anxiety disorder and substance use disorder. Results were similar, except there was no longer a main effect of emotion ($F_{(4.42, 3831.30)} = 0.50, p = .76$, partial $\eta^2 = .001$). There was no main effect of intensity ($F_{(1, 687)} = 1.12, p = .29$, partial $\eta^2 = .001$). There was a significant main effect of group ($F_{(2, 867)} = 3.11, p = .045$, partial $\eta^2 = .01$), with the CD/HCU group underperforming relative to both other groups, but again there was no significant interaction between emotion and group ($F_{(1, 867)} = 0.65, p = .75$, partial $\eta^2 = .001$).

5.3.4 Exploratory Analyses of Sex Differences in Emotion Recognition

We also explored sex differences in emotion recognition abilities. Although females outperformed males with 75% versus 72% accuracy respectively, there was no

significant main effect of sex ($F_{(1, 867)} = 1.20, p = .27$, partial $\eta^2 = .001$). There was also no significant sex × emotion interaction ($F_{(3.69, 3196.87)} = 2.36, p = .06$, partial $\eta^2 = .003$), no significant sex × group interaction ($F_{(2, 867)} = 0.29, p = .75$, partial $\eta^2 = .001$), and no significant sex × group × emotion interaction ($F_{(7.38, 3196.87)} = 0.44, p = .88$, partial $\eta^2 = .001$). There was thus no evidence that sex differences obscured differences in emotion recognition abilities between youths with CD/HCU, youths with CD/LCU and TD youths.

5.3.5 Exploratory Analyses of Group Differences in Error Type

We explored group differences in the number of remote prototype errors, *i.e.*, confusion of an emotion with a non-adjacent emotion, which was never blended with the dominant emotion. These error types can be considered 'complete' errors because the emotion identified by the participant was not to any degree present in the face shown. Group differences were explored using a repeated-measures ANCOVA, with emotion as the within-subjects factor and group as the between-subjects factor. Sex and site were included as additional between-subjects factors of no interest, and mean-centred pubertal status and IQ scores were used as covariates. The results were very similar to those for accuracy. There was a significant main effect of group ($F_{(2, 867)} = 3.12, p = .045$, partial $\eta^2 = .01$), with the CD/HCU group making more remote prototype errors than the other groups (percentage of remote prototype errors: 6% CD/HCU, 5% CD/LCU, 4% TD; both $p < .05$, Bonferroni correction). There was a significant main effect of emotion, matching that observed in the accuracy analyses ($F_{(2.68, 2324.42)} = 55.15, p < .001$, partial $\eta^2 = .06$). However, there was no emotion × group interaction (F

$_{(48.25, 867)} = .90$, $p = .67$, partial $\eta^2 = .02$). Again, this pattern of errors does not suggest particular difficulties with negative emotions for youths with CD/HCU.

5.3.6 Angle-GMLVQ Classifier Performance

Angle-GMLVQ classifier performance is shown in **Table 18**. Differences in performance between the Mixed-TD, HCU-TD and LCU-TD models were small. In terms of accuracy, classifier performance was better for the HCU-TD (0.63) and LCU-TD (0.61) models than for the Mixed-TD model (0.59). However, in terms of MCERs, only the HCU-TD model (0.40) outperformed the Mixed-TD model (0.41). This discrepancy reflects the much more accurate classification of TD participants, combined with the larger size of the TD group compared to the CD groups in the HCU-TD and LCU-TD models. This resulted in improvements in accuracy that were largely driven by better identification of TD participants. After accounting for this imbalance, only the CD/HCU participants were better classified when the CD/mixed group was split into CD/HCU and CD/LCU groups. When these groups were distinguished from each other directly (HCU-LCU model), performance was slightly below chance.

Table 18. Angle-GMLVQ model performance (mean (95% confidence intervals of the mean))

	Mixed-TD	HCU-TD	LCU-TD	F (*p*), partial η^2	HCU-LCU
Accuracy	0.59 (0.59, 0.59) [a]	0.63 (0.62, 0.63) [b]	0.61 (0.61, 0.61) [c]	247.01 (<.001), .14	0.49 (0.48, 0.49)
PPV	0.61 (0.60, 0.61) [a]	0.45 (0.44, 0.45) [b]	0.32 (0.32, 0.32) [c]	8863.80 (<.001), .86	0.48 (0.48, 0.49)
NPV	0.58 (0.57, 0.58) [a]	0.74 (0.74, 0.74) [b]	0.79 (0.79, 0.79) [c]	12390.29 (<.001), .89	0.49 (0.48, 0.49)
TPR	0.50 (0.50, 0.51) [a]	0.51 (0.51, 0.52) [b]	0.48 (0.48, 0.49) [c]	49.83 (<.001), .03	0.46 (0.46, 0.47)
TNR	0.67 (0.67, 0.67) [a]	0.68 (0.68, 0.69) [b]	0.65 (0.65, 0.66) [c]	67.96 (<.001), .04	0.51 (0.51, 0.52)
MCER	0.41 (0.41, 0.41) [a]	0.40 (0.40, 0.41) [b]	0.43 (0.43, 0.44) [c]	134.82 (<.001), .08	0.51 (0.51, 0.52)

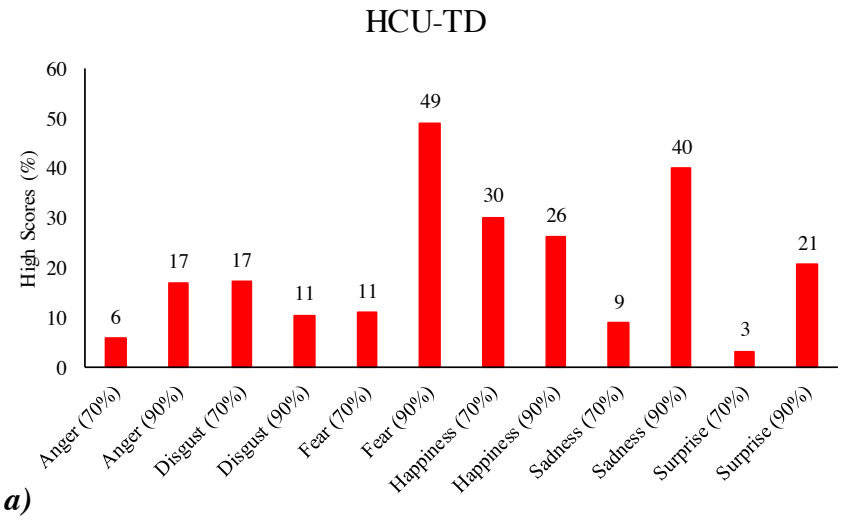
Notes: Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. PPV = positive predictive value, NPV = negative predictive value, TPR = true positive rate, TNR = true negative rate, MCER = macro-averaged classification error rate. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < 0.05$, Bonferroni corrected). Note that the HCU-LCU model (column 6) was not included in statistical tests as comparisons between this and other models were not relevant to hypotheses

5.3.7 Feature Relevance

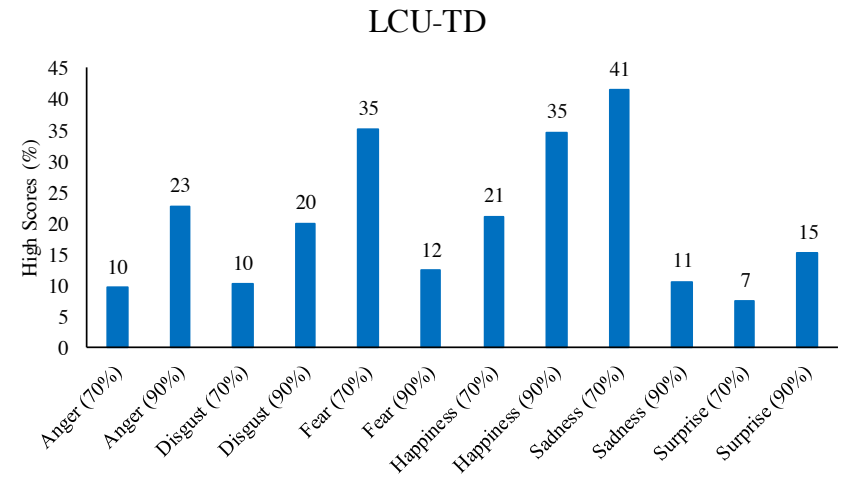
Feature relevance scores are shown in **Figure 22**. Fear and sadness were generally the most relevant emotions, with surprise also highly relevant to the HCU-LCU model[21]. For the HCU-TD and LCU-TD models, the 90% intensity trials were generally more relevant than their 70% intensity counterparts. This pattern suggests that TD youths as well as youths with CD sometimes struggled to identify the less intense expressions. However, TD youths typically benefited from an increase in expression intensity, whereas youths with CD did not always do so. Interestingly, there were two emotions in each of these models where the 70% intensity trials were more relevant than the 90%

[21] Interestingly, an exploratory investigation of 'false positives' revealed that surprise and fear were mistaken for each in approximately 25% of incorrect responses, across both intensities and both CD groups.

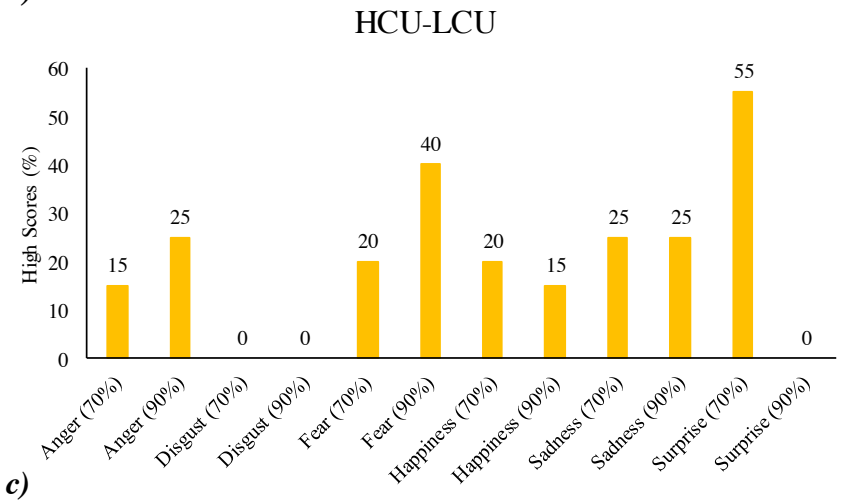
intensity trials; these were disgust and happiness in the HCU-TD model, and fear and sadness in the LCU-TD model. Although this finding is not straightforward to interpret, it appears that youths with CD/LCU might have benefited from an increased fear and sadness intensity, while youths with CD/HCU were less likely to do so. Likewise, youths with CD/HCU might have benefited from an increased disgust and happiness intensity, while youths with CD/LCU did not appear to benefit to the same extent. However, in the HCU-LCU model there was no consistency as to whether 70% or 90% intensity trials were more relevant, presumably reflecting the similar performance of these groups.



a)



b)



c)

Figure 22. Feature relevance for ***a)*** HCU-TD model, ***b)*** LCU-TD model and ***c)*** HCU-LCU models. Bars show percentage of re-samplings in which feature relevance was in the top 20% of relevance scores across all re-samplings with MCER ≤ 0.40

5.3.8 SVM Classifiers

Of all the SVM classifiers tested, the linear SVM performed best for each model (see **Table 19**). MCER values for the linear SVMS were 0.42 for the Mixed-TD, HCU-TD and LCU-TD models and 0.49 for the HCU-LCU model. In each case the corresponding Angle-GMLVQ classifier exhibited significantly better performance (independent samples t-tests, all $p < .001$), except for the HCU-LCU model, where the SVM model was the better performer ($t_{(1998)} = 9.21, p < .001$). Across all models, differences in performance for the SVM and Angle-GMLVQ classifiers were small (*c.f.* **Table 18**).

Table 19. Performance (MCER) for linear and non-linear SVM models (mean (95% confidence intervals of the mean))

	Mixed-TD	**HCU-TD**	**LCU-TD**	**HCU-LCU**
Linear	0.42 (0.42, 0.43) [a]	0.42 (0.42, 0.42) [a]	0.44 (0.44, 0.44) [a]	0.49 (0.49, 0.50)
2nd order polynomial	0.45 (0.45, 0.45) [b]	0.45 (0.45, 0.46) [b]	0.47 (0.47, 0.48) [b]	0.49 (0.49, 0.50)
3rd order polynomial	0.46 (0.46, 0.47) [c]	0.44 (0.44, 0.45) [c]	0.49 (0.48, 0.49) [c, d]	0.49 (0.49, 0.50)
4th order polynomial	0.46 (0.46, 0.47) [c]	0.44 (0.44, 0.45) [c]	0.49 (0.49, 0.49) [c]	0.50 (0.49, 0.50)
5th order polynomial	0.47 (0.47, 0.47) [d]	0.46 (0.45, 0.46) [b]	0.48 (0.48, 0.48) [d]	0.49 (0.49, 0.50)
6th order polynomial	0.49 (0.49, 0.50) [d]	0.47 (0.47, 0.48) [d]	0.49 (0.49, 0.49) [c]	0.50 (0.49, 0.50)
F (*p*)	370.32 (*<.001*)	148.10 (*<.001*)	192.21 (*<.001*)	1.30 (*.26*)

Notes: MCER = macro-averaged classification error rate. Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. Within each column, models with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

5.4 Discussion

The purpose of this study was to investigate differences in emotion recognition abilities in youths with CD/HCU, youths with CD/LCU and TD youths. First, we hypothesised that youths with CD would perform more poorly than TD youths. Second, and more specifically, we hypothesised that youths with CD/HCU would perform more poorly across all emotions than youths with CD/LCU, who in turn would perform more poorly than TD youths. Youths with CD did indeed underperform relative to TD youths, but contrary to our second hypothesis, youths with CD/HCU and CD/LCU performed similarly. Third, we explored whether youths with CD/HCU would exhibit especially poor performance for negative emotions. This did not appear to be the case; there was no significant interaction between emotion and group. Fourth, in the classification analyses, only the HCU-TD model significantly outperformed the Mixed-TD model, and the HCU-LCU model performed poorly. Finally, in partial support of our last hypothesis, the most relevant emotions for the HCU-TD model (and indeed all models) were negative emotions, although happiness and surprise were nonetheless more relevant than several negative emotions.

5.4.1 Emotion Recognition Difficulties in CD

These findings support previous evidence that emotion recognition difficulties are common in CD (*e.g.*, Sully *et al.*, 2015). In contrast to a previous all-female sample (Pajer *et al.*, 2010), but in line with a larger FemNAT-CD mixed-sex sample (Kohls *et al.*, 2019), such difficulties were seen in females as well as males. Difficulties with anger and disgust recognition in psychopathy have been hypothetically linked to orbitofrontal cortex dysfunction (*e.g.*, Blair & Cipolotti, 2000; Blair, 2003), while

difficulties with sadness, fear and (to a lesser extent) happiness have been linked to amygdala dysfunction (Blair, 2003). The pervasive difficulties seen in the current sample thus indicate a potential role for both the orbitofrontal cortex and amygdala in emotion recognition difficulties in CD.

5.4.2 Similarity between CD/HCU and CD/LCU

Our findings also align with those studies that reported similar emotion recognition abilities in CD/HCU and CD/LCU (Rehder *et al*., 2017; Schwenck *et al*., 2012). However, contrary to these findings, neurocognitive models and several previous experimental findings indicate specific difficulties with negative emotions – especially fear and sadness – in CD/HCU (*e.g*., Blair, 2003; Blair *et al*., 2005; Blair *et al*., 2001; Stevens *et al*., 2001; Blair & Coles, 2000; Bowen *et al*., 2014). Interestingly, although anger recognition difficulties have been associated with psychopathy, they appear to be more consistently related to the antisocial behaviour dimension rather than CU traits (Blair & Coles, 2000; Bowen *et al*., 2014). This fits with hypothesised orbitofrontal cortex dysfunction underlying reactive aggression, which is common to both CD/HCU and CD/LCU (*e.g*., Blair, 2003; Blair *et al*., 2005). In the current sample, youths with CD/HCU and CD/LCU did not differ significantly on reactive aggression, which might relate to their similarity in anger recognition abilities. Even more unusually, the CD/HCU and CD/LCU groups did not differ significantly on proactive aggression, which is hypothesised to be linked to dysfunctional amygdala responsivity to fearful and sad expressions (Blair, 2003). This similar behavioural profile of the CD/HCU and CD/LCU groups is unusual compared to previous studies (*e.g*., Blair *et al*., 2001), and might go some way to explaining the similar emotion recognition profiles observed in

CD/HCU and CD/LCU in the current sample. It could thus be argued that CU traits alone, in the absence of other psychopathic features such as proactive aggression, are not necessarily associated with poor fear and sadness recognition. Finally, it should be noted that differences in emotion recognition between the CD/LCU and TD groups were reduced to non-significance, with a decreased effect size, after controlling for comorbid diagnoses. This suggests that for youths with CD/LCU in particular, other psychopathologies might play a role in emotion recognition difficulties. However, we interpret this finding with caution, because the CD/LCU group lay intermediate to the other groups in both analyses, and it is possible that a similar decrease in group differences between all three groups would cause the CD/LCU group to 'drop' below the significance level that originally distinguished them from the TD group.

5.4.3 Distinguishing between Individuals based on Emotion Recognition Abilities

In line with these group level differences, only the HCU-TD model significantly outperformed the Mixed-TD model when accounting for imbalanced class sizes. The HCU-LCU model also performed poorly (MCER = 0.51). Notably, however, the HCU-TD model slightly but significantly outperformed the LCU-TD model. This highlights a tendency for youths with CD/HCU to perform more poorly than youths with CD/LCU, despite the lack of significant differences at the group level. Interestingly, however, classification of TD youths was consistently better than classification of youths with CD. These differences in performance indicate that, even for youths with CD/HCU, a large proportion (approximately 50%) exhibited emotion recognition accuracies that were closer to the 'prototypical' TD youth than to the prototype for their own group. Thus, impoverished emotion recognition abilities are far from consistently present in

CD. Given the site differences observed in the current sample, as well as previous work by Rehder *et al*. (2017), cultural background is an interesting potential influence in emotion recognition abilities in CD.

The three most relevant emotions for both the HCU-TD and LCU-TD models were fear, happiness and sadness. Fear was also the second most relevant emotion in the HCU-LCU model, after surprise (which was frequently mistaken for fear). These relevance scores were broadly in line with the magnitude of group level differences for each emotion. Interestingly, recognition of the most relevant emotions across models – fear, sadness and happiness – are those theoretically linked to amygdala dysfunction in psychopathy (Blair, 2003). The relevance of these emotions to both the HCU-TD and LCU-TD models is thus notable. However, despite these similarities, there were indications that youths with CD/LCU were to some extent able to 'catch up' with TD youths for more intense fear and sadness, whereas youths with CD/HCU were not able to do so (as judged from the relative relevance of 70% and 90% intensity trials for these emotions). It thus appears that youths with CD/HCU benefitted relatively little from increased fear and sadness intensity compared to the CD/LCU group. This fits with a previous report of reduced sensitivity to gradually intensifying fearful expressions in youths with CD/HCU (Blair *et al*., 2001). However, there are likely to be complex interactions between CD and CU traits, emotions and expression intensity that will require further investigation before firm conclusions can be drawn.

5.4.4 Strengths and Limitations

This study has several strengths, including a large, well characterised, mixed-sex sample and the use of both univariate and multivariate analyses, illustrating diversity in

emotion recognition abilities at both the group and individual level. However, some limitations should also be noted. First, the Hexagon task uses only static facial expressions, and the identity of the face is the same across all expressions. This results in a fairly limited set of stimuli. The ecological validity of the task would be improved by the inclusion of a range of facial identities at multiple intensities. This would also be helpful for further investigating the intensity effects described above. Second, the TD participants were selected on the basis of having no current diagnosable mental disorders, meaning that they were not completely representative of the non-CD population. This limitation is common to Chapters 3-5, in line with the FemNAT-CD inclusion criteria, but is notable here because of the complex associations between internalising disorders and emotion recognition difficulties (*e.g.*, Schepman *et al.*, 2012; Guyer *et al.*, 2007). Finally, the relative performance of the CD/HCU, CD/LCU and TD groups differed across sites. The small number of participants at some sites precluded a detailed investigation of cultural, linguistic or national differences. Given the dominance of the English-speaking world in CD and psychopathy research, cultural differences are an important topic for future research.

5.4.5 Summary and Conclusions

In summary, these findings indicate that youths with CD have difficulty in recognising facial emotions relative to TD youths, and these difficulties are pervasive rather than limited to negative emotions. However, there is a large degree of overlap in emotion recognition abilities between TD youths and those with CD. There was minimal evidence for differences in emotion recognition abilities between youths with CD/HCU and CD/LCU. Nonetheless, at the individual level, youths with CD/HCU were more

easily distinguished from TD youths on the basis of emotion recognition ability than were those with CD/LCU. There were also indications that youths with CD/HCU might benefit less than those with CD/LCU from an increase in expression intensity for fear and sadness. In conclusion, emotion recognition difficulties are common but far from universal in both CD/HCU and CD/LCU. This highlights heterogeneity within CD and calls for further investigation of exacerbating and ameliorating factors in emotion recognition ability.

CHAPTER 6: GENERAL DISCUSSION

This concluding chapter provides a short summary of the findings from Chapters 3-5, a discussion of their implications, strengths and weaknesses, and future directions. An overview of the main findings from each chapter is provided in **Table 20**.

Table 20. Summary of Chapters 3-5

	Chapter 3	**Chapter 4**	**Chapter 5**
Topic	Parenting	Grey matter volume	Facial emotion recognition
Sample size	756	452	924
Univariate (group level) analyses			
Summary of findings	CD/HCU & CD/LCU experience less positive and more negative parenting than TD. CD/HCU experience less positive parenting than CD/LCU	Group differences only when controlling for ADHD. Whole brain: CD < TD & CD/HCU < TD in right rolandic operculum/insula. Regions of interest: left insula: CD < TD & CD/HCU < TD. Left orbitofrontal cortex: CD/HCU < TD	CD/HCU & CD/LCU poorer than TD across emotions. No group × emotion interaction
Angle-GMLVQ performance (MCER, accuracy)			
Mixed-TD	0.29, 0.71	0.52, 0.48	0.41, 0.59
HCU-TD	0.26, 0.75	0.44, 0.55	0.40, 0.63
LCU-TD	0.33, 0.69	0.43, 0.58	0.43, 0.61
HCU-LCU	0.42, 0.58	0.35, 0.65	0.51, 0.49

Notes: Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. Angle-GMLVQ = Angle-Based Generalised Matrix Learning Vector Quantisation, MCER = macro-averaged classification error rate

6.1 Summary of Findings

6.1.1 Chapter 3: Parenting Behaviour

The work presented in this chapter was motivated by a large and growing body of literature that suggests differential relationships between CD, CU traits and positive

versus negative parenting (*e.g.*, Waller *et al.*, 2013). In this Chapter, we demonstrated group level differences in several aspects of positive and negative parenting. Youths with CD – regardless of subtype – experienced high levels of negative parenting and low levels of positive parenting relative to TD youths. However, only positive parenting differed significantly between the CD/HCU and CD/LCU groups. At the individual level, the group status of youths (*i.e.*, CD/HCU, CD/LCU or TD) could be predicted with error rates (MCER) between 0.26 and 0.42, corresponding to overall accuracies of 0.75 to 0.58. Parental involvement (but not positive reinforcement) was highly relevant when distinguishing youths with CD/HCU from any other group. Negative parenting was also highly relevant when distinguishing CD groups from the TD group, but not when distinguishing CD/HCU from CD/LCU.

6.1.2 Chapter 4: Grey Matter Volume

This chapter was motivated by previous research suggesting lower grey matter volume in youths with CD, as well as in youths with CD/HCU relative to those with CD/LCU (Raschle *et al.*, 2015; Noordermeer *et al.*, 2016; Rogers & De Brito, 2016; Sebastian *et al.*, 2016). We demonstrated lower grey matter volumes in both CD generally, and CD/HCU specifically, in the right insula (and operculum) relative to TD youths. Youths with CD/HCU also exhibited lower volumes relative to TD youths in the left orbitofrontal cortex. However, these differences were only significant after accounting for comorbid ADHD diagnoses. Classifier performance ranged from MCERs of 0.52 to 0.35, corresponding to overall accuracies of 0.48 to 0.65. Due to a discrepancy between the direction of group differences in the VBM analyses (CD/HCU < TD) and the mean

volume in each region (CD/HCU > TD > CD/LCU), the most relevant features were those with the least difference in the VBM analyses (*e.g.*, the amygdala).

6.1.3 Chapter 5: Facial Emotion Recognition

This chapter was motivated by research suggesting that youths with CD/HCU might experience specific difficulties with recognition of negative emotions, relative to youths with CD/LCU (Blair *et al.*, 2001; Stevens *et al.*, 2001; Bowen *et al.*, 2014). Although youths with CD performed more poorly than TD youths across all emotions, there were no significant differences in performance for youths with CD/HCU and youths with CD/LCU (expect when controlling for comorbid diagnoses; here, the CD/HCU group differed from the other groups). Neither was there evidence that youths with CD/HCU, or CD generally, were poorer at recognising negative emotions specifically. Classifier performance ranged from MCERs of 0.40 to 0.51, corresponding to overall accuracies of 0.49 to 0.63. Across models, fear was consistently high in relevance. Sadness was also highly relevant when distinguishing youths with either CD/HCU or CD/LCU from TD youths.

6.2 Different Developmental Pathways to CD

The findings from Chapters 3-5 point to both similarities and differences between CD/HCU and CD/LCU. Neurobiologically, youths with CD/HCU and youths with CD/LCU exhibit subtle, widespread differences in grey matter volume, falling on opposite sides of TD youths. These are coupled with differences in parenting that are largely (but not exclusively) a question of magnitude, and minimal evidence for differences in emotion recognition ability. Together, these findings are suggestive of

different, but overlapping, developmental pathways to similar outcomes, influenced by different underlying vulnerabilities.

In a recent review of the literature, Viding and McCrory (2019) provide a conceptual framework for understanding the development of atypical social affiliation in CD/HCU. They highlight evidence for atypical affiliative tendencies from a very young age in children who go on to develop high levels of CU traits. For example, as young as five weeks old, reduced tracking of the mother's face was associated with increased levels of CU traits at 2.5 years old (Bedford, Pickles, Sharp, Wright, & Hill, 2015). In later childhood, reduced eye contact was also associated with CU traits, and seemed to be driven by the child rather than the caregiver (Dadds, Jambrak, Pasalich, Hawes, & Brennan, 2011; Dadds *et al*., 2012; Dadds, Allen, McGregor, Woolgar, Viding, & Scott, 2014). There is also evidence that children with elevated CU traits experience less affective resonance in response to a positive emotion (hearing laughter; O'Nions, Lima, Scott, Roberts, McCrory, & Viding, 2017). Children's CU traits appear to contribute to reductions in positive parenting (Flom *et al*., 2019; Muñoz *et al*., 2011; Salihovic *et al*., 2012) as well as vice versa (Hyde *et al*., 2016; Waller *et al*., 2016; Waller *et al*., 2017), and reductions in positive parenting have in turn been linked to reduced affective empathy (Kochanska *et al*., 2005). Low affective empathy has been linked to atypical function in a network of brain regions, including the anterior insula (*e.g*., Fan, Duncan, de Greck, & Northoff, 2011; Kanske, Böckler, Trautwein, & Singer, 2015), which was a key region of grey matter reductions in CD/HCU in this sample (Chapter 4). Interpreting the present findings in the light of this literature, we speculate that neurobiological vulnerabilities (*e.g*., atypical grey matter volume of the anterior insula) might contribute to the reductions in positive parenting seen here in CD/HCU. By

contrast, youths with CD/LCU might have different – and perhaps fewer – neurobiological vulnerabilities. In these youths, it is possible that low positive (and high negative) parenting is more of a driving factor behind behaviour problems. Interestingly, however, and perhaps counterintuitively, poor emotion recognition appears to be independent of infants' eye gaze and maternal sensitivity as a risk factor for CD/HCU (Bedford, Wagner, Rehder, Propper, Willoughby, & Mills-Koonce, 2017). Elucidating these developmental pathways to CD/HCU and CD/LCU, and from there to psychopathy, ASPD or cessation of antisocial behaviour in adulthood, requires further investigation (Viding & McCrory, 2019). More generally, the interplay between neurobiological risk factors and exacerbating or ameliorative social environments is an important topic for future research (see Barker, Walton, and Cecil (2018) and Cecil *et al*. (2014) for work in this area).

6.3 Similarities between Subtypes

Despite evidence for different developmental pathways to CD/HCU and CD/LCU, there are still many similarities in presentation between the subtypes. In Chapters 3-5, both youths with CD/HCU and youths with CD/LCU exhibited differences from TD youths at the group level, including higher levels of negative parenting, lower levels of positive parenting and difficulties in identifying all six of the basic emotional facial expressions. Youths with CD as a whole also exhibited localised reductions in grey matter volume in the insulae. Second, excepting Chapter 4 (grey matter volume differences), youths with CD were distinguished from TD youths at above-chance levels, while CD/HCU was distinguished from CD/LCU rather poorly. Thus, in terms of emotion recognition abilities in particular, and to a lesser extent parenting, CD could be distinguished quite

easily from TD, without making the distinction between CD/HCU and CD/LCU. That there are similarities as well as differences between subtypes is a point that should not be overlooked, and demonstrates that despite different developmental pathways, there are also common risk factors in the development of both subtypes of CD.

6.4 Limitations of the CU Subtypes for Addressing Heterogeneity and Alternative Approaches

6.4.1 Need for a Clinical Threshold for CD/HCU

The experimental findings presented in this thesis make clear that there remains considerable overlap between CD/HCU and CD/LCU. In addition to genuine shared characteristics, a methodological decision that might partially explain this outcome is the reliance on sample-specific thresholds for CD/HCU (*i.e.*, tertile and median splits). Although this approach is common (*e.g.*, Wootton *et al.*, 1997), it naturally leads to definitions of CD/HCU that reflect the distribution of CU traits in the overall sample. In a largely community-based sample, such as this one, the threshold for CD/HCU is likely to be relatively low[22]. Although CU traits and psychopathic traits appear to be dimensional (Clark, 2007; Murrie *et al.*, 2007; Edens *et al.*, 2006), it is nonetheless possible that some of the characteristics associated with CD/HCU manifest only at the extreme end of the CU spectrum. This could result if their relationships with CU traits are non-linear, as has been demonstrated for psychopathy and risk for schizophrenia (Abu-Akel, Heinke, Gillespie, Mitchell, & Bo, 2015). Alternatively, Schaich Borg, Kahn, Sinnott-Armstrong, Kurzban, Robinson and Kiehl, (2013) suggest that

[22] The exclusive use of parent ratings, as opposed to combined parent, teacher and/or child ratings, is another factor that might contribute to differences in thresholds between studies.

interactions between different components of psychopathic traits might lead to a psychopathy taxon, despite the constituent traits being dimensional in nature. If this is the case, then there remains an urgent need to define an objective, research-informed clinical threshold for CD/HCU that can be applied consistently across populations (see Kimonis *et al*., 2014).

6.4.2 Distinct Components of CU Traits

Although CU traits have usually been treated as a unitary construct, recent research suggests that they can be divided into distinct subcomponents. For example, a twin study using the ICU provided evidence for two specific sub-factors; an unemotional factor and a combined callous-uncaring factor, with a low genetic correlation between these two dimensions (Henry, Pingault, Boivin, Rijsdijk, & Viding, 2016). Several recent neuroimaging studies have also demonstrated a role for callous traits specifically, which are sometimes more predictive of group differences than overall CU traits (*e.g*., Lockwood, Sebastian, McCrory, Hyde, Gu, De Brito, & Viding, 2013; Rogers *et al*., 2019). These studies highlight the importance of investigating the aetiology and clinical characteristics associated with the various components of CU traits, as well as the overall CU construct.

6.4.3 Demographic Heterogeneity

Another factor that might contribute to overlap between subtypes is the demographic heterogeneity of the sample. Although sex and age effects were minimal in the current research, there is evidence that CD and CU traits can manifest differently depending on these factors. For example, females with CD/HCU engage in relatively more covert and

manipulative behaviours than overt aggression, as compared to males (O'Keefe, Carr, & McQuaid, 1998). There is also evidence for differences in the age of onset and persistence of CD in males and females (Fontaine, Carbonneau, Vitaro, Barker, & Tremblay, 2009). Given the over-representation of females in FemNAT-CD, the potential for sex differences is an important factor to consider when comparing the current findings with those from male- dominated samples. In addition, data on socioeconomic status for the current sample were not available at the time when this thesis was completed. Ideally, socioeconomic status should be taken into account.

Cultural differences are another oft-neglected area for potential differences in presentation of CD and CU traits. Researchers have demonstrated cultural differences in the prevalence of externalising behaviours (Zwirs, Burger, Schulpen, & Buitelaar, 2006) and their treatment (Zwirs *et al*., 2006; Safer & Malever, 2000). There are also indications that the association between CU traits and emotion recognition abilities might differ between ethnic groups (Rehder *et al*., 2017), a finding that has some parallels with fear recognition in psychopathy (*e.g*., Baskin-Sommers, Newman, Sathasivam, & Curtin, 2011). However, CD samples recruited in China appear to show similar neurobiological characteristics to western samples (*e.g*., Zhang *et al*., 2018a). Data on ethnicity were not available for the FemNAT-CD sample at the time when the analyses presented in this thesis were conducted, but there were indications of site differences in emotion recognition (though group sizes were very small, *e.g*., n = 25; see Appendix C). Clearly, cultural differences in the presentation of CD/HCU and CD/LCU – and perhaps even the extent to which this distinction is meaningful in non-westernised cultures – is an important direction for future research.

6.4.4 Alternative Subtyping Approaches

Finally, recent years have witnessed a shift in focus towards subtyping methods based on broader concepts of psychopathy. First, there is continued interest in primary versus secondary variants of psychopathy (Karpman, 1948). This distinction is thought to map onto 'idiopathic' psychopathy (primary psychopathy) versus environmentally induced (secondary) psychopathy, and appears to manifest in differences in anxiety levels (Lykken, 1957). Physiological measures of anxiety tend to be negatively correlated with CU traits (*e.g.*, Fanti *et al*., 2016), and youths with CD/HCU can be further divided into those with high versus low anxiety (*e.g.*, Zwaanswijk, van Geel, Andershed, Fanti, & Vedder, 2018; Fanti, Demetriou, & Kimonis, 2013). There is evidence that within CD/HCU, amygdala reactivity to threat differs in youths with high versus low anxiety (Fanti, Konikou, Cohn, Popma, & Brazil, 2019). Regarding heritability, CD/HCU with low anxiety does not appear to differ from CD/HCU with high anxiety (Humayun, Kahn, Frick, & Viding, 2014). More recently, however, Cecil, McCrory, Barker, Guiney and Viding (2018) demonstrated significantly higher levels of childhood maltreatment, attachment insecurity, psychological distress and symptomatology (*e.g.*, suicidal ideation, ADHD), in youths with elevated CU traits and anxiety compared to those with elevated CU traits but low anxiety. This study provides some evidence for anxiety-based primary and secondary variants in childhood CU traits, with environmental adversity (*e.g.*, maltreatment) being more strongly associated with the secondary variant. Consistent with these findings, there is some evidence that trauma exposure is an important factor in distinguishing primary from secondary variants. For example, Meffert *et al*. (2018) found that for youths with low trauma exposure, right

amygdala BOLD signal to fearful faces was negatively associated with CU traits. By contrast, it was positively associated with CU traits in youths high in trauma exposure.

Drawing on a similar theoretical framework, Cecil *et al*. (2014) demonstrated differences in risk factors for CU traits in youths with conduct problems and high versus low internalising problems. In a longitudinal study of 84 children from birth to 13 years, Cecil and colleagues demonstrated that for youths with low (below-median) internalising problems, oxytocin receptor gene methylation at birth was associated with higher CU traits at age 13, as well as lower levels of victimisation throughout childhood. In turn, pre-natal risk factors such as criminal behaviour of the mother were themselves associated with higher levels of oxytocin receptor gene methylation at birth. By contrast, there was no association between oxytocin receptor gene methylation at birth and later CU traits in children with high levels of internalising problems. In this group, interpersonal pre-natal risk factors such as intimate partner violence were associated with later CU traits. Although this study focused on gene methylation rather than genetic (DNA) differences, it provides evidence for potentially different developmental pathways to CU traits within CD, based on levels of internalising problems. Viding and McCrory (2018) further highlight the many complex neurocognitive and environmental risk factors that interact to produce these different developmental pathways to CU.

Third, some researchers have proposed incorporating other dimensions of psychopathic traits into the criteria for CD subtypes (Salekin, 2016). For example, da Silva, Salekin, & Rijo (2019) suggest that grandiose-manipulative and impulsive-irresponsible traits, in addition to CU traits, should be used to subtype CD. The Child Problematic Traits Inventory (CPTI) was developed to assess these traits in children, and a validation study

suggested that the combination of the three psychopathic traits was a stronger predictor of conduct problems than any individual trait alone (Colins, Andershed, Frogner, López-Romero, Veen, & Andershed, 2014; see also López-Romero, Maneiro, Colins, Andershed, & Romero, 2019). It is likely that a greater understanding of psychopathic traits in childhood will lead to improvements in CD subtyping. These attempts do, however, have to be balanced against ethical concerns regarding the downward extension of psychopathy to young children, which can be highly stigmatising (Colins *et al*., 2014).

6.5 Summary and Conclusions

The findings presented in this thesis point to neurobiological differences, as well as differences in parenting environment, that can reliably distinguish CD/HCU and CD/LCU from each other, and from TD, at both the group and individual level. Nonetheless, there remains considerable overlap between the two subtypes in all of the measures presented in this thesis. Future research should address demographic heterogeneity (*e.g*., sex and cultural differences), refine and develop the criteria used for specifying subtypes, and elucidate the different developmental pathways to these subtypes. It is hoped that further efforts in this area will eventually lead to substantially greater therapeutic success for youths with the most severe and complex forms of behavioural disorders.

APPENDIX A: GROUP DIFFERENCES IN GREY MATTER VOLUME USING AAL REGIONS OF INTEREST

A.1 Region of Interest Selection

We selected key regions from the WFU PickAtlas toolbox's AAL atlas (Maldjian *et al*., 2003). Thirteen bilateral regions of interest were selected, and masks created for each region using the WFU PickAtlas toolbox. Frontal areas not defined in the AAL atlas were defined using Brodmann areas from the same toolbox. The selected regions were the anterior insula, amygdala, caudate, orbitofrontal cortex, pallidum, putamen and ventromedial prefrontal cortex. These areas are central to theoretical conceptualisations of psychopathy and antisocial behaviour (Blair, 2013; Glenn & Yang, 2012; Blair *et al*., 2005; Blair, 2007). The orbitofrontal and ventromedial prefrontal cortices were defined as Brodmann areas 11, 12 and 47, and 25 and 32 respectively, in line with Hooker and Knight (2006). The anterior portion of the insula was extracted from the AAL insula mask in MATLAB R2016a, by selecting all voxels lying anterior to the midpoint of the insula. For the classification analyses only, the bilateral regions were split into left and right hemisphere regions. The number of voxels in each region was: 2494 (left anterior insula), 2430 (right anterior insula), 391 (left amygdala), 580 (right amygdala), 2053 (left caudate), 2141 (right caudate), 4165 (left orbitofrontal cortex), 4200 (right orbitofrontal cortex), 528 (left pallidum), 571 (right pallidum), 1999 (left putamen), 2201 (right putamen) and 4270 (ventromedial prefrontal cortex).

A.2 VBM Region of Interest Analyses

There were no significant clusters differentiating any groups in the main models (*i.e.*, covarying for IQ, sex, pubertal stage, site of data collection, and total intracranial volume). Consistent with the whole brain analyses (Chapter 4, section 4.3.2), when controlling for ADHD diagnoses, youths with CD/HCU exhibited decreased grey matter volume relative to TD youths in a small cluster in the right anterior insula (x = 47, y = 0, z = 3, $Z = 3.79$, $k = 4$; $p_{(\text{FWE-corrected})} = .031$).

A.3 Mean Grey Matter Volume Differences in Regions of Interest

Mean grey matter volumes in each AAL region of interest are displayed in **Table A1**. Interestingly, across all regions, mean grey matter volumes were highest for the CD/HCU group and lowest for the CD/LCU group. This is in contrast to the direction of differences in the significant clusters identified in the VBM analyses. While the CD/HCU and CD/LCU groups differed significantly across all regions, the CD/HCU group differed significantly from the TD group only in the right amygdala, right orbitofrontal cortex, left and right pallidum and left putamen. The CD/LCU group, by contrast, differed from the TD group in the left and right anterior insulae, left amygdala and ventromedial prefrontal cortex.

Table A1. Group differences in mean grey matter volumes in AAL regions of interest (mean (95% confidence intervals of the mean))

Region of interest	CD/HCU (*n* = 113)	CD/LCU (*n* = 113)	TD (*n* = 226)	F (*p*), partial η^2
Left anterior insula	0.77 (0.76, 0.78) [a]	0.74 (0.73, 0.75) [b]	0.76 (0.75, 0.77) [a]	8.01 (< *.001*), .03
Right anterior insula	0.63 (0.62, 0.64) [a]	0.61 (0.60, 0.62) [b]	0.63 (0.62, 0.64) [a]	7.43 (*.001*), .03
Left amygdala	0.65 (0.64, 0.66) [a]	0.63 (0.62, 0.64) [b]	0.65 (0.64, 0.66) [a]	6.99 (*.001*), .03
Right amygdala	0.86 (0.85, 0.87) [a]	0.82 (0.80, 0.83) [b]	0.83 (0.82, 0.84) [b]	10.44 (<*.001*), .04
Left caudate	0.78 (0.77, 0.79) [a]	0.74 (0.72, 0.75) [b]	0.76 (0.75, 0.76) [a, b]	16.31 (<*.001*), .07
Right caudate	0.66 (0.64, 0.67) [a]	0.63 (0.62, 0.65) [b]	0.65 (0.64, 0.66) [a, b]	3.97 (*.02*), .02
Left orbitofrontal cortex	0.66 (0.65, 0.67) [a]	0.64 (0.62, 0.65) [b]	0.66 (0.65, 0.66) [a, b]	3.59 (*.03*), .02
Right orbitofrontal cortex	0.51 (0.50, 0.53) [a]	0.49 (0.48, 0.50) [b]	0.50 (0.49, 0.50) [b]	6.55 (*.002*), .03
Left pallidum	0.58 (0.57, 0.59) [a]	0.55 (0.54, 0.56) [b]	0.56 (0.55, 0.57) [b]	7.48 (*.001*), .03
Right pallidum	0.75 (0.74, 0.77) [a]	0.72 (0.71, 0.73) [b]	0.73 (0.72, 0.74) [b]	6.23 (*.002*), .03
Left putamen	0.75 (0.74, 0.76) [a]	0.72 (0.70, 0.73) [b]	0.73 (0.72, 0.74) [b]	7.76 (<*.001*), .03
Right putamen	0.77 (0.76, 0.78) [a]	0.74 (0.72, 0.75) [b]	0.77 (0.76, 0.78) [a]	9.75 (<*.001*), .04
Ventromedial prefrontal cortex	0.83 (0.82, 0.85) [a]	0.80 (0.78, 0.81) [b]	0.82 (0.81, 0.83) [a]	6.81 (*.001*), .03

Notes: AAL=Automated Anatomical Labelling, CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < .05$, Bonferroni corrected)

A.4 Angle-GMLVQ Classifier Performance and Feature Relevance

Angle-GMLVQ classifier performance is shown in **Table A2**. Across all models, performance was extremely similar to the main models presented in Chapter 4 (section 4.3.8).

Table A2. Angle-GMLVQ model performance using AAL regions (mean (95% confidence intervals of the mean))

	Mixed-TD	HCU-TD	LCU-TD	F (*p*), partial η^2	HCU-LCU
Accuracy	0.49 (0.49, 0.50) [a]	0.55 (0.54, 0.55) [b]	0.59 (0.59, 0.60) [c]	764.14 (< *.001*), .34	0.64 (0.63, 0.64)
PPV	0.49 (0.49, 0.50) [a]	0.38 (0.38, 0.39) [b]	0.42 (0.42, 0.43) [c]	934.99 (< *.001*), .38	0.64 (0.63, 0.64)
NPV	0.49 (0.49, 0.50) [a]	0.71 (0.70, 0.71) [b]	0.74 (0.74, 0.74) [c]	6085.36 (< *.001*), .80	0.65 (0.64, 0.65)
TPR	0.50 (0.49, 0.50) [a]	0.55 (0.54, 0.55) [b]	0.58 (0.58, 0.59) [c]	195.98 (< *.001*), .12	0.65 (0.65, 0.66)
TNR	0.49 (0.49, 0.50) [a]	0.55 (0.55, 0.56) [b]	0.60 (0.59, 0.60) [c]	446.49 (< *.001*), .23	0.62 (0.61, 0.63)
MCER	0.51 (0.50, 0.51) [a]	0.45 (0.45, 0.45) [b]	0.41 (0.41, 0.41) [c]	682.64 (< *.001*), .31	0.36 (0.36, 0.37)

Notes: Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. PPV = positive predictive value, NPV = negative predictive value, TPR = true positive rate, TNR = true negative rate, MCER = macro-averaged classification error rate. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < 0.05$, Bonferroni corrected). Note that the HCU-LCU model (column 6) was not included in statistical tests as comparisons between this and other models were not relevant to hypotheses

Feature relevance scores are shown in **Figure A1**. The most relevant features were the left pallidum (HCU-TD model) and right putamen (LCU-TD and HCU-LCU models). Interestingly, the anterior insula was also among the higher scoring features in the HCU-TD model (left anterior insula) and LCU-TD model (right anterior insula). This contrasted with the feature relevance scores presented in Chapter 4 (4.3.9), where

regions with significant clusters in the VBM analyses did not rank highly for relevance. This might be partially explained by the AAL regions capturing more of the group differences observed in the VBM analyses, relative to the meta-analysis-based regions in Chapter 4.

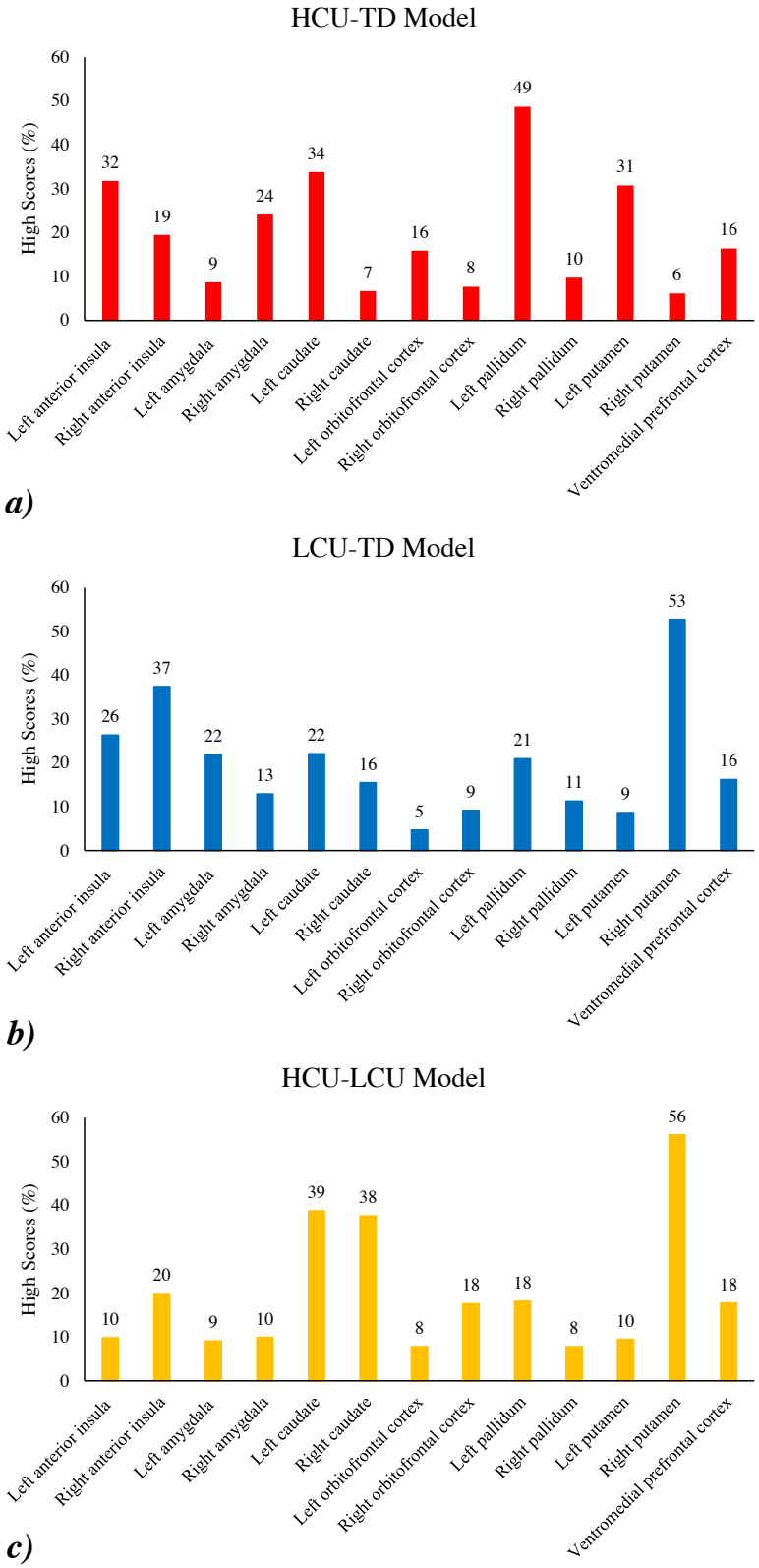


Figure A1. Feature relevance for ***a)*** HCU-TD model, ***b)*** LCU-TD model and ***c)*** HCU-LCU models using AAL regions. Bars show percentage of re-samplings in which feature relevance was in the top 20% of relevance scores across all re-samplings with MCER ≤ 0.40

APPENDIX B: ANGLE-GMLVQ CLASSIFICATION ANALYSES USING GREY MATTER VOLUMES WITH VARIANCE ASSOCIATED WITH ADHD DIAGNOSES REMOVED

The classification analyses presented in Chapter 4 (section 4.3.8 – 4.3.9) were repeated after regressing out variance associated with ADHD diagnoses in addition to IQ, sex, pubertal status, site of data collection and total intracranial volume. Classifier results are shown in **Table B1** and feature relevance scores are shown in **Figure B1**.

As before, the HCU-TD and LCU-TD models both outperformed the Mixed-TD model, indicating that the division into CD/HCU and CD/LCU subtypes corresponded to reliable (albeit small) differences in grey matter volume. Compared to the main classifiers presented in Chapter 4, however, performance dropped for the HCU-TD and HCU-LCU models, while performance for the LCU-TD model was virtually unchanged. These results are consistent with ADHD as a contributing factor to the neurobiological characteristics of the CD/HCU group, but less so to the CD/LCU group. Feature relevance scores remained largely unchanged from the main analyses.

Table B1. Angle-GMLVQ model performance, after regressing out variance associated with ADHD diagnoses (mean (95% confidence intervals of the mean))

	Mixed-TD	HCU-TD	LCU-TD	F (*p*), partial η^2	HCU-LCU
Accuracy	0.49 (0.49, 0.50) [a]	0.52 (0.52, 0.52) [b]	0.58 (0.57, 0.58) [c]	532.89 (< *.001*), .26	0.52 (0.52, 0.52)
PPV	0.49 (0.49, 0.50) [a]	0.36 (0.35, 0.36) [b]	0.40 (0.40, 0.41) [c]	1279.55 (< *.001*), .46	0.36 (0.35, 0.36)
NPV	0.49 (0.49, 0.50) [a]	0.69 (0.68, 0.69) [b]	0.72 (0.72, 0.72) [c]	4653.32 (< *.001*), .76	0.69 (0.68, 0.69)
TPR	0.47 (0.47, 0.48) [a]	0.54 (0.53, 0.54) [b]	0.54 (0.53, 0.54) [b]	141.31 (< *.001*), .09	0.54 (0.53, 0.54)
TNR	0.51 (0.51, 0.52) [a]	0.51 (0.50, 0.51) [a]	0.60 (0.59, 0.60) [b]	363.71 (< *.001*), .20	0.51 (0.50, 0.51)
MCER	0.51 (0.50, 0.51) [a]	0.48 (0.47, 0.48) [b]	0.43 (0.43, 0.44) [c]	370.10 (< *.001*), .20	0.48 (0.47, 0.48)

Notes: Mixed-TD = model classifying youths with conduct disorder with mixed levels of callous unemotional traits and typically developing youths, HCU-TD = model classifying youths with conduct disorder with high levels of callous unemotional traits and typically developing youths, LCU-TD = model classifying youths with conduct disorder and low levels of callous-unemotional traits and typically developing youths. PPV = positive predictive value, NPV = negative predictive value, TPR = true positive rate, TNR = true negative rate, MCER = macro-averaged classification error rate. Groups with different superscript indices differ significantly in post-hoc comparisons ($p < 0.05$, Bonferroni corrected). Note that the HCU-LCU model (column 6) was not included in statistical tests as comparisons between this and other models were not relevant to hypotheses

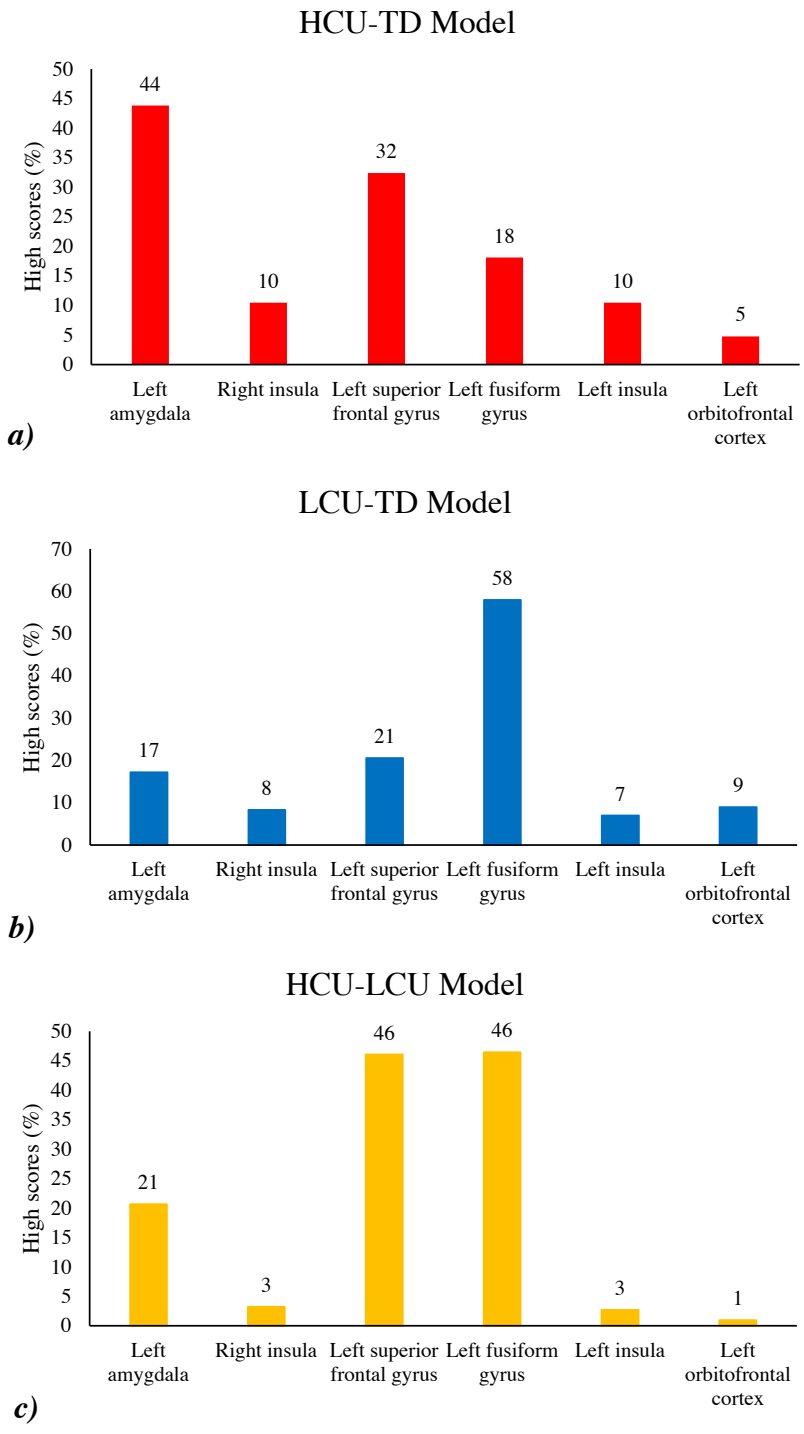


Figure B1. Feature relevance for ***a)*** HCU-TD model, ***b)*** LCU-TD model and ***c)*** HCU-LCU models, with variance associated with ADHD diagnoses removed. Bars show percentage of re-samplings in which feature relevance was in the top 20% of relevance scores across all re-samplings with MCER ≤ 0.40

APPENDIX C: SITE DIFFERENCES IN EMOTION RECOGNITION

ANCOVAs were repeated with data from each site separately. As before, interactions between group, emotion and intensity (*i.e.*, 70% versus 90% expression intensity) were investigated using repeated measures ANCOVAs with group as a between-subjects factor and emotion and intensity as within-subjects factors. Sex was included as a factor of no interest, and IQ and pubertal status (both grand-mean-centred) were entered as covariates. Mean accuracies per group at each site (collapsed across all emotions and intensities) are displayed in **Table C1**. It should be noted that group sizes were very small at some sites; consequently, these findings should be interpreted cautiously. Accuracies for each of the six emotions, per group at each site, are shown in **Figures C1-C10**. After correcting for multiple comparisons, the only significant group by emotion interactions were for the Athens sub-sample. Here, The CD/HCU group performed significantly more poorly than the TD group for anger, disgust and surprise, and significantly more poorly than the CD/LCU group for surprise (all $p < .001$, Bonferroni-corrected).

Although the TD participants usually outperformed those with CD, there were no other consistent patterns across sites, and no noticeable cultural patterns between sites. Again, however, it must be stressed that the sample sizes were often too small for results to be interpreted confidently.

Table C1. Group sizes, clinical characteristics and emotion recognition accuracy per site

Site	Group sizes	Mean CD symptoms	Mean CU score	Mean accuracy (%)	Group main effect (F, *p*, partial η^2)	Group × emotion interaction (F, *p*, partial η^2)
Frankfurt	22 *CD/HCU*	5.45 *CD/HCU*	47.23 *CD/HCU*	67 *CD/HCU*	4.80	1.38
(n = 145)	33 *CD/LCU*	4.76 *CD/LCU*	23.73 *CD/LCU*	69 *CD/LCU*	*.01*	*.22*
	90 *TD*	0.01 *TD*	17.20 *TD*	78 *TD*	.07	.02
Aachen	51 *CD/HCU*	5.59 *CD/HCU*	46.59 *CD/HCU*	69 *CD/HCU*	7.85	0.84
(n = 218)	37 *CD/LCU*	4.95 *CD/LCU*	25.30 *CD/LCU*	62 *CD/LCU*	*< .001*	*.56*
	130 *TD*	0.11 *TD*	15.52 *TD*	76 *TD*	.07	.01
Amsterdam	16 *CD/HCU*	6.44 *CD/HCU*	46.31 *CD/HCU*	76 *CD/HCU*	5.05	1.33
(n = 83)	18 *CD/LCU*	6.00 *CD/LCU*	23.39 *CD/LCU*	74 *CD/LCU*	*.01*	*.24*
	49 *TD*	0.00 *TD*	14.41 *TD*	80 *TD*	0.12	.03
Southampton	16 *CD/HCU*	7.31 *CD/HCU*	47.69 *CD/HCU*	72 *CD/HCU*	3.32	1.30
(n = 95)	25 *CD/LCU*	7.16 *CD/LCU*	23.88 *CD/LCU*	77 *CD/LCU*	*.04*	*.25*
	54 *TD*	0.31 *TD*	16.70 *TD*	84 *TD*	.07	.03
Basel	3 *CD/HCU*	6.67 *CD/HCU*	45.67 *CD/HCU*	87 *CD/HCU*	1.0	0.84
(n = 25)	9 *CD/LCU*	6.66 *CD/LCU*	19.78 *CD/LCU*	81 *CD/LCU*	*.40*	*.55*
	13 *TD*	0.00 *TD*	16.31 *TD*	88 *TD*	.09	.08
Birmingham	19 *CD/HCU*	6.95 *CD/HCU*	51.11 *CD/HCU*	63 *CD/HCU*	5.00	0.61
(n = 124)	24 *CD/LCU*	4.96 *CD/LCU*	23.79 *CD/LCU*	68 *CD/LCU*	*.01*	*.75*
	81 *TD*	0.20 *TD*	16.79 *TD*	81 *TD*	.08	.01
Barcelona	8 *CD/HCU*	3.75 *CD/HCU*	49.13 *CD/HCU*	67 *CD/HCU*	2.31	0.85
(n = 27)	6 *CD/LCU*	3.67 *CD/LCU*	27.33 *CD/LCU*	69 *CD/LCU*	*.12*	*.54*
	13 *TD*	0.00 *TD*	14.31 *TD*	78 *TD*	.18	.08
Bilbao	19 *CD/HCU*	7.58 *CD/HCU*	49.37 *CD/HCU*	77 *CD/HCU*	0.56	1.84
(n = 87)	13 *CD/LCU*	7.23 *CD/LCU*	24.23 *CD/LCU*	75 *CD/LCU*	*.57*	*.08*
	55 *TD*	0.00 *TD*	14.25 *TD*	85 *TD*	.01	.05
Budapest	11 *CD/HCU*	6.45 *CD/HCU*	46.36 *CD/HCU*	78 *CD/HCU*	2.51	0.61
(n = 28)	4 *CD/LCU*	6.75 *CD/LCU*	21.25 *CD/LCU*	86 *CD/LCU*	*.11*	*.72*
	13 *TD*	0.00 *TD*	17.92 *TD*	89 *TD*	0.20	.06
Athens	27 *CD/HCU*	5.52 *CD/HCU*	46.93 *CD/HCU*	56 *CD/HCU*	14.52	3.95
(n = 92)	14 *CD/LCU*	5.07 *CD/LCU*	23.29 *CD/LCU*	82 *CD/LCU*	*< .001*	*< .001*
	51 *TD*	0.00 *TD*	13.75 *TD*	80 *TD*	.26	.09

Notes: CU = callous-unemotional, CD/HCU = conduct disorder with high levels of callous-unemotional traits, CD/LCU = conduct disorder with low levels of callous-unemotional traits, TD = typically developing

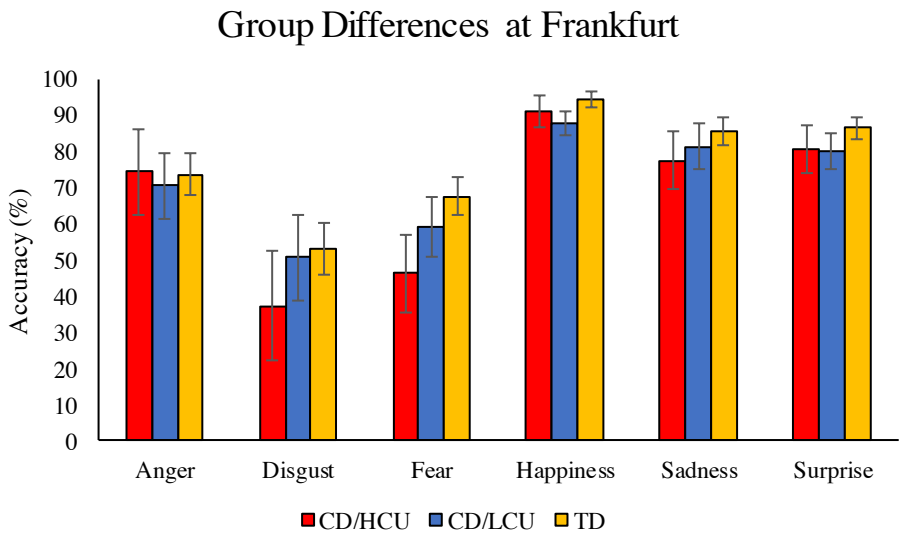


Figure C1. Group differences for each emotion at Frankfurt

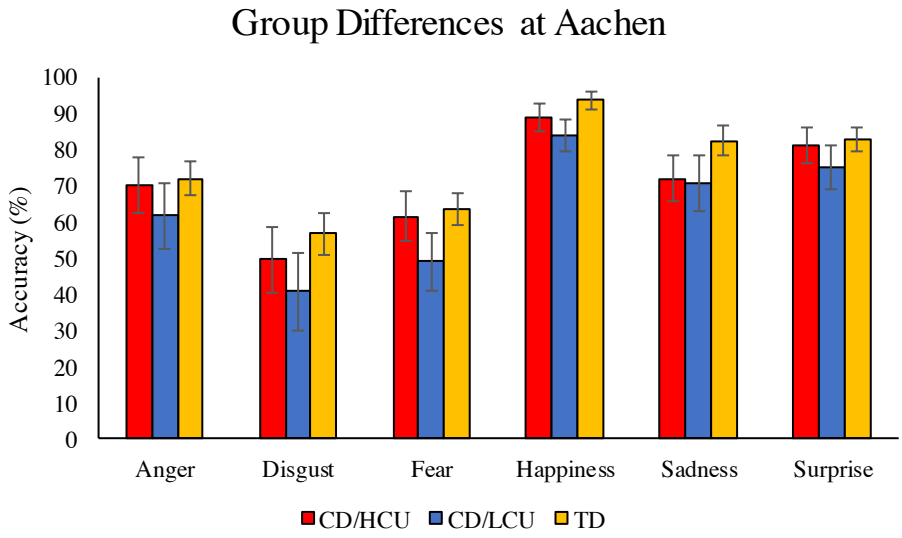


Figure C2. Group differences for each emotion at Aachen

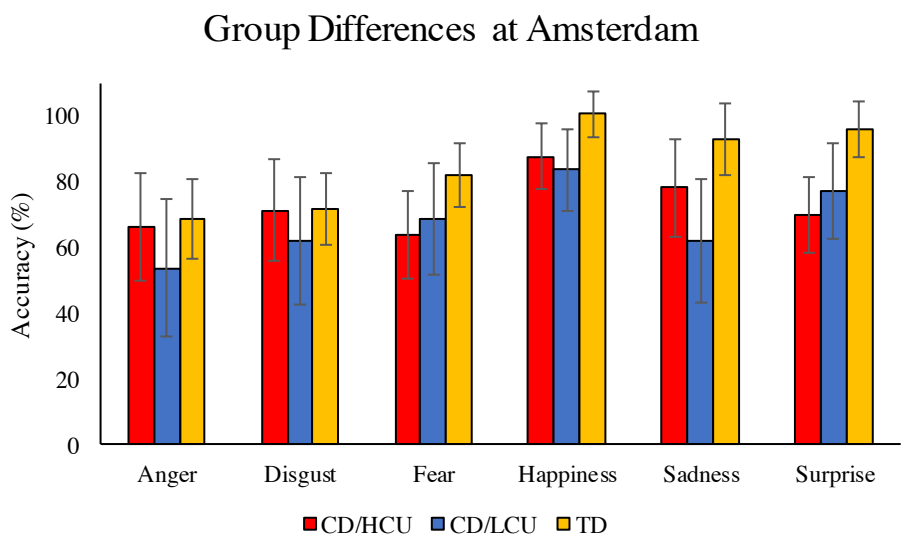


Figure C3. Group differences for each emotion at Amsterdam

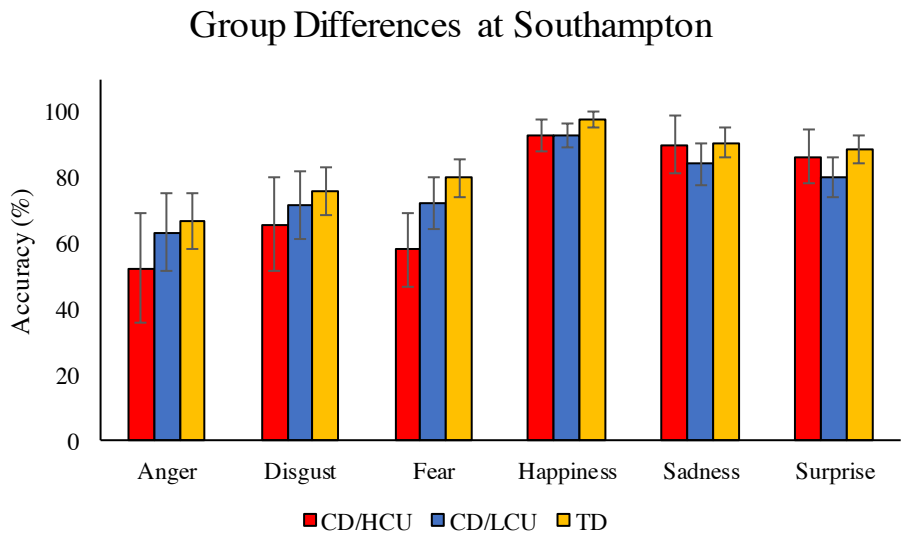


Figure C4. Group differences for each emotion at Southampton

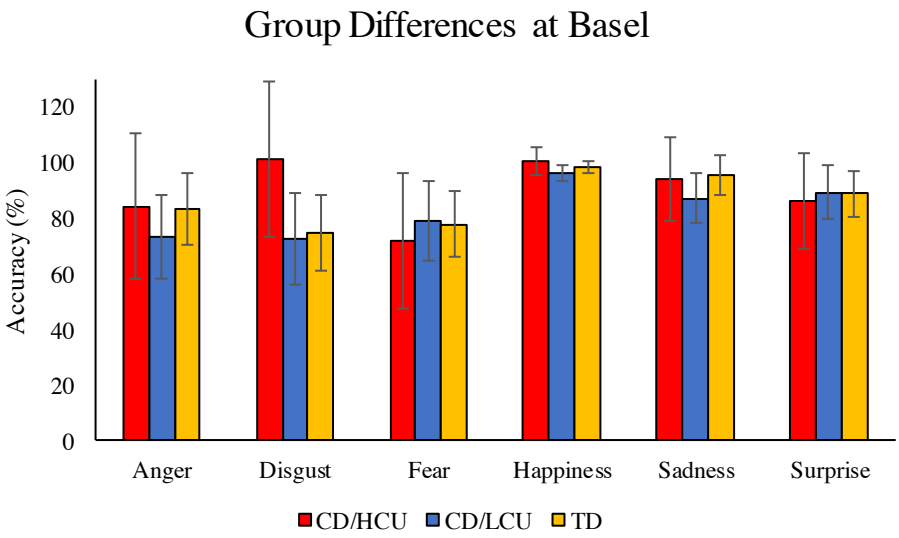


Figure C5. Group differences for each emotion at Basel

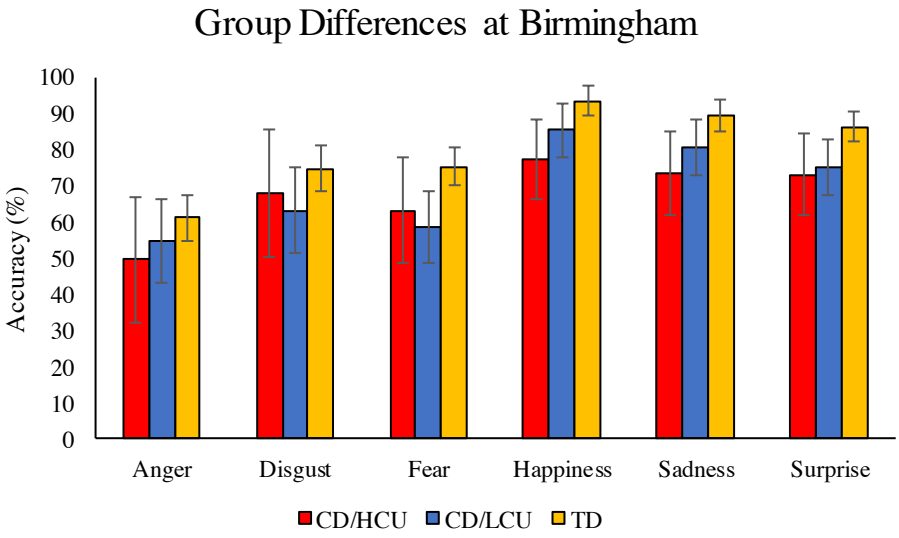


Figure C6. Group differences for each emotion at Birmingham

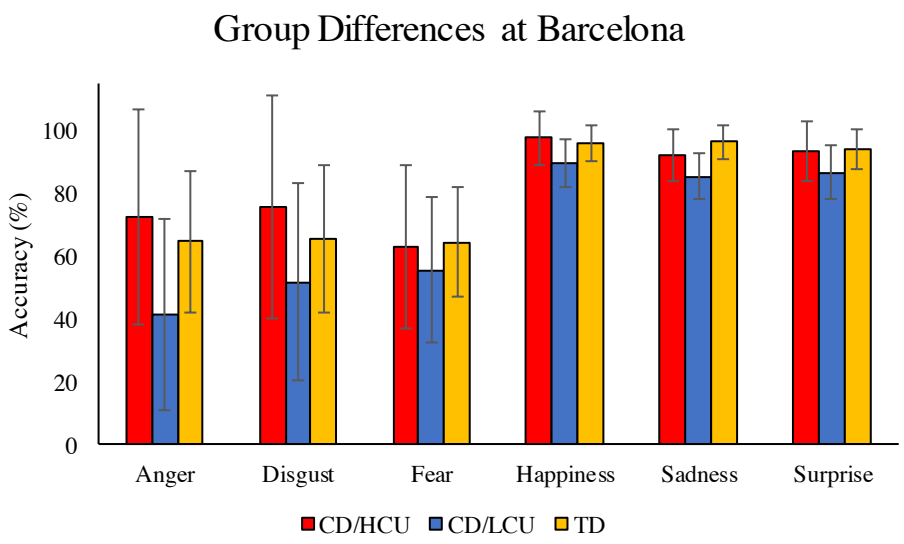


Figure C7. Group differences for each emotion at Barcelona

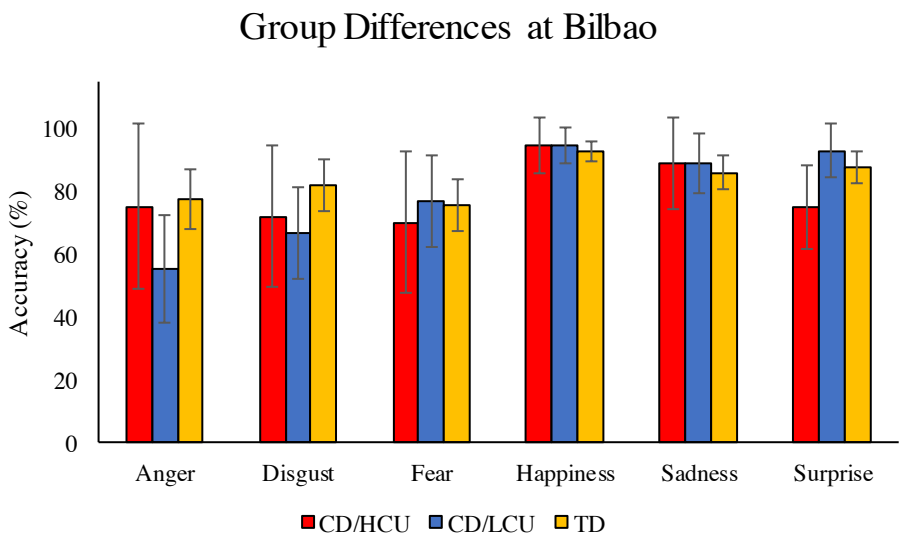


Figure C8. Group differences for each emotion at Bilbao

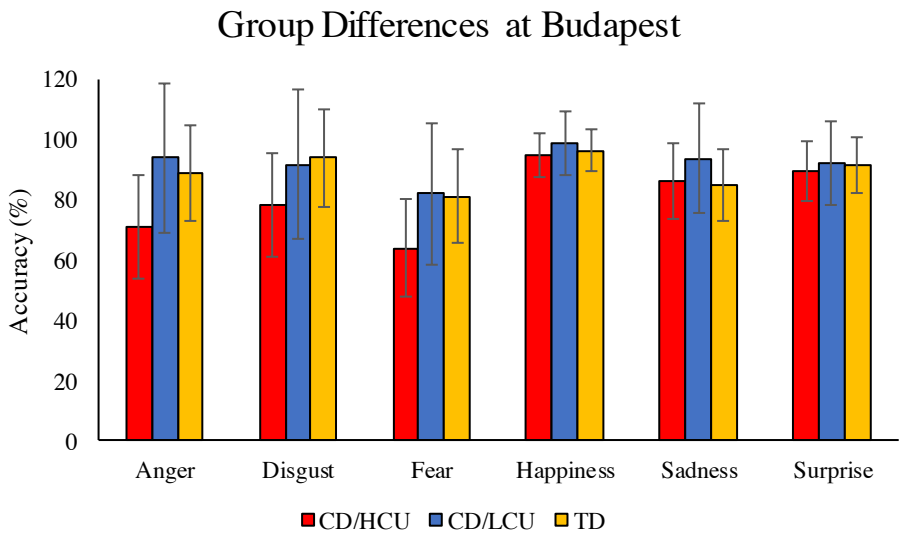


Figure C9. Group differences for each emotion at Budapest

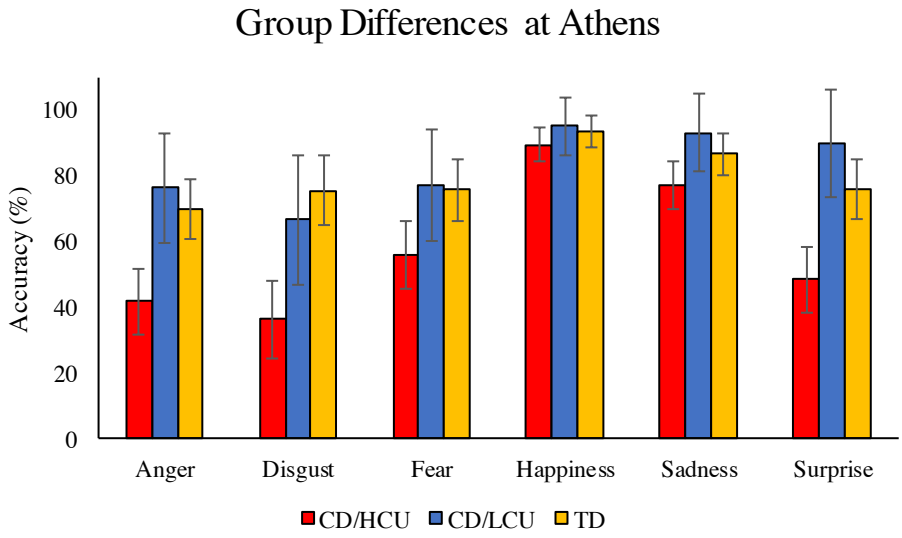


Figure C10. Group differences for each emotion at Athens

REFERENCES

Abu-Akel, A., Heinke, D., Gillespie, S. M., Mitchell, I. J., & Bo, S. (2015). Metacognitive impairments in schizophrenia are arrested at extreme levels of psychopathy: The cut-off effect. *Journal of Abnormal Psychology, 124*(4), 1102-1109. doi: http://dx.doi.org/10.1037/abn0000096

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Alahmadi, H. H., Shen, Y., Fouad, S., Luft, C. D. B., Bentham, P., Kourtzi, Z., & Tino, P. (2016). Classifying cognitive profiles using machine learning with privileged information in mild cognitive impairment. *Frontiers in Computational Neuroscience, 10*, 117. doi: https://doi.org/10.3389/fncom.2016.00117

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Health Disorders: DSM-5* (5th ed.). Washington, DC: American Psychiatric Publishing.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Health Disorders: DSM-IV-TR* (4th ed., Text Revision). Washington, DC: American Psychiatric Publishing.

Andershed, H. A., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: A new assessment tool. In *Psychopaths: Current International Perspectives* (ed. E Blaauw, L Sheridan), pp. 131–58. The Hague: Elsevier.

Aoki, Y., Inokuchi, R., Nakao, T., & Yamasue, H. (2013). Neural bases of antisocial behavior: a voxel-based meta-analysis. *Social Cognitive and Affective Neuroscience, 9*(8), 1223-1231. doi: https://doi.org/10.1093/scan/nst104

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95-113. doi: https://doi.org/10.1016/j.neuroimage.2007.07.007

Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Evaluation measures for ordinal regression. In *Proceedings of the Ninth International Conference on Intelligent System Design and Applications* (pp. 283–287). San Mateo, CA: IEEE Computer Society.

Bardone, A. M., Moffitt, T. E., Caspi, A., Dickson, N., & Silva, P. A. (1996). Adult mental health and social outcomes of adolescent girls with depression and conduct disorder. *Development and Psychopathology, 8*(4), 811-829. doi: https://doi.org/10.1017/S0954579400007446

Barker, E. D., Oliver, B. R., Viding, E., Salekin, R. T., & Maughan, B. (2011). The impact of prenatal maternal risk, fearless temperament and early parenting on adolescent callous-unemotional traits: A 14-year longitudinal investigation. *Journal of Child Psychology and Psychiatry, 52*(8), 878-888. doi: https://doi.org/10.1111/j.1469-7610.2011.02397.xl

Barker, E. D., Walton, E., & Cecil, C. A. (2018). Annual Research Review: DNA methylation as a mediator in the association between risk exposure and child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, *59*(4), 303-322. doi: https://doi.org/10.1111/jcpp.12782

Baskin-Sommers, A. R., Newman, J. P., Sathasivam, N., & Curtin, J. J. (2011). Evaluating the generalizability of a fear deficit in psychopathic African American offenders. *Journal of Abnormal Psychology*, *120*(1), 71-78. doi: http://dx.doi.org/10.1037/a0021225

Beauchaine, T. P., Hinshaw, S. P., & Pang, K. L. (2010). Comorbidity of attention-deficit/hyperactivity disorder and early-onset conduct disorder: Biological, environmental, and developmental mechanisms. *Clinical Psychology: Science and Practice, 17*(4), 327-336. doi: https://doi.org/10.1111/j.1468-2850.2010.01224.x

Bedford, R., Pickles, A., Sharp, H., Wright, N., & Hill, J. (2015). Reduced face preference in infancy: a developmental precursor to callous-unemotional traits? *Biological Psychiatry*, *78*(2), 144-150. doi: https://doi.org/10.1016/j.biopsych.2014.09.022

Bedford, R., Wagner, N. J., Rehder, P. D., Propper, C., Willoughby, M. T., & Mills-Koonce, R. W. (2017). The role of infants' mother-directed gaze, maternal sensitivity, and emotion recognition in childhood callous unemotional behaviours. *European Child & Adolescent Psychiatry*, *26*(8), 947-956. doi: 10.1007/s00787-017-0967-1

Bellucci, G., Feng, C., Camilleri, J., Eickhoff, S. B., & Krueger, F. (2018). The role of the anterior insula in social norm compliance and enforcement: Evidence from coordinate-based and functional connectivity meta-analyses. *Neuroscience & Biobehavioral Reviews, 92*, 378-389. doi: https://doi.org/10.1016/j.neubiorev.2018.06.024

Bergstrøm, H., & Farrington, D.P. (2018). Grandiose-manipulative, callous-unemotional, and daring-impulsive: the prediction of psychopathic traits in adolescence and their outcomes in adulthood. *Journal of Psychopathology and Behavioral Assessment 40*(2), 149-158. doi: https://doi.org/10.1007/s10862-018-9674-6

Blair, R. J. R. (2003). Facial expressions, their communicatory functions and neuro–cognitive substrates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 561-572. doi: https://doi.org/10.1098/rstb.2002.1220

Blair, R. J. R. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry, 71*(6), 727-731. doi: http://dx.doi.org/10.1136/jnnp.71.6.727

Blair, R. J. R. (2004). The roles of orbital frontal cortex in the modulation of antisocial behavior. *Brain and Cognition, 55*(1), 198-208. doi: https://doi.org/10.1016/S0278-2626(03)00276-8

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, *11*(9), 387-392. doi: https://doi.org/10.1016/j.tics.2007.07.003

Blair, R. J. R. (2010). Psychopathy, frustration, and reactive aggression: the role of ventromedial prefrontal cortex. *British Journal of Psychology, 101*(3), 383-399. doi: https://doi.org/10.1348/000712609X418480

Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience, 14*(11), 786. doi: https://doi.org/10.1038/nrn3577

Blair, R. J., & Cipolotti, L. (2000). Impaired social response reversal: A case of 'acquired sociopathy'. *Brain*, *123*(6), 1122-1141. doi: https://doi.org/10.1093/brain/123.6.1122

Blair, J., Mitchell, D., & Blair, K. (2005). The Psychopath: Emotion and the Brain. Oxford, UK: Blackwell Publishing

Blair, R. J. R., & Coles, M. (2000). Expression recognition and behavioural problems in early adolescence. *Cognitive Development*, *15*(4), 421-434. doi: https://doi.org/10.1016/S0885-2014(01)00039-9

Blair, R. J. R., Colledge, E., & Mitchell, D. G. V. (2001). Somatic markers and response reversal: Is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Child Psychology, 29*(6), 499-511. doi: https://doi.org/10.1023/A:1012277125119

Blefari, M. L., Martuzzi, R., Salomon, R., Bello-Ruiz, J., Herbelin, B., Serino, A., & Blanke, O. (2017). Bilateral Rolandic operculum processing underlying heartbeat awareness reflects changes in bodily self-consciousness. *European Journal of Neuroscience, 45*(10), 1300-1312. doi: 10.1111/ejn.13567.

Bowen, K. L., Morgan, J. E., Moore, S. C., & van Goozen, S. H. (2014). Young offenders' emotion recognition dysfunction across emotion intensities: explaining variation using psychopathic traits, conduct disorder and offense severity. *Journal of Psychopathology and Behavioral Assessment, 36*(1), 60-73. doi: https://doi.org/10.1007/s10862-013-9368-z

Bunte, K., Baranowski, E. S., Arlt, W., & Tino, P. (2016). Relevance learning vector quantization in variable dimensional spaces. In *New Challenges in Neural Computation (NC 2), Workshop of the GI-Fachgruppe Neuronale Netze and the German Neural Networks Society in connection to GCPR*, 20-23.

Byrd, A. L., Hawes, S. W., Burke, J. D., Loeber, R., & Pardini, D. A. (2018). Boys with conduct problems and callous-unemotional traits: neural response to reward and punishment and associations with treatment response. *Developmental Cognitive Neuroscience, 30*, 51-59. doi: https://doi.org/10.1016/j.dcn.2017.12.004

Calder, A. J. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13*(5), 699-745. doi: https://doi.org/10.1080/026432996381890

Caputo, A. A., Frick, P. J., & Brodsky, S. L. (1999). Family violence and juvenile sex offending: potential mediating roles of psychopathic traits and negative attitudes toward women. *Criminal Justice and Behavior, 26*(3), 338–356. doi: https://doi.org/10.1177/0093854899026003004

van Casteren, M., & Davis, M. H. (2007). Match: A program to assist in matching the conditions of factorial experiments. *Behavior Research Methods*, *39*(4), 973-978. doi: https://doi.org/10.3758/BF03192992

Cecil, C. A., Lysenko, L. J., Jaffee, S. R., Pingault, J. B., Smith, R. G., Relton, C. L., ... & Barker, E. D. (2014). Environmental risk, Oxytocin Receptor Gene (OXTR) methylation and youth callous-unemotional traits: a 13-year longitudinal study. *Molecular Psychiatry*, *19*(10), 1071. doi: https://doi.org/10.1038/mp.2014.95

Cecil, C. A., McCrory, E. J., Barker, E. D., Guiney, J., & Viding, E. (2018). Characterising youth with callous–unemotional traits and concurrent anxiety: evidence for a high-risk clinical group. *European Child & Adolescent Psychiatry*, *27*(7), 885-898. doi: https://doi.org/10.1007/s00787-017-1086-8

Centifanti, L. C., Shaw, H., Atherton, K. J., Thomson, N. D., MacLellan, S., & Frick, P. J. (2019). CAPE for measuring callous-unemotional traits in disadvantaged families: a cross-sectional validation study [version 1; peer review: awaiting peer review]. *F1000Research*, 8, 1027. doi: https://doi.org/10.12688/f1000research.19605.1)

Chinchilla, M. A., & Kosson, D. S. (2016). Psychopathic traits moderate relationships between parental warmth and adolescent antisocial and other high-risk behaviors. *Criminal Justice and Behavior, 43*(6), 722-738. doi: 10.1177/0093854815617216

Christian, R. E., Frick, P. J., Hill, N. L., Tyler, L., & Frazer, D. R. (1997). Psychopathy and conduct problems in children: II. Implications for subtyping children with conduct problems. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(2), 233-241. doi: https://doi.org/10.1097/00004583-199702000-00014

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: perennial issues and an emerging reconceptualization. *Annual Review of Psychology, 58*, 227–57. doi: https://doi.org/10.1146/annurev.psych.57.102904.190200

Clark, J. E., & Frick, P. J. (2018). Positive parenting and callous-unemotional traits: their association with school behavior problems in young children. *Journal of Clinical Child & Adolescent Psychology, 47*(sup1), S242-S254. doi: https://doi.org/10.1080/15374416.2016.1253016

Cleckley, H. (1941). The Mask of Sanity (1st ed.). Saint Louis: Mosby

Cohn, M. D., Viding, E., McCrory, E., Pape, L., van den Brink, W., Doreleijers, T. A., ... & Popma, A. (2016). Regional grey matter volume and concentration in at-risk adolescents: Untangling associations with callous-unemotional traits and conduct disorder symptoms. *Psychiatry Research: Neuroimaging*, *254*, 180-187. doi: https://doi.org/10.1016/j.pscychresns.2016.07.003

Coid, J., & Ullrich, S. (2010). Antisocial personality disorder is on a continuum with psychopathy. *Comprehensive Psychiatry, 51*(4), 426-433. doi: https://doi.org/10.1016/j.comppsych.2009.09.006

Colins, O. F., Andershed, H., Frogner, L., López-Romero, L., Veen, V., & Andershed, A. K. (2014). A new measure to assess psychopathic personality in children: The Child Problematic Traits Inventory. *Journal of Psychopathology and Behavioral Assessment*, *36*(1), 4-21. doi: https://doi.org/10.1007/s10862-013-9385-y

Cope, L. M., Ermer, E., Gaudet, L. M., Steele, V. R., Eckhardt, A. L., Arbabshirani, M. R., … & Kiehl, K. A. (2014). Abnormal brain structure in youth who commit homicide. *NeuroImage: Clinical, 4*, 800-807. doi: https://doi.org/10.1016/j.nicl.2014.05.002

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. doi: 10.1001/archgenpsychiatry.2009.85

Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*(1), 59-70. doi: https://doi.org/10.1038/nrn2555

Crum, K. I., Waschbusch, D. A., Bagner, D. M., & Coxe, S. (2015). Effects of callous–unemotional traits on the association between parenting and child conduct problems. *Child Psychiatry & Human Development, 46*(6), 967-980. doi: https://doi.org/10.1007/s10578-015-0535-1

Dadds, M. R., Allen, J. L., McGregor, K., Woolgar, M., Viding, E., & Scott, S. (2014). Callous-unemotional traits in children and mechanisms of impaired eye contact during expressions of love: A treatment target? *Journal of Child Psychology and Psychiatry, 55*(7), 771-780. doi: https://doi.org/10.1111/jcpp.12155

Dadds, M. R., Allen, J. L., Oliver, B. R., Faulkner, N., Legge, K., Moul, C., ... & Scott, S. (2012). Love, eye contact and the developmental origins of empathy v. psychopathy. *The British Journal of Psychiatry, 200*(3), 191-196. doi: https://doi.org/10.1192/bjp.bp.110.085720

Dadds, M. R., Hunter, K., Hawes, D. J., Frost, A. D., Vassallo, S., Bunn, P., ... & El Masry, Y. (2008). A measure of cognitive and affective empathy in children using parent ratings. *Child Psychiatry and Human Development*, *39*(2), 111-122. doi: https://doi.org/10.1007/s10578-007-0075-4

Dadds, M. R., Jambrak, J., Pasalich, D., Hawes, D. J., & Brennan, J. (2011). Impaired attention to the eyes of attachment figures and the developmental origins of psychopathy. *Journal of Child Psychology and Psychiatry, 52*(3), 238-245. doi: https://doi.org/10.1111/j.1469-7610.2010.02323.x

Dadds, M. R., & Salmon, K. (2003). Punishment insensitivity and parenting: Temperament and learning as interacting risks for antisocial behavior. *Clinical Child and Family Psychology Review*, *6*(2), 69-86. doi: https://doi.org/10.1023/A:1023762009877

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: https://doi.org/10.1016/j.neubiorev.2012.08.006

De Brito, S. A., Mechelli, A., Wilke, M., Laurens, K. R., Jones, A. P., Barker, G. J., ... & Viding, E. (2009). Size matters: Increased grey matter in boys with conduct problems and callous–unemotional traits. *Brain*, *132*(4), 843-852. doi: https://doi.org/10.1093/brain/awp011

Dotterer, H. L., Hyde, L. W., Swartz, J. R., Hariri, A. R., & Williamson, D. E. (2017). Amygdala reactivity predicts adolescent antisocial behavior but not callous-unemotional traits. *Developmental Cognitive Neuroscience*, 24, 84-92. doi: https://doi.org/10.1016/j.dcn.2017.02.008

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

Eekhout, I., de Vet, H. C., Twisk, J. W., Brand, J. P., de Boer, M. R., & Heymans, M. W. (2014). Missing data in a multi-item instrument were best handled by multiple imputation at the item score level. *Journal of Clinical Epidemiology, 67*(3), 335-342. doi: https://doi.org/10.1016/j.jclinepi.2013.09.009

Ekman, P., & Friesen, W.V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologists

Elizur, Y., & Somech, L. Y. (2018). Callous-unemotional traits and effortful control mediate the effect of parenting intervention on preschool conduct problems. *Journal of Abnormal Child Psychology*, *46*, 1631-1642. doi: https://doi.org/10.1007/s10802-018-0412-z

Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. *Cerebral Cortex*, *10*(3), 308-317. doi: https://doi.org/10.1093/cercor/10.3.308

Elster, A.D. (2018). *Questions and Answers in MRI*. http://mriquestions.com

Enebrink, P., Andershed, H., & Långström, N. (2005). Callous–unemotional traits are associated with clinical severity in referred boys with conduct problems. *Nordic Journal of Psychiatry, 59*(6), 431-440. doi: https://doi.org/10.1080/08039480500360690

Eriksson, S. H., Free, S. L., Thom, M., Symms, M. R., Martinian, L., Duncan, J. S., & Sisodiya, S. M. (2009). Quantitative grey matter histological measures do not correlate with grey matter probability values from in vivo MRI in the temporal lobe. *Journal of Neuroscience Methods, 181*(1), 111-118. doi: https://doi.org/10.1016/j.jneumeth.2009.05.001

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006a). Psychometric properties of the Alabama parenting questionnaire. *Journal of Child and Family Studies*, *15*(5), 595-614. doi: https://doi.org/10.1007/s10826-006-9036-y

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006b). Callous-unemotional traits in a community sample of adolescents. *Assessment, 13*(4), 454-469. doi: https://doi.org/10.1177/1073191106287354

Fairchild, G., Van Goozen, S. H., Calder, A. J., Stollery, S. J., & Goodyer, I. M. (2009). Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. *Journal of Child Psychology and Psychiatry*, *50*(5), 627-636. doi: https://doi.org/10.1111/j.1469-7610.2008.02020.x

Fairchild, G., Hagan, C. C., Walsh, N. D., Passamonti, L., Calder, A. J., & Goodyer, I. M. (2013). Brain structure abnormalities in adolescent girls with conduct disorder. *Journal of Child Psychology and Psychiatry, 54*(1), 86-95. doi: https://doi.org/10.1111/j.1469-7610.2012.02617.x

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*(43). doi: https://doi.org/10.1038/s41572-019-0095-y

Fairchild, G., Stobbe, Y., Van Goozen, S. H., Calder, A. J., & Goodyer, I. M. (2010). Facial expression recognition, fear conditioning, and startle modulation in female subjects with conduct disorder. *Biological Psychiatry*, *68*(3), 272-279. doi: https://doi.org/10.1016/j.biopsych.2010.02.019

Fairchild, G., Toschi, N., Hagan, C. C., Goodyer, I. M., Calder, A. J., & Passamonti, L. (2015). Cortical thickness, surface area, and folding alterations in male youths with conduct disorder and varying levels of callous–unemotional traits. *NeuroImage: Clinical, 8*, 253-260. doi: https://doi.org/10.1016/j.nicl.2015.04.018

Falk, A. E., & Lee, S. S. (2012). Parenting behavior and conduct problems in children with and without attention-deficit/hyperactivity disorder (ADHD): moderation by callous-unemotional traits. *Journal of Psychopathology and Behavioral Assessment, 34*(2), 172-181. doi: https://doi.org/10.1007/s10862-011-9268-z

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 903-911. doi: https://doi.org/10.1016/j.neubiorev.2010.10.009

Fanti, K. A. (2013). Individual, social, and behavioral factors associated with co-occurring conduct problems and callous-unemotional traits. *Journal of Abnormal Child Psychology, 41*(5), 811-824. doi: 10.1007/s10802-013-9726-z

Fanti, K. A., Demetriou, C. A., & Kimonis, E. R. (2013). Variants of callous-unemotional conduct problems in a community sample of adolescents. *Journal of Youth and Adolescence, 42*(7), 964-979. doi: https://doi.org/10.1007/s10964-013-9958-9

Fanti, K. A., Frick, P. J., & Georgiou, S. (2009). Linking callous-unemotional traits to instrumental and non-instrumental forms of aggression. *Journal of Psychopathology and Behavioral Assessment, 31*(4), 285. doi: 10.1007/s10862-008-9111-3

Fanti, K. A., Konikou, K., Cohn, M., Popma, A., & Brazil, I. A. (2019). Amygdala functioning during threat acquisition and extinction differentiates antisocial subtypes. *Journal of Neuropsychology*. doi: https://doi.org/10.1111/jnp.12183

Fanti, K. A., Panayiotou, G., Lazarou, C., Michael, R., & Georgiou, G. (2016). The better of two evils? Evidence that children exhibiting continuous conduct problems high or low on callous–unemotional traits score on opposite directions on physiological and behavioral measures of fear. *Development and Psychopathology, 28*(1), 185-198. doi: https://doi.org/10.1017/S0954579415000371

Faraone, S. V., Biederman, J., Jetton, J. G., & Tsuang, M. T. (1997). Attention deficit disorder and conduct disorder: longitudinal evidence for a familial subtype. *Psychological Medicine, 27*(2), 291-300. doi: https://doi.org/10.1097/00004583-199804000-00022

Farrington, D. P., & Loeber, R. (2000). Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health, 10*(2), 100-122. doi: https://doi.org/10.1002/cbm.349

FemNAT-CD – The Neurobiology and Treatment of Adolescent Female Conduct Disorder: The Central Role of Emotion Processing. (n.d.). Retrieved from http://www.femnat-cd.eu/index.htm

Flach, P. (2012). *Machine Learning: the Art and Science of Algorithms that Make Sense of Data*. UK: Cambridge University Press.

Flom, M., White, D., Ganiban, J., & Saudino, K. J. (2019). Longitudinal Links Between Callous-Unemotional Behaviors and Parenting in Early Childhood: A Genetically Informed Design. *Journal of the American Academy of Child & Adolescent Psychiatry*. doi: https://doi.org/10.1016/j.jaac.2019.03.013

Follette, W. C., & Houts, A. C. (1996). Models of scientific progress and the role of theory in taxonomy development: A case study of the DSM. *Journal of Consulting and Clinical Psychology*, *64*(6), 1120-1132. doi: http://dx.doi.org/10.1037/0022-006X.64.6.1120

Fontaine, N., Carbonneau, R., Vitaro, F., Barker, E. D., & Tremblay, R. E. (2009). Research review: A critical review of studies on the developmental trajectories of antisocial behavior in females. *Journal of Child Psychology and Psychiatry*, *50*(4), 363-385. doi: https://doi.org/10.1111/j.1469-7610.2008.01949.x

Fontaine, N. M., McCrory, E. J., Boivin, M., Moffitt, T. E., & Viding, E. (2011). Predictors and outcomes of joint trajectories of callous–unemotional traits and conduct problems in childhood. *Journal of Abnormal Psychology*, *120*(3), 730-742. doi: 10.1037/a0022620

Fouad, S., & Tino, P. (2012). Adaptive Metric Learning Vector Quantization for Ordinal Classification. *Neural Computation, 24*(11), 2825-2825. doi: 10.1162/NECO_a_00358

Freitag, C. M., Konrad, K., Stadler, C., De Brito, S. A., Popma, A., Herpertz, S. C., ... & Fairchild, G. (2018). Conduct disorder in adolescent females: current state of research and study design of the FemNAT-CD consortium. *European Child & Adolescent Psychiatry, 27*(9), 1077-1093. doi: 10.1007/s00787-018-1172-6

Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology, 31*(4), 457-470. doi: 10.1023/A:1023899703866

Frick, P. J., & Ellis, M. (1999). Callous-unemotional traits and subtypes of conduct disorder. *Clinical Child and Family Psychology Review, 2*(3), 149-168. doi: 10.1023/A:1021803005547

Frick, P. J., Kimonis, E. R., Dandreaux, D. M., & Farrell, J. M. (2003). The 4 year stability of psychopathic traits in non-referred youth. *Behavioral Sciences & the Law, 21*(6), 713-736. doi: 10.1002/bsl.568

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Annual research review: a developmental psychopathology approach to understanding callous-unemotional traits in children and adolescents with serious conduct problems. *Journal of Child Psychology and Psychiatry, 55*(6), 532-548. doi: 10.1111/jcpp.12152

Frick, P. J., Stickle, T. R., Dandreaux, D. M., Farrell, J. M., & Kimonis, E. R. (2005). Callous–unemotional traits in predicting the severity and stability of conduct problems and delinquency. *Journal of Abnormal Child Psychology, 33*(4), 471-487. doi: 10.1007/s10648-005-5728-9

Frick, P. J., & Viding, E. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology, 21*(04), 1111-1131. doi:10.1017/S0954579409990071

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: https://doi.org/10.1111/j.1469-7610.2007.01862.x

Frogner, L., Andershed, A. K., & Andershed, H. (2018). Psychopathic personality works better than CU traits for predicting fearlessness and ADHD symptoms among children with conduct problems. *Journal of Psychopathology and Behavioral Assessment, 40*(1), 26-39. doi: https://doi.org/10.1007/s10862-018-9651-0

Gaser, C., & Dahnke, R. (2016). CAT-a computational anatomy toolbox for the analysis of structural MRI data. *Organisation for Human Brain Mapping*, *2016*. Available at http://www.neuro.uni-jena.de/hbm2016/GaserHBM2016.pdf

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., ... & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience, 2*(10), 861-863. doi: https://doi.org/10.1038/13158

Glenn, A. L., & Yang, Y. (2012). The potential role of the striatum in antisocial behavior and psychopathy. *Biological Psychiatry*, *72*(10), 817-822. doi: https://doi.org/10.1016/j.biopsych.2012.04.027

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... & Rapoport, J. L. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, *101*(21), 8174-8179. Doi: https://doi.org/10.1073/pnas.0402680101

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Friston, K. J., & Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage, 14*(1), 21-36. doi: https://doi.org/10.1006/nimg.2001.0786

Guyer, A. E., McClure, E. B., Adler, A. D., Brotman, M. A., Rich, B. A., Kimes, A. S., ... & Leibenluft, E. (2007). Specificity of facial expression labeling deficits in childhood psychopathology. *Journal of Child Psychology and Psychiatry*, *48*(9), 863-871. doi: |https://doi.org/10.1111/j.1469-7610.2007.01758.x

Hare, R. D. (2003). *The Hare Psychopathy Checklist-Revised: (PCL-R) 2nd edition*. London, UK: Pearson Assessment

Hare, R. D. (1999). *Without Conscience: The Disturbing World of the Psychopaths Among us*. NY: Guilford Press.

Henry, J., Dionne, G., Viding, E., Vitaro, F., Brendgen, M., Tremblay, R. E., & Boivin, M. (2018). Early warm-rewarding parenting moderates the genetic contributions to callous–unemotional traits in childhood. *Journal of Child Psychology and Psychiatry*, *59*(12), 1282-1288. doi: https://doi.org/10.1111/jcpp.12918|

Henry, J., Pingault, J. B., Boivin, M., Rijsdijk, F., & Viding, E. (2016). Genetic and environmental aetiology of the dimensions of Callous-Unemotional traits. *Psychological Medicine*, *46*(2), 405-414. doi: https://doi.org/10.1017/S0033291715001919

Herpers, P. C., Klip, H., Rommelse, N. N., Greven, C. U., & Buitelaar, J. K. (2016). Associations between high callous–unemotional traits and quality of life across youths with non-conduct disorder diagnoses. *European Child & Adolescent Psychiatry, 25*(5), 547-555. doi: 10.1007/s00787-015-0766-5

Högström, J., Enebrink, P., & Ghaderi, A. (2013). The moderating role of child callous-unemotional traits in an Internet-based parent-management training program. *Journal of Family Psychology, 27*(2), 314. doi: 10.1037/a0031883

Hooker, C.I, & Knight, R.T. (2006). *The Role of Lateral Orbitofrontal Cortex in the Inhibitory Control of Emotion*. Available from http://scholar.harvard.edu/files/chooker/files/hooker_knight_2002.pdf

Humayun, S., Kahn, R. E., Frick, P. J., & Viding, E. (2014). Callous-unemotional traits and anxiety in a community sample of 7-year-olds. *Journal of Clinical Child & Adolescent Psychology, 43*(1), 36-42. doi: https://doi.org/10.1080/15374416.2013.814539

Hutton, C., Draganski, B., Ashburner, J., & Weiskopf, N. (2009). A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. *Neuroimage, 48*(2), 371-380. doi: https://doi.org/10.1016/j.neuroimage.2009.06.043

Hyde, L. W., Waller, R., Trentacosta, C. J., Shaw, D. S., Neiderhiser, J. M., Ganiban, J. M., ... & Leve, L. D. (2016). Heritable and nonheritable pathways to early callous-unemotional behaviors. *American Journal of Psychiatry, 173*(9), 903-910. doi: https://doi.org/10.1176/appi.ajp.2016.15111381

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York, NY: Springer.

Janssen, R. J., Mourão-Miranda, J., & Schnack, H. G. (2018). Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(9), 798-808. doi: https://doi.org/10.1016/j.bpsc.2018.04.004

Japkowicz, N. (2000). Learning from imbalanced data sets: a comparison of various strategies. In *AAAI Workshop on Learning from Imbalanced Data Sets* (Vol. 68, pp. 10-15).

Jezior, K.L., McKenzie, M.E., & Lee, S.S. (2016). Narcissism and callous-unemotional traits prospectively predict child conduct problems. *Journal of Clinical Child & Adolescent Psychology, 45*(5), 579-590. doi: 10.1080/15374416.2014.982280

Jones, A. P., Laurens, K. R., Herba, C. M., Barker, G. J., & Viding, E. (2009). Amygdala hypoactivity to fearful faces in boys with conduct problems and callous-unemotional traits. *American Journal of Psychiatry, 166*(1), 95-102. doi: https://doi.org/10.1176/appi.ajp.2008.07071050

Kahn, R. E., Frick, P. J., Youngstrom, E., Findling, R. L., & Youngstrom, J. K. (2012). The effects of including a callous–unemotional specifier for the diagnosis of conduct disorder. *Journal of Child Psychology and Psychiatry*, *53*(3), 271-282. doi: https://doi.org/10.1111/j.1469-7610.2011.02463.x

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *Neuroimage*, *122*, 6-19. doi: https://doi.org/10.1016/j.neuroimage.2015.07.082

Karpman, B. (1948). The myth of the psychopathic personality. *American Journal of Psychiatry*, *104*(9), 523-534.

Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., ... & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980-988. doi: https://doi.org/10.1097/00004583-199707000-00021

Kersten, L., Vriends, N., Steppan, M., Raschle, N. M., Praetzlich, M., Oldenhof, H., ... & Martinelli, A. (2017). Community violence exposure and conduct problems in children and adolescents with conduct disorder and healthy controls. *Frontiers in Behavioral Neuroscience*, *11*, 219. doi: https://doi.org/10.3389/fnbeh.2017.00219

Kessler, R. C., Foster, C. L., Saunders, W. B., & Stang, P. E. (1995). Social consequences of psychiatric disorders, I: Educational attainment. *American Journal of Psychiatry, 152*(7), 1026-1032.

Kimonis, E. R., Fanti, K., Goldweber, A., Marsee, M. A., Frick, P. J., & Cauffman, E. (2014). Callous-unemotional traits in incarcerated adolescents. *Psychological Assessment, 26*(1), 227-237. doi: http://dx.doi.org/10.1037/a0034585

Kochanska, G., Forman, D. R., Aksan, N., & Dunbar, S. B. (2005). Pathways to conscience: early mother–child mutually responsive orientation and children's moral emotion, conduct, and cognition. *Journal of Child Psychology and Psychiatry, 46*(1), 19-34. doi: 10.1111/j.1469-7610.2004.00348.x

Kochanska, G., Kim, S., Boldt, L. J., & Yoon, J. E. (2013). Children's callous-unemotional traits moderate links between their positive relationships with parents at preschool age and externalizing behavior problems at early school age. *Journal of Child Psychology and Psychiatry, 54*(11), 1251-1260. doi: https://doi.org/10.1111/jcpp.12084

Kohavi, R., & Provost, F. (1998) Glossary of Terms: Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. *Machine Learning, 30*, 271-274. Available online at: http://robotics.stanford.edu/~ronnyk/glossary.html

Kohls, G., Baumann, S., Gundlach, M., Scharke, W., Bernhard, A., Martinelli, A., ... & Jansen, L. (2019). Investigating sex differences in emotion recognition, learning, and regulation among youths with conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*. doi: https://doi.org/10.1016/j.jaac.2019.04.003

Kohonen, T. (1995). *Self-Organising Maps*. Germany: Springer-Verlag.

Kosson, D. S., Lorenz, A. R., & Newman, J. P. (2006). Effects of comorbid psychopathy on criminal offending and emotion processing in male offenders with antisocial personality disorder. *Journal of Abnormal Psychology, 115*(4), 798. doi: 10.1037/0021-843X.115.4.798

Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology, 72*(5), 341-372. doi: https://doi.org/10.1016/j.pneurobio.2004.03.006

Lahey, B. B., Loeber, R., Quay, H. C., Applegate, B., Shaffer, D., Waldman, I., ... & Dulcan, M. K. (1998). Validity of DSM-IV subtypes of conduct disorder based on age of onset. *Journal of the American Academy of Child & Adolescent Psychiatry, 37*(4), 435-442. doi: https://doi.org/10.1097/00004583-199804000-00022

Liu, Y., & De, A. (2015). Multiple imputation by fully conditional specification for dealing with missing data in a large epidemiologic study. *International Journal of Statistics in Medical Research, 4*(3), 287. doi: 10.6000/1929-6029.2015.04.03.7

Lockwood, P. L., Sebastian, C. L., McCrory, E. J., Hyde, Z. H., Gu, X., De Brito, S. A., & Viding, E. (2013). Association of callous traits with reduced neural response to others' pain in children with conduct problems. *Current Biology*, *23*(10), 901-905. doi: https://doi.org/10.1016/j.cub.2013.04.018

Lockwood-Estrin, G., Thom, M., Focke, N. K., Symms, M. R., Martinian, L., Sisodiya, S. M., ... & Eriksson, S. H. (2012). Correlating 3T MRI and histopathology in patients undergoing epilepsy surgery. *Journal of Neuroscience Methods, 205*(1), 182-189. doi: https://doi.org/10.1016/j.jneumeth.2011.12.014

Loeber, R., Burke, J. D., Lahey, B. B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorder: a review of the past 10 years, part I. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(12), 1468-1484. doi: https://doi.org/10.1097/00004583-200012000-00007

López-Romero, L., Maneiro, L., Colins, O. F., Andershed, H., & Romero, E. (2019). Psychopathic Traits in Early Childhood: Further Multi-informant Validation of the Child Problematic Traits Inventory (CPTI). *Journal of Psychopathology and Behavioral Assessment*, 1-9. doi: https://doi.org/10.1007/s10862-019-09735-0

Lozier, L. M., Cardinale, E. M., VanMeter, J. W., & Marsh, A. A. (2014). Mediation of the relationship between callous-unemotional traits and proactive aggression by amygdala response to fear among children with conduct problems. *JAMA psychiatry, 71*(6), 627-636. doi: 10.1001/jamapsychiatry.2013.4540

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *The Journal of Abnormal and Social Psychology*, *55*(1), 6-10.

Lynam, D. R., Caspi, A., Moffitt, T. E., Loeber, R., & Stouthamer-Loeber, M. (2007). Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. *Journal of Abnormal Psychology, 116*(1), 155. doi: 10.1037/0021-843X.116.1.155

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, *19*(3), 1233-1239. doi: https://doi.org/10.1016/S1053-8119(03)00169-1

Marsh, A. A., Finger, E. C., Mitchell, D. G., Reid, M. E., Sims, C., Kosson, D. S., ... & Blair, R. J. R. (2008). Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. *American Journal of Psychiatry, 165*(6), 712-720. doi: https://doi.org/10.1176/appi.ajp.2007.07071145

Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. *Journal of Child Psychology and Psychiatry, 45*(3), 609-621. doi: https://doi.org/10.1111/j.1469-7610.2004.00250.x

McDonough-Caplan, H., Klein, D. N., & Beauchaine, T. P. (2018). Comorbidity and continuity of depression and conduct problems from elementary school to adolescence. *Journal of Abnormal Psychology, 127*(3), 326. doi: http://dx.doi.org/10.1037/abn0000339

Mechelli, A., Price, C. J., Friston, K. J., & Ashburner, J. (2005). Voxel-based morphometry of the human brain: methods and applications. *Current Medical Imaging Reviews, 1*(2), 105-113. doi: https://doi.org/10.2174/1573405054038726

Meffert, H., Thornton, L. C., Tyler, P. M., Botkin, M. L., Erway, A. K., Kolli, V., ... & Blair, R. J. R. (2018). Moderation of prior exposure to trauma on the inverse relationship between callous-unemotional traits and amygdala responses to fearful expressions: an exploratory study. *Psychological Medicine*, *48*(15), 2541-2549. doi: https://doi.org/10.1017/S0033291718000156

Millon, T., Simonsen, E., Birket-Smith, M., & Davis, R. D. (Eds.). (2002). *Psychopathy: Antisocial, Criminal, and Violent Behavior*. NY: Guilford Press.

Mills-Koonce, W. R., Wagner, N. J., Willoughby, M. T., Stifter, C., Blair, C., Granger, D. A., & Family Life Project Key Investigators. (2015). Greater fear reactivity and psychophysiological hyperactivity among infants with later conduct problems and callous-unemotional traits. *Journal of Child Psychology and Psychiatry, 56*(2), 147-154. doi: https://doi.org/10.1111/jcpp.12289|

Moffitt, T. E., Caspi, A., Dickson, N., Silva, P., & Stanton, W. (1996). Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and Psychopathology, 8*(2), 399-424. doi: https://doi.org/10.1017/S0954579400007161

Molinuevo, B., Pardo, Y., & Torrubia, R. (2011). Psychometric analysis of the Catalan version of the Alabama Parenting Questionnaire (APQ) in a community sample. *The Spanish Journal of Psychology*, *14*(2), 944-955. doi: https://doi.org/10.5209/rev_SJOP.2011.v14.n2.40

Muñoz, L. C., Pakalniskiene, V., & Frick, P. J. (2011). Parental monitoring and youth behavior problems: moderation by callous-unemotional traits over time. *European Child & Adolescent Psychiatry*, *20*(5), 261-269. doi: 10.1007/s00787-011-0172-6

Muratori, P., Lochman, J. E., Lai, E., Milone, A., Nocentini, A., Pisano, S., ... & Masi, G. (2016). Which dimension of parenting predicts the change of callous unemotional traits in children with disruptive behavior disorder? *Comprehensive Psychiatry*, *69*, 202-210. doi: https://doi.org/10.1016/j.comppsych.2016.06.002

Murrie, D. C., Marcus, D. K., Douglas, K. S., Lee, Z., Salekin, R. T., & Vincent, G. (2007). Youth with psychopathy features are not a discrete class: a taxometric analysis. *Journal of Child Psychology and Psychiatry, 48*(7), 714–723. doi: https://doi.org/10.1111/j.1469-7610.2007.01734.x

Namkung, H., Kim, S. H., & Sawa, A. (2017). The Insula: An Underestimated Brain Area in Clinical Neuroscience, Psychiatry, and Neurology. *Trends in Neurosciences, 40*(4), 200-207. doi:10.1016/j.tins.2017.02.002

Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the National Comorbidity Survey Replication. *Psychological Medicine, 36*(5), 699-710. doi: https://doi.org/10.1017/S0033291706007082

Noordermeer, S. D., Luman, M., & Oosterlaan, J. (2016). A systematic review and meta-analysis of neuroimaging in oppositional defiant disorder (ODD) and conduct disorder (CD) taking attention-deficit hyperactivity disorder (ADHD) into account. *Neuropsychology Review*, *26*(1), 44-72. doi: https://doi.org/10.1007/s11065-015-9315-8

O'Brien, B. S., & Frick, P. J. (1996). Reward dominance: associations with anxiety, conduct problems, and psychopathy in children. *Journal of Abnormal Child Psychology, 24*(2), 223-240. doi: https://doi.org/10.1007/BF01441486

O'Keefe, J. J., Carr, A., & McQuaid, P. (1998). Conduct disorder in girls and boys: The identification of distinct psychosocial profiles. *The Irish Journal of Psychology*, *19*(2-3), 368-385. doi: https://doi.org/10.1080/03033910.1998.10558197

O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. UK: Penguin Random House.

O'Nions, E., Lima, C. F., Scott, S. K., Roberts, R., McCrory, E. J., & Viding, E. (2017). Reduced laughter contagion in boys at risk for psychopathy. *Current Biology*, *27*(19), 3049-3055. doi: https://doi.org/10.1016/j.cub.2017.08.062

Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews, 36*(4), 1140-1152. doi: https://doi.org/10.1016/j.neubiorev.2012.01.004

Oxford, M., Cavell, T. A., & Hughes, J. N. (2003). Callous/unemotional traits moderate the relation between ineffective parenting and child externalizing problems: A partial replication and extension. *Journal of Clinical Child and Adolescent Psychology, 32*(4), 577-585. doi: https://doi.org/10.1207/S15374424JCCP3204_10

Pajer, K., Leininger, L., & Gardner, W. (2010). Recognition of facial affect in girls with conduct disorder. *Psychiatry Research*, *175*(3), 244-251. doi: https://doi.org/10.1016/j.psychres.2009.06.003

Pardini, D. A., Lochman, J. E., & Powell, N. (2007). The development of callous-unemotional traits and antisocial behavior in children: Are there shared and/or unique predictors? *Journal of Clinical Child and Adolescent Psychology*, *36*(3), 319-333. doi: https://doi.org/10.1080/15374410701444215

Pasalich, D. S., Dadds, M. R., Hawes, D. J., & Brennan, J. (2011). Do callous-unemotional traits moderate the relative importance of parental coercion versus warmth in child conduct problems? An observational study. *Journal of Child Psychology and Psychiatry, 52*(12), 1308-1315. doi: https://doi.org/10.1111/j.1469-7610.2011.02435.x|

Petersen, A. C., Crockett, L., Richards, M., & Boxer, A. (1988). A self-report measure of pubertal status: Reliability, validity, and initial norms. *Journal of Youth and Adolescence*, *17*(2), 117-133. doi: https://doi.org/10.1007/BF01537962

Pinel, P. (1806/2017). *A Treatise on Insanity*. (D.D Davis, Trans.) Miami, FL: HardPress.

Poeppl, T. B., Donges, M. R., Mokros, A., Rupprecht, R., Fox, P. T., Laird, A. R., ... & Eickhoff, S. B. (2019). A view behind the mask of sanity: meta-analysis of aberrant brain activity in psychopaths. *Molecular Psychiatry, 24*(3), 463-470. doi: https://doi.org/10.1038/s41380-018-0122-5

Prince's Trust (2010). *The Cost of Exclusion: Counting the Cost of Youth Disadvantage in the UK*. Available from: www.princestrust.org .uk/pdf/COE_full_report.pdf.

Rahimi, A., & Recht, B. (2017). Reflections on Random Kitchen Sinks. *Acceptance speech for Test-of-time awards, NIPS 2017*. Text available from http://www.argmin.net/2017/12/05/kitchen-sinks/

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., ... & Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, *32*(2), 159-171. doi: 10.1002/ab.20115

Raschle, N. M., Menks, W. M., Fehlbaum, L. V., Tshomba, E., & Stadler, C. (2015). Structural and functional alterations in right dorsomedial prefrontal and left insular cortex co-localize in adolescents with aggressive behaviour: an ALE meta-analysis. *PLoS One, 10*(9), e0136553. doi: https://doi.org/10.1371/journal.pone.0136553

Raschle, N. M., Menks, W. M., Fehlbaum, L. V., Steppan, M., Smaragdi, A., Gonzalez-Madruga, K., … & Stadler, C. (2018). Callous-unemotional traits and brain structure: Sex-specific effects in anterior insula of typically-developing youths. *NeuroImage: Clinical*, *17*, 856-864. doi: https://doi.org/10.1016/j.nicl.2017.12.015

Ray, J. V., Frick, P. J., Thornton, L. C., Wall Myers, T. D., Steinberg, L., & Cauffman, E. (2017). Callous–unemotional traits predict self-reported offending in adolescent boys: the mediating role of delinquent peers and the moderating role of parenting practices. *Developmental Psychology*, *53*(2), 319-328. doi: http://dx.doi.org/10.1037/dev0000210

Raznahan, A., Shaw, P., Lalonde, F., Stockman, M., Wallace, G. L., Greenstein, D., ... & Giedd, J. N. (2011). How does your cortex grow? *Journal of Neuroscience, 31*(19), 7174-7177. doi: https://doi.org/10.1523/JNEUROSCI.0054-11.2011

Rehder, P. D., Mills-Koonce, W. R., Willoughby, M. T., Garrett-Peters, P., Wagner, N. J., & Family Life Project Key Investigators. (2017). Emotion Recognition Deficits among Children with Conduct Problems and Callous-Unemotional Behaviors. *Early Childhood Research Quarterly*, *41*, 174-183. doi: https://doi.org/10.1016/j.ecresq.2017.07.007

Richers, J. E., & Cicchetti, D. (1993). Mark Twain meets DSM-III-R: Conduct disorder, development, and the concept of harmful dysfunction. *Development and Psychopathology*, *5*(1-2), 5-29. doi: https://doi.org/10.1017/S0954579400004235

Robins, L. N. (1978). Sturdy childhood predictors of adult antisocial behaviour: Replications from longitudinal studies. *Psychological medicine, 8*(4), 611-622. doi: https://doi.org/10.1017/S0033291700018821

Rogers, J. C., & De Brito, S. A. (2016). Cortical and subcortical gray matter volume in youths with conduct problems: a meta-analysis. *JAMA Psychiatry, 73*(1), 64-72. doi: 10.1001/jamapsychiatry.2015.2423

Rogers, J. C., Gonzalez-Madruga, K., Kohls, G., Baker, R. H., Clanton, R. L., Pauli, R., ... & Smaragdi, A. (2019). White Matter Microstructure in Youths With Conduct Disorder: Effects of Sex and Variation in Callous Traits. *Journal of the American Academy of Child & Adolescent Psychiatry*. doi: https://doi.org/10.1016/j.jaac.2019.02.019

Romeo, R., Knapp, M., & Scott, S. (2006). Economic cost of severe antisocial behaviour in children–and who pays it. *The British Journal of Psychiatry, 188*(6), 547–553. doi:10.1192/bjp.bp.104.007625

Rowe, R., Maughan, B., Moran, P., Ford, T., Briskman, J., & Goodman, R. (2010). The role of callous and unemotional traits in the diagnosis of conduct disorder. *Journal of Child Psychology and Psychiatry*, *51*(6), 688-695. doi: https://doi.org/10.1111/j.1469-7610.2009.02199.x

Safer, D. J., & Malever, M. (2000). Stimulant treatment in Maryland public schools. *Pediatrics, 106*(3), 533-539. doi: 10.1542/peds.106.3.533

Salihovic, S., Kerr, M., Özdemir, M., & Pakalniskiene, V. (2012). Directions of effects between adolescent psychopathic traits and parental behavior. *Journal of Abnormal Child Psychology*, *40*(6), 957-969. doi: 10.1007/s10802-012-9623-x

Salekin, R. T. (2016). Psychopathy in childhood: Toward better informing the DSM–5 and ICD-11 conduct disorder specifiers. *Personality Disorders: Theory, Research, and Treatment, 7*(2), 180-191. http://dx.doi.org/10.1037/per0000150

Sato, A., & Yamada, K. (1996). Generalized learning vector quantization. In *Advances in Neural Information Processing* (pp. 423-429).

Sato, J. R., de Oliveira-Souza, R., Thomaz, C. E., Basilio, R., Bramati, I. E., Amaro Jr, E., ... & Moll, J. (2011). Identification of psychopathic individuals using pattern classification of MRI images. *Social Neuroscience, 6*(5-6), 627-639. doi: https://doi.org/10.1080/17470919.2011.562687

Schaich Borg, J., Kahn, R. E., Sinnott-Armstrong, W., Kurzban, R., Robinson, P. H., & Kiehl, K. A. (2013). Subcomponents of psychopathy have opposing correlations with punishment judgments. *Journal of Personality and Social Psychology, 105*(4), 667-687. doi: 10.1037/a0033485

Scerbo, A., Raine, A., O'Brien, M., Chan, C. J., Rhee, C., & Smiley, N. (1990). Reward dominance and passive avoidance learning in adolescent psychopaths. *Journal of Abnormal Child Psychology, 18*(4), 451-463. doi: https://doi.org/10.1007/BF00917646

Schepman, K., Taylor, E., Collishaw, S., & Fombonne, E. (2012). Face emotion processing in depressed children and adolescents with and without comorbid conduct disorder. *Journal of Abnormal Child Psychology*, *40*(4), 583-593. doi: https://doi.org/10.1007/s10802-011-9587-2

Schneider, P., Biehl, M., & Hammer, B. (2009). Adaptive relevance in matrices in learning vector quantization. *Neural Computation 21*(12), 3532-3561. doi: https://doi.org/10.1162/neco.2009.11-08-908

Schwenck, C., Mergenthaler, J., Keller, K., Zech, J., Salehi, S., Taurines, R., ... & Freitag, C. M. (2012). Empathy in children with autism and conduct disorder: Group-specific profiles and developmental aspects. *Journal of Child Psychology and Psychiatry*, *53*(6), 651-659. doi: https://doi.org/10.1111/j.1469-7610.2011.02499.x|

Sculley, D., Snoek, J., Wiltschko, A., & Rahimi, A. (2018). Winner's curse? On pace, progress, and empirical rigor. *Workshop Track – ICLR 2018*. Available at https://openreview.net/forum?id=rJWF0Fywf

Sebastian, C. L., De Brito, S. A., McCrory, E. J., Hyde, Z. H., Lockwood, P. L., Cecil, C. A., & Viding, E. (2016). Grey matter volumes in children with conduct problems and varying levels of callous-unemotional traits. *Journal of Abnormal Child Psychology, 44*(4), 639-649. doi: https://doi.org/10.1007/s10802-015-0073-0

Sebastian, C. L., McCrory, E. J., Cecil, C. A., Lockwood, P. L., De Brito, S. A., Fontaine, N. M., & Viding, E. (2012). Neural responses to affective and cognitive theory of mind in children with conduct problems and varying levels of callous-unemotional traits. *Archives of General Psychiatry, 69*(8), 814-822. doi: 10.1001/archgenpsychiatry.2011.2070

Sebastian, C. L., McCrory, E. J., Dadds, M. R., Cecil, C. A. M., Lockwood, P. L., Hyde, Z. H., ... & Viding, E. (2014). Neural responses to fearful eyes in children with conduct problems and varying levels of callous–unemotional traits. *Psychological Medicine*, *44*(1), 99-109. doi: https://doi.org/10.1017/S0033291713000482

Seijas, R., Servera, M., García-Banda, G., Barry, C. T., & Burns, G. L. (2018). Evaluation of a four-item DSM–5 Limited Prosocial Emotions specifier scale within and across settings with Spanish children. *Psychological Assessment, 30*(4), 474-485. doi: http://dx.doi.org/10.1037/pas0000496

Sethi, A., O'Nions, E., McCrory, E., Bird, G., & Viding, E. (2018). An fMRI investigation of empathic processing in boys with conduct problems and varying levels of callous-unemotional traits. *NeuroImage: Clinical*, *18*, 298-304. doi: https://doi.org/10.1016/j.nicl.2018.01.027

Short, R. M., Sonuga-Barke, E. J., Adams, W. J., & Fairchild, G. (2016). Does comorbid anxiety counteract emotion recognition deficits in conduct disorder? *Journal of Child Psychology and Psychiatry*, *57*(8), 917-926. doi: https://doi.org/10.1111/jcpp.12544

da Silva, D. R., Salekin, R. T., & Rijo, D. (2019). Psychopathic severity profiles: A latent profile analysis in youth samples with implications for the diagnosis of conduct disorder. *Journal of Criminal Justice*, *60*, 74-83. doi: https://doi.org/10.1016/j.jcrimjus.2018.12.003

Simonoff, E., Elander, J., Holmshaw, J., Pickles, A., Murray, R., & Rutter, M. (2004). Predictors of antisocial personality: Continuities from childhood to adult life. *The British Journal of Psychiatry, 184*(2), 118-127. doi: https://doi.org/10.1192/bjp.184.2.118

Steele, V. R., Rao, V., Calhoun, V. D., & Kiehl, K. A. (2017). Machine learning of structural magnetic resonance imaging predicts psychopathic traits in adolescent offenders. *NeuroImage, 145*, 265-273. doi: https://doi.org/10.1016/j.neuroimage.2015.12.013

Sterzer, P., Stadler, C., Poustka, F., & Kleinschmidt, A. (2007). A structural neural deficit in adolescents with conduct disorder and its association with lack of empathy. *NeuroImage, 37*(1), 335-342. doi: https://doi.org/10.1016/j.neuroimage.2007.04.043

Stevens, D., Charman, T., & Blair, R. J. R. (2001). Recognition of emotion in facial expressions and vocal tones in children with psychopathic tendencies. *The Journal of Genetic Psychology*, *162*(2), 201-211. doi: https://doi.org/10.1080/00221320109597961

Sully, K., Sonuga-Barke, E. J., & Fairchild, G. (2015). The familial basis of facial emotion recognition deficits in adolescents with conduct disorder and their unaffected relatives. *Psychological Medicine*, *45*(9), 1965-1975. doi: https://doi.org/10.1017/S0033291714003080

Trouton, A., Spinath, F.M., & Plomin, R. (2002). Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behaviour problems in childhood. *Twin Research, 5*(5), 444–448. doi: https://doi.org/10.1375/twin.5.5.444

Viding, E., Blair, R. J. R., Moffitt, T. E., & Plomin, R. (2005). Evidence for substantial genetic risk for psychopathy in 7-year-olds. *Journal of Child Psychology and Psychiatry, 46*(6), 592-597. doi: 10.1111/j.1469-7610.2004.00393.x

Viding, E., Fontaine, N. M., Oliver, B. R., & Plomin, R. (2009). Negative parental discipline, conduct problems and callous–unemotional traits: Monozygotic twin differences study. *The British Journal of Psychiatry, 195*(5), 414-419. doi: https://doi.org/10.1192/bjp.bp.108.061192

Viding, E., Jones, A. P., Paul, J. F., Moffitt, T. E., & Plomin, R. (2008). Heritability of antisocial behaviour at 9: do callous-unemotional traits matter? *Developmental Science*, *11*(1), 17-22. doi: https://doi.org/10.1111/j.1467-7687.2007.00648.x

Viding, E., & McCrory, E. J. (2018). Understanding the development of psychopathy: progress and challenges. *Psychological Medicine, 48*(4), 566-577. doi: https://doi.org/10.1017/S0033291717002847

Viding, E., & McCrory, E. (2019). Towards understanding atypical social affiliation in psychopathy. *The Lancet Psychiatry, 6*(5), 437-444. doi: https://doi.org/10.1016/S2215-0366(19)30049-5

Viding, E., Sebastian, C. L., Dadds, M. R., Lockwood, P. L., Cecil, C. A., De Brito, S. A., & McCrory, E. J. (2012). Amygdala response to preattentive masked fear in children with conduct problems: the role of callous-unemotional traits. *American Journal of Psychiatry, 169*(10), 1109-1116. doi: https://doi.org/10.1176/appi.ajp.2012.12020191

Wall, T. D., Frick, P. J., Fanti, K. A., Kimonis, E. R., & Lordos, A. (2016). Factors differentiating callous-unemotional children with and without conduct problems. *Journal of Child Psychology and Psychiatry, 57*(8), 976-983. doi: https://doi.org/10.1111/jcpp.12569|

Wallace, G. L., White, S. F., Robustelli, B., Sinclair, S., Hwang, S., Martin, A., & Blair, R. J. R. (2014). Cortical and subcortical abnormalities in youths with conduct disorder and elevated callous-unemotional traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(4), 456-465. doi: https://doi.org/10.1016/j.jaac.2013.12.008

Waller, R., Gardner, F., & Hyde, L. W. (2013). What are the associations between parenting, callous–unemotional traits, and antisocial behavior in youth? A systematic review of evidence. *Clinical Psychology Review, 33*(4), 593-608. doi: https://doi.org/10.1016/j.cpr.2013.03.001

Waller, R., Gardner, F., Shaw, D. S., Dishion, T. J., Wilson, M. N., & Hyde, L. W. (2015). Callous-unemotional behavior and early-childhood onset of behavior problems: The role of parental harshness and warmth. *Journal of Clinical Child & Adolescent Psychology, 44*(4), 655-667. doi: https://doi.org/10.1080/15374416.2014.886252

Waller, R., Gardner, F., Viding, E., Shaw, D. S., Dishion, T. J., Wilson, M. N., & Hyde, L. W. (2014). Bidirectional associations between parental warmth, callous unemotional behavior, and behavior problems in high-risk preschoolers. *Journal of Abnormal Child Psychology*, *42*(8), 1275-1285. doi: https://doi.org/10.1007/s10802-014-9871-z

Waller, R., Hyde, L. W., Klump, K. L., & Burt, S. A. (2018). Parenting is an environmental predictor of callous-unemotional traits and aggression: A monozygotic twin differences study. *Journal of the American Academy of Child & Adolescent Psychiatry, 57*(12), 955-963. doi: https://doi.org/10.1016/j.jaac.2018.07.882

Waller, R., Shaw, D. S., Neiderhiser, J. M., Ganiban, J. M., Natsuaki, M. N., Reiss, D., ... & Hyde, L. W. (2017). Toward an understanding of the role of the environment in the development of early callous behavior. *Journal of Personality, 85*(1), 90-103. doi: 10.1111/jopy.12221

Waller, R., Trentacosta, C. J., Shaw, D. S., Neiderhiser, J. M., Ganiban, J. M., Reiss, D., ... & Hyde, L. W. (2016). Heritable temperament pathways to early callous–unemotional behaviour. *The British Journal of Psychiatry, 209*(6), 475–482. doi: 10.1192/bjp.bp.116.181503

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV)*. San Antonio, TX: Psychological Corporation.

White, S. F., VanTieghem, M., Brislin, S. J., Sypher, I., Sinclair, S., Pine, D. S., ... & Blair, R. J. R. (2015). Neural correlates of the propensity for retaliatory behavior in youths with disruptive behavior disorders. *American Journal of Psychiatry, 173*(3), 282-290. doi: https://doi.org/10.1176/appi.ajp.2015.15020250

Whitwell, J. L. (2009). Voxel-based morphometry: an automated technique for assessing structural changes in the brain. *Journal of Neuroscience, 29*(31), 9661-9664. doi: https://doi.org/10.1523/JNEUROSCI.2160-09.2009

Wilke, M., Holland, S. K., Altaye, M., & Gaser, C. (2008). Template-O-Matic: a toolbox for creating customized pediatric templates. *NeuroImage*, *41*(3), 903-913. doi: https://doi.org/10.1016/j.neuroimage.2008.02.056

Wootton, J. M., Frick, P. J., Shelton, K. K., & Silverthorn, P. (1997). Ineffective parenting and childhood conduct problems: the moderating role of callous-unemotional traits. *Journal of Consulting and Clinical Psychology*, *65*(2), 301. doi: http://dx.doi.org/10.1037/0022-006X.65.2.292.b

Zhang, J., Cao, W., Wang, M., Wang, N., Yao, S., & Huang, B. (2018a). Multivoxel pattern analysis of structural MRI in children and adolescents with conduct disorder. *Brain Imaging and Behavior,* 1-8. doi: https://doi.org/10.1007/s11682-018-9953-6

Zhang, J., Liu, W., Zhang, J., Wu, Q., Gao, Y., Jiang, Y., ... & Huang, B. (2018b). Distinguishing adolescents with conduct disorder from typically developing youngsters based on pattern classification of brain structural MRI. *Frontiers in Human Neuroscience, 12*, *152*. doi: 10.3389/fnhum.2018.00152

Zwaanswijk, W., van Geel, M., Andershed, H., Fanti, K. A., & Vedder, P. (2018). Variants of Psychopathy and the Dependence on Gender, Age, and Ethnic Background. *Journal of Personality Disorders*, *32*(6), 721-737. doi: https://doi.org/10.1521/pedi_2017_31_319

Zwirs, B. W., Burger, H., Schulpen, T. W., & Buitelaar, J. K. (2006). Different treatment thresholds in non-western children with behavioral problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(4), 476-483. doi: https://doi.org/10.1097/01.chi.0000192251.46023.5a