Neurobehavioral correlates of disinhibitory psychopathology

Carl Delfin

Centre for Ethics, Law and Mental Health
Department of Psychiatry and Neurochemistry
Institute of Neuroscience and Physiology
Sahlgrenska Academy
University of Gothenburg
Gothenburg, 2020

UNIVERSITY OF GOTHENBURG

Cover illustration by Pia Moberg.

The inside of the front cover shows Växjö, Sweden, as depicted in *Svecia Antiqua et Hodierna*; a collection of copper engravings made between 1690 and 1710 by architect and draftsman Erik Dahlbergh. Across the small hill above the horse-drawn carriage, one can glimpse the old lunatic asylum, built around 1540. An even older asylum, located closer to the cathedral, was first mentioned in 1318.

The inside of the back cover shows the outer wall of Sweden's first clinic for mentally disordered offenders, known colloquially as 'the Special'. Beginning in 1795, the old lunatic asylum in Växjö was relocated and new buildings were erected further away from the city center. After much deliberation, the Special opened its doors in 1906; it remained a national clinic for mentally disordered offenders for almost 80 years, until it was demolished and replaced by more modern buildings in 1983-84. Printed with permission from photographer Hans Runesson.

This thesis is typeset using LaTeX and Yihui Xie's R package bookdown. The typeface you are reading now is Vollkorn, designed by Friedrich Althausen. Fira Sans, designed by Erik Spiekermann and Ralph du Carrois, is used for sans-serif text.

Neurobehavioral correlates of disinhibitory psychopathology

carl.delfin@gu.se

ISBN 978-91-8009-088-9 (PRINT)
ISBN 978-91-8009-089-6 (PDF)
http://hdl.handle.net/2077/65469

Printed in Borås, Sweden, 2020
Printed by Stema Specialtryck AB

Mein Werk bestehe aus zwei Teilen: aus dem, der hier vorliegt, und aus alledem, was ich *nicht* geschrieben habe. Und gerade dieser zweite Teil ist der Wichtige.

Ludwig Wittgenstein

Neurobehavioral correlates of disinhibitory psychopathology

Carl Delfin

Centre for Ethics, Law and Mental Health, Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden

Abstract

Disinhibitory psychopathology refers to maladaptive behavioral expressions stemming from problems with impulse control. Despite a robust association with antisocial and criminal behavior, knowledge about the neurobehavioral correlates of disinhibitory psychopathology is still lacking. The aims of this thesis were to (1) quantify the prevalence of disinhibitory psychopathology, (2) examine associations between disinhibitory psychopathology and neurocognitive function as well as (3) brain structure and function, and (4) explore how neurobehavioral variables associated with disinhibitory psychopathology may be used in the prediction of recidivism. Four studies, with participants recruited among offenders, mentally disordered offenders, and young adults of the general population, were conducted. Each study used a different, specific set of methods, including clinical and self-report assessments, file review, and register data, as well as neurocognitive tasks probing inhibitory control and neuroimaging techniques such as electrophysiological recordings and structural brain scans.

The prevalence of disinhibitory psychopathology was similar to or even higher than previous national and international estimates. Disinhibitory psychopathology was associated with neurocognitive impairments, most prominently an impulsive approach to planning and problem-solving and a reduced capacity for inhibitory control, and with neurobiological alterations in regions involved in monitoring and evaluation of behavior, inhibitory control, working memory, and attention. Finally, a set of neurobehavioral variables associated with disinhibitory psychopathology increased the accuracy of recidivism prediction.

In conclusion, this thesis confirms the importance of disinhibitory psychopathology as a clinical construct. It adds to a scarce literature on mentally disordered offenders and provides much needed evidence of specific neurobehavioral correlates that may be used to guide the development of novel diagnostic frameworks and treatment strategies, and that may be useful for targeted interventions in forensic settings.

Keywords: Disinhibition, psychopathology, crime, recidivism, mentally disordered offenders, event-related potentials, magnetic resonance imaging

ISBN 978-91-8009-088-9 (PRINT)
ISBN 978-91-8009-089-6 (PDF)
http://hdl.handle.net/2077/65469

Sammanfattning på svenska

Disinhibition — en nedsatt förmåga att hämma impulser — är starkt kopplad till olika antisociala och kriminella beteenden. Ökad kunskap om specifika neurobiologiska och beteendemässiga korrelat till disinhibiton är nödvändig för att möjliggöra individanpassad vård och behandling, samt för fortsatt utveckling av innovativa diagnostiska ramverk. I den här avhandlingen undersöks (1) förekomsten av olika former av maladaptiva beteenden präglade av disinhibition, (2) kopplingen mellan maladaptiva beteenden präglade av disinhibition och neurokognitiva funktioner samt (3) hjärnans struktur och funktion och slutligen (4) huruvida olika beteendemässiga och neurobiologiska variabler kopplade till disinhibition kan användas för att förbättra träffsäkerheten i bedömningar om risken för återfall i brott. Avhandlingen omfattar fyra delstudier, där deltagare rekryterades bland kriminalvårdsklienter, rättspsykiatriska patienter, samt bland unga vuxna i den allmänna befolkningen. Varje enskild studie använde en bred uppsättning av olika metoder, till exempel klinisk bedömning, självrapportering, journalgranskning och registerdata, men också neurokognitiva test av impulskontrollförmåga samt olika sätt att undersöka hjärnans struktur och funktion, exempelvis elektroencefalografi och strukturell hjärnavbildning.

Förekomsten av maladaptiva beteenden präglade av disinhibition var likartad med, och i vissa avseenden även högre än vad som framkommit i tidigare nationella och internationella studier. Olika maladaptiva beteenden präglade av disinhibition var kopplade till nedsatt neurokognitiv funktion, främst i form av impulsiv planering och problemlösning samt nedsatt förmåga till responsinhibering. Kopplingar framkom även till neurobiologiska förändringar i hjärnregioner involverade i övervakning och utvärdering av beteende, impulskontroll, arbetsminne, och uppmärksamhet. Slutligen visade sig en kombination av beteendemässiga och neurobiologiska mått kopplade till disinhibition avsevärt förbättra träffsäkerheten gällande sannolikheten att återfalla i brott.

Avhandlingen bidrar med nya data till ett fält med knapphändig forskningslitteratur, och understryker vikten av disinhibition som ett kliniskt användbart begrepp. Maladaptiva beteenden präglade av disinhibition bör uppmärksammas i större utsträckning än vad som görs i dagsläget, både i forensiska sammanhang och bland unga vuxna i den allmänna befolkningen. Avhandlingens resultat kan användas som underlag för fortsatt utveckling av nya diagnostiska ramverk och för framtida studier om individanpassad vård och behandling, baserat på kunskap om förändringar i hjärnans struktur och funktion.

List of Studies

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I Delfin, C., Andiné, P., Hofvander, B., Billstedt, E., & Wallinius, M. (2018). Examining associations between psychopathic traits and executive functions in incarcerated violent offenders. *Frontiers in Psychiatry, 9*, 310.

II Delfin, C., Krona, H., Andiné, P., Ryding, E., Wallinius, M., & Hofvander, B. (2019). Prediction of recidivism in a long-term follow-up of forensic psychiatric patients: Incremental effects of neuroimaging data. *PLoS ONE, 14*(5), 1–21.

III Delfin, C., Ruzich, E., Wallinius, M., Björnsdotter, M., & Andiné, P. Trait disinhibition and NoGo event-related potentials in violent mentally disordered offenders and healthy controls. Submitted.

IV Delfin, C., Andiné, P., Wallinius, M., & Björnsdotter, M. Structural brain correlates of the externalizing spectrum in young adults. Submitted.

Acknowledgements

Peter Andiné, my main supervisor, has been a bedrock of support and encouragement throughout this project. With a calm demeanor one can only acquire from decades of experience, your kind guidance and profound knowledge of forensic psychiatry made sure that this project stayed on course until the end. Thank you.

Märta Wallinius, my first co-supervisor, has allowed me to wander off into the unknown, and brought me back when necessary. Your extensive clinical expertise and keen eye for the big picture has been a constant rescue. Thank you.

Malin Björnsdotter, my second co-supervisor, introduced me to the intricate world of neuroimaging research. Ever positive and down to earth, you have been a treasured source of knowledge, inspiration, and ideas. Thank you.

I have been fortunate to have the faith and support of my employer, the Regional Forensic Psychiatric Clinic in Växjö, throughout this project. A special thanks to *Martin Lindgren*, *Tina Fogelklou*, *David Wirdelöv*, *Clas Bengtsson*, and *Ann-Sofie Karlsson*. You made this journey possible, and your enthusiasm for research and devotion to improving forensic psychiatric care has been both invaluable and admirable.

A heartfelt thanks to everyone at the Clinic's Research Department, which has steadily grown since this project began. I am particularly grateful to *Natalie Laporte*, who has endured my endless monologues about methods and statistics, *Andreas Söderberg*, always ready to discuss the ins and outs of psychiatry, politics, and life in general, *Johan Berlin*, a much needed ally in the march towards open science, the "newcomers" *Fernando González Moraga*, *Jessica Revelj*, and *Stéphanie Klein Tuente*, who brought with them laughter and inspiration, *Eva Lindström*, a fountain of wisdom, and *Christel Karlsson*, who diligently brings order into chaos.

A special shout-out to my colleagues and fellow risk assessors — *Jacob Jansson* and *Magnus Rosén* in particular — who has provided much needed banter and dark humor.

I am grateful to everyone at the Centre for Ethics, Law and Mental Health (CELAM). I haven't visited as much as I had planned or hoped, but everytime I did manage to show up, you treated me with a kindness and warmth one can only hope to receive. I am especially grateful to *Susanna Radovic*, *Henrik Anckarsäter*, *Thomas Nilsson*, and *Alessio Degl'innocenti* for your kind support throughout this project. Thank you, also, to *Åse Holl* and *Stefan Axelsson*, for help with all practical matters, big and small.

I am greatly indebted to my fellow co-authors — *Emily Ruzich*, *Hedvig Krona*, *Björn Hofvander*, *Eva Billstedt*, and *Erik Ryding* — who went to great lengths in order to answer questions and help me figure out even the tiniest of details. It has been a pleasure working with you.

I have received a lot of help and support throughout this project, especially during data collection, and I sincerely thank everyone who has contributed. I am especially grateful to *Eirini Alexiou*, *Oskar Holmberg*, and the staff at the Rågården Forensic Psychiatric Clinic in Gothenburg. *Miriam Hermansdotter* and *Linnea Huld* were instrumental in making sure that everything ran smoothly during those long weekends with the scanner. I also wish to extend my sincere gratitude to *Marie Thorstensson Levander* and her colleagues at the Department of Criminology, Malmö University, who offered generous support in a time of need and asked for nothing in return.

Since long before this project began, I have had the privilege to come across a number of individuals that, while not directly involved in this project, have had a profound impact on the way I think about research. I hope that you know who you are, and that you recognize bits and pieces of your influence in this thesis.

Thank you to my family, and to my friends, for remaining curious and supportive without quite understanding what it is that I do, and when — if ever — I was going to wrap things up.

I am ever so grateful to everyone who devoted their time to participate. Without you, of course, there would be no thesis. Thank you!

Finally, thank you *Karolin*, for your love, kindness, and patience. You inspire me daily, and your warmhearted support and encouragement is what kept me going.

Contents

Abbreviations vii

1 Introduction 1
- 1.1 Defining disinhibitory psychopathology . . . 2
- 1.2 Disinhibitory versus externalizing psychopathology . . . 3
- 1.3 Measuring disinhibitory psychopathology . . . 4
  - 1.3.1 Diagnostic manuals . . . 4
  - 1.3.2 Alternative nosological frameworks . . . 5
  - 1.3.3 Self-report assessment . . . 6
- 1.4 Disinhibitory psychopathology and psychopathy . . . 7
  - 1.4.1 Early conceptualizations of psychopathy . . . 7
  - 1.4.2 The Psychopathy Checklist . . . 8
  - 1.4.3 Other conceptualizations of psychopathy . . . 9
- 1.5 Disinhibitory psychopathology and crime . . . 10
  - 1.5.1 Persistent offenders . . . 10
  - 1.5.2 Risk assessment and recidivism prediction . . . 11
- 1.6 Disinhibitory psychopathology and neurocognition . . . 13
  - 1.6.1 Executive functions . . . 13
  - 1.6.2 Response inhibition . . . 14
- 1.7 Disinhibitory psychopathology and neurobiology . . . 15
  - 1.7.1 Structural overview of the human brain . . . 15
  - 1.7.2 Brain regions implicated in disinhibition . . . 16

2 Aims 21
- 2.1 General aim . . . 21
- 2.2 Specific aims . . . 21

3 Methods 23
- 3.1 Ethics . . . 24
- 3.2 Study I . . . 24
  - 3.2.1 Participants and procedures . . . 24
  - 3.2.2 Psychopathic traits . . . 25
  - 3.2.3 Executive functions . . . 25
  - 3.2.4 Data analysis . . . 26
- 3.3 Study II . . . 26
  - 3.3.1 Participants and procedures . . . 26
  - 3.3.2 Baseline measures . . . 27

3.3.3 Follow-up data . . . 27
3.3.4 Neuroimaging data . . . 28
3.3.5 Data analysis . . . 28
3.4 Study III . . . 30
3.4.1 Participants and procedures . . . 30
3.4.2 Clinical data and self-report assessments . . . 31
3.4.3 Response inhibition . . . 32
3.4.4 Event-related potentials . . . 32
3.4.5 Data analysis . . . 33
3.5 Study IV . . . 34
3.5.1 Participants and procedures . . . 34
3.5.2 Clinical data and self-report assessments . . . 34
3.5.3 Response inhibition . . . 34
3.5.4 Magnetic resonance imaging . . . 35
3.5.5 Data analysis . . . 36

4 Results 37
4.1 Prevalence of disinhibitory psychopathology . . . 37
4.1.1 Study I . . . 37
4.1.2 Study II . . . 37
4.1.3 Study III . . . 38
4.1.4 Study IV . . . 38
4.2 Disinhibitory psychopathology and neurocognition . . . 39
4.2.1 Study I . . . 39
4.2.2 Study III . . . 40
4.2.3 Study IV . . . 40
4.3 Disinhibitory psychopathology and neurobiology . . . 41
4.3.1 Study II . . . 41
4.3.2 Study III . . . 41
4.3.3 Study IV . . . 42
4.4 Neurobehavioral variables in recidivism prediction . . . 43

5 Discussion 45
5.1 Comments on main findings . . . 45
5.1.1 Prevalence of disinhibitory psychopathology . . . 45
5.1.2 Disinhibitory psychopathology and neurocognition 49
5.1.3 Disinhibitory psychopathology and neurobiology . 53
5.1.4 Neurobehavioral variables in recidivism prediction 57
5.2 General discussion . . . 59
5.2.1 From basic research to clinical application . . . 59
5.2.2 Implications for future research . . . 61

5.2.3 Strengths and limitations 63
5.2.4 Ethical considerations 65
5.3 Summary and conclusions 66

A The Swedish criminal justice system 67

B Frequentist vs. Bayesian inference 71

C Introduction to Bayesian statistical modeling 77

D Bayesian models used in Study III and IV 83

References 87

Abbreviations

ACC	Anterior Cingulate Cortex
AMPD	Alternative DSM-5 Model for Personality Disorders
BF	Bayes Factor
CANTAB	Cambridge Neuropsychological Test Automated Battery
CAPP	Comprehensive Assessment of Psychopathic Personality
CI	Confidence Interval
CrI	Credible Interval
DLPFC	Dorsolateral Prefrontal Cortex
DSM	Diagnostic and Statistical Manual for Mental Disorders
EEG	Electroencephalography
EF	Executive Function
ERP	Event-Related Potential
ESI	Externalizing Spectrum Inventory
ESI-BF	Externalizing Spectrum Inventory-Brief Form
FPI	Forensic Psychiatric Investigation
FWE	Family-Wise Error
HDI	Highest Density Interval
HiTOP	Hierarchical Taxonomy of Psychopathology
ICD	International Classification of Diseases
IED	Intra/Extra Dimensional Shift
JAGS	Just Another Gibbs Sampler
MCMC	Markov Chain Monte Carlo
MRI	Magnetic Resonance Imaging
NHST	Null Hypothesis Significance Testing
OFC	Orbitofrontal Cortex
PCL-R	Psychopathy Checklist-Revised
PCL:SV	Psychopathy Checklist: Screening Version
PPI-R	Psychopathic Personality Inventory-Revised
QI	Quantile Interval
rCBF	Regional Cerebral Blood Flow
RDoC	Research Domain Criteria
SOC	Stockings of Cambridge
SST	Stop Signal Task
SWM	Spatial Working Memory
tDCS	Transcranial Direct Current Stimulation
TriPM	Triarchic Psychopathy Measure

1. Introduction

Forty years ago, Gorenstein & Newman (1980) coined the term 'disinhibitory psychopathology' to describe a set of separate yet related syndromes that demonstrate deficits in inhibition, failure of self-control, and excessive rule-breaking or norm-violation. They used the term 'disinhibition' in a deliberately vague and descriptive sense; being disinhibited, they suggested, means being unable to control immediate urges, thus disregarding long-term goals in favour of instant gratification. Furthermore, Gorenstein & Newman (1980) proposed that these syndromes, which included hyperactivity, impulsivity, alcoholism, as well as antisocial behavior and psychopathy, share the same genetic origin and reflect similar central nervous system abnormalities. Since then, research has indeed found evidence of a common genetic influence underlying several mental disorders characterized by disinhibition, including ADHD, conduct disorder, substance use disorders, and antisocial personality disorder, as well as personality traits such as novelty seeking and neurocognitive deficits such as impaired response inhibition (Hicks et al., 2013; Kendler et al., 2016; Young et al., 2000, 2009).

Disinhibitory psychopathology is robustly associated with crime and recidivism, and thus of crucial importance to the criminal justice system (de Carvalho et al., 2013; McReynolds et al., 2010; Wibbelink et al., 2017). For instance, males who in adulthood are diagnosed with antisocial personality disorder may follow a similar developmental trajectory that begins with oppositional defiant disorder and conduct disorder during childhood — sometimes further complicated by co-occuring ADHD — followed by substance abuse in adolescence, and that later results in incarceration and recidivism (Beauchaine et al., 2017). As such, research that may lead to novel, disinhibition-focused treatment and intervention strategies is increasingly encouraged (Mullins-Sweatt et al., 2019).

Recent years has also seen increased interest in incorporating neurobiological findings into forensic mental health practice, such as in psychological assessment, risk assessment, and recidivism prediction (Patrick et al., 2019; van Dongen & Franken, 2019), alongside initiatives highlighting the benefits of dimensional approaches to psychopathology research (Kotov et al., 2017). Clarifying the biological aspects of psychopathological constructs, and how they relate to observed behavior, promotes a more comprehensive understanding of their etiology,

and can help guide research towards promising targets for treatment (Perkins, Latzman, et al., 2020). Still, knowledge about the neurobiological and behavioral — or *neurobehavioral* — correlates of disinhibitory psychopathology is still lacking, especially in forensic mental health populations.

1.1 Defining disinhibitory psychopathology

The term 'disinhibition' likely originates from Russian physiologist Ivan Pavlov's (1849-1936) work on classical conditioning, in which he described disinhibition as the "inhibition of an inhibition" (Pavlov, 1927, p. 67). In this context, disinhibition refers to the recurrence of a conditioned response during its extinction phase. Later, in the early 1960s, Russian neuropsychologist Alexander Luria (1902-1977) described experiments carried out by one of his students, Evgenia Homskaya (1929-2004). In one experiment, participants were instructed to press a rubber bulb in response to red signals, and to withhold pressing the bulb in response to green signals. When the signals were made shorter in length and presented at an accelerated rate, the participants began to make mistakes; they pressed the rubber bulb on green signals, although often accompanying the incorrect response by exclaiming "Wrong!". The experimental setup bears striking resemblance to the modern Go/NoGo task, and Luria called this inability to refrain from pressing the bulb the "disinhibition of inhibitory reactions" (Luria, 1961, pp. 112–113).

In contemporary research, and in the context of disinhibitory psychopathology, the term 'disinhibition' is used in a broader sense; it may be defined as a general propensity towards impulse control problems, characterized by deficits in planning and foresight, impaired ability to regulate affect and urges, and an insistence on immediate gratification (Krueger & South, 2009; Patrick et al., 2009). Furthermore, although there are many definitions of 'psychopathology' (for an overview, see Adams et al., 2002), the term usually denotes a pattern of abnormal or maladaptive behavioral expressions that deviate from cultural norms and expectations. Thus, in broad agreement with its original use by Gorenstein & Newman (1980), this thesis will use the term 'disinhibitory psychopathology' to refer to maladaptive behavioral expressions of disinhibition, such as a lack of responsibility, impatience and impulsivity that often leads to negative consequences, anger and reactive aggression, and proneness to substance abuse and engagement in norm-violating and antisocial activities (Krueger & South, 2009; Patrick et al., 2009).

1.2 Disinhibitory versus externalizing psychopathology

The terms 'disinhibitory psychopathology' and 'externalizing psychopathology' are sometimes, and seemingly arbitrarily, used interchangeably (e.g., Iacono et al., 1999). Likewise, the term 'disinhibition' is sometimes used synonymously with terms such as 'externalizing proneness' (e.g., Venables, Foell, Yancey, Kane, et al., 2018). While similar, it can be argued that there is an important but perhaps often overlooked difference between these two terms.

The idea of an 'externalizing spectrum' of behaviors dates back to Achenbach (1966), who observed that behavioral symptoms of various mental disorders form two coherent clusters. He labelled these clusters Internalizing, describing problems within the self, and Externalizing, describing conflict with the surrounding environment. Among the behavioral symptoms included in the Externalizing cluster were disobedience, stealing, lying, and fighting, as well as cruelty and inadequate guilt feelings (Achenbach, 1966). Today, externalizing spectrum *disorders* typically refer to mental disorders that express distress outwards, such as ADHD, conduct disorder, substance use disorders, and antisocial personality disorder, while externalizing spectrum *behaviors* often refer to reckless, impulsive, violent, and antisocial tendencies (e.g., Beauchaine et al., 2017).

However, while the externalizing spectrum refers to disorders and behaviors largely characterized by disinhibitory psychopathology, and while disinhibition is believed to be the core feature of *all* externalizing spectrum disorders and behaviors (Krueger & South, 2009), there are aspects of the externalizing spectrum that are *not* directly related to disinhibition, such as callousness and lack of empathy (Venables & Patrick, 2012). In the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017, see Section 1.3.2), for instance, callousness is a trait associated with 'Antagonistic Externalizing', whereas 'Disinhibited Externalizing' is characterized by impulse control problems. Thus 'externalizing' and 'externalizing psychopathology' could be considered to be broader constructs than 'disinhibition' and 'disinhibitory psychopathology'.

1.3 Measuring disinhibitory psychopathology

1.3.1 Diagnostic manuals

The *Diagnostic and Statistical Manual for Mental Disorders* (e.g., DSM-5; American Psychiatric Association, 2013) groups antisocial, borderline, narcissistic, and histrionic personality disorder into the so-called Cluster B category of personality disorders. These personality disorders are all characterized by disinhibitory tendencies, including impulsivity, recklessness, and difficulties in regulating behavior and emotions (Casillas & Clark, 2002; Taylor et al., 2006). Of the four, antisocial personality disorder, which is defined by a chronic pattern of unlawful, reckless, impulsive, and irresponsible behavior that begins in adolescence and persists into adulthood, is probably the clinical diagnosis that is closest to the concept of disinhibitory psychopathology (McKinley et al., 2018).

Nonetheless, no current evidence indicates that personality disorders are categorical in nature, nor that there are a specific number of different personality disorders (Hopwood, 2018). After decades of persistent and vocal encouragement to abandon the categorical approach to personality disorders (e.g., Widiger & Simonsen, 2005; Widiger & Trull, 2007), the DSM-5 includes a so-called 'Alternative DSM-5 Model for Personality Disorders' (AMPD). Within the AMPD (p. 780), the trait domain 'Disinhibition' is defined as:

> Orientation toward immediate gratification, leading to impulsive behavior driven by current thoughts, feelings, and external stimuli, without regard for past learning or consideration of future consequences.

Similarly, the latest revision of the *International Classification of Diseases* (ICD-11; World Health Organization, 2020, Section 6D11.3) includes Disinhibition in its list of five prominent personality trait domains:

> The core feature of the Disinhibition trait domain is the tendency to act rashly based on immediate external or internal stimuli (i.e., sensations, emotions, thoughts), without consideration of potential negative consequences. Common manifestations of Disinhibition, not all of which may be present in a given individual at a given time, include: impulsivity; distractibility; irresponsibility; recklessness; and lack of planning.

These clinical characterizations of disinhibition largely correspond to the description of disinhibition by Gorenstein & Newman (1980), and together they signal an increased interest both in disinhibitory psychopathology as well as in moving away from traditional, categorical diagnoses (Krueger & Tackett, 2015; Mullins-Sweatt et al., 2019). Beyond personality disorders, several DSM-5 diagnostic criteria for ADHD and intermittent explosive disorder align with the construct of disinhibitory psychopathology. For instance, ADHD is characterized by disruptive and inappropriate behaviors, including restlessness and impatience, whereas intermittent explosive disorder entails recurrent failures to control aggressive impulses.

1.3.2 Alternative nosological frameworks

As a response to criticism against the DSM approach to classification, the US National Institute of Mental Health launched the Research Domain Criteria (RDoC) initiative in 2009 (Kozak & Cuthbert, 2016). The RDoC is founded on the view that the use of categorical classification systems has thwarted attempts by neuroscientists and geneticists to develop a robust and useful theory of psychopathology, and that while classification may be a clinical necessity, it should not lure us into thinking that mental disorders themselves are discrete entities (Clark et al., 2017; Kozak & Cuthbert, 2016). Within RDoC, the construct most closely resembling disinhibition is called 'cognitive control', defined as the ability to inhibit unwanted behavior (Clark et al., 2017). Despite its criticism against the DSM, however, the RDoC is envisioned as a framework for research rather than clinical practice.

The HiTOP, mentioned in Section 1.2, was designed to be a viable alternative to the DSM that is readily incorporated into clinical practice. Specifically, the HiTOP framework is based on the idea of a superspectrum of general psychopathology that is parsed hierarchically into five levels, from spectra at the top to traits at the bottom. Disinhibition is represented in a spectrum labelled 'Disinhibited externalizing', which entails acting on impulse or in response to a current stimulus, with little consideration of consequences (Krueger et al., 2018). Notably, both the Disinhibited externalizing spectrum and a spectrum labelled 'Antagonistic externalizing' (characterized by callousness and deceitfulness) lead to the 'Antisocial behavior' subfactor.

1.3.3 Self-report assessment

Despite decades of research, there are few, unifying models that link together traits and behaviors characterized by disinhibition. A promising development, therefore, is the Externalizing Spectrum Inventory (ESI; Krueger et al., 2007); a comprehensive self-report instrument that allows for dimensional assessment of disinhibitory tendencies, antisocial behaviors, and substance abuse. The ESI, with its 415 items, has often been deemed too excessive, however, resulting in the use of different shortened versions. To ameliorate this, Patrick et al. (2013) developed a brief form (ESI-BF) that allows for more efficient assessment, with shorter subfactors ($\sim$ 20 items each) able to index different manifestations of externalizing behavior. As illustrated in Figure 1.1, the ESI-BF contains three (moderately correlated) subfactors, each made up of several lower-order facet scales: the General Disinhibition subfactor, reflecting the core propensity towards impulse control problems characterized by insistence on immediate gratification, deficient behavioral restraint, and lack of planfulness and foresight; the Callous-Aggression subfactor, reflecting destructive, antagonistic, and aggressive tendencies; and the Substance Abuse subfactor, reflecting both recreational as well as problematic substance use (Patrick et al., 2009, 2013).

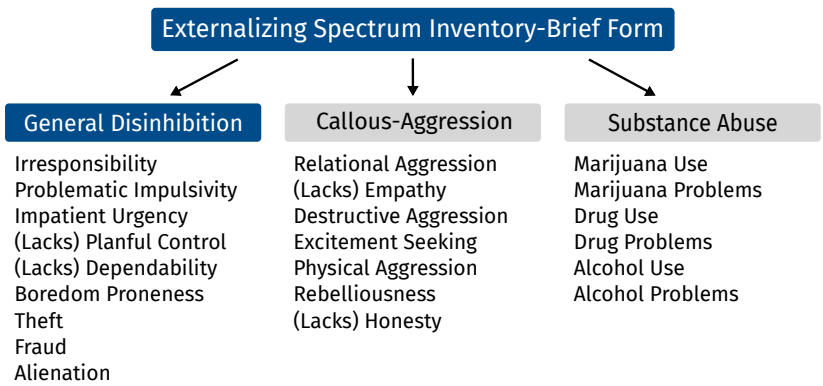

Figure 1.1: The three subfactors of the Externalizing Spectrum Inventory-Brief Form (Patrick et al., 2013). The left side highlights the subfactor that is primarily associated with disinhibitory psychopathology.

The ESI-BF — and the General Disinhibition subfactor in particular — is consistent both with the early description of disinhibition by Gorenstein & Newman (1980) and with the 'Disinhibition' domains of the AMPD and ICD-11. Moreover, the ESI-BF represents a promising, dimensional alternative to categorical approaches to psychopathology (Krueger & Tackett,

2015) that is recommended for use within the HiTOP framework (Kotov et al., 2018). Unfortunately, few studies to date have utilized the ESI-BF, with research in offender samples especially lacking.

1.4 Disinhibitory psychopathology and psychopathy

Psychopathic traits, described as a "prescription for the commission of antisocial and criminal acts" (Hare & Neumann, 2009, p. 796), have been linked to both general and violent crime — perhaps most notably severe forms of violence — as well as to recidivism and institutional problems (Fox & DeLisi, 2019; Kiehl & Hoffman, 2011; Leistico et al., 2008). While Gorenstein & Newman (1980, p. 302) noted that psychopathy probably is the "prototypical syndrome of disinhibition", contemporary researchers likely would argue that psychopathy entails more than just disinhibition. Still, most if perhaps not all conceptualizations of psychopathy highlight disinhibitory tendencies to some degree, and the disinhibitory features of psychopathy have been associated with a propensity for both reactive and proactive forms of violence (Blais et al., 2014; van Dongen et al., 2017).

1.4.1 Early conceptualizations of psychopathy

Modern conceptualizations of psychopathy stem, to a large extent, from American psychiatrist Hervey Cleckley's (1903-1984) idea of a severe pathology masked by an outward appearance of an ordinary, well-functioning individual. Cleckeley's 1941 book *The Mask of Sanity* was derived from his own experiences working in an inpatient psychiatric hospital, and contains a list of 16 specific criteria that characterized psychopathic individuals, including "superficial charm and good intelligence", "inadequately motivated antisocial behavior", "failure to follow any life plan", and "pathologic egocentricity and incapacity for love" (Cleckley, 1988, p. 338).

Although present, Cleckley did not emphasize impulsivity or aggression in his descriptions. Other early scholars, however, outlined such tendencies as key components of the psychopathy construct. For instance, some argued that two primary clinical features must be present in order to constitute psychopathy: a lack of affection towards other humans, and a liability to act on impulse and without forethought (Craft, 1966, p. 5). Similarly, McCord & McCord (1964, pp. 8, 10, 87) described psychopathic individuals as "highly impulsive" and capable of "brutal aggression", "sub-

ject to aggressive explosions", with a "dangerously disruptive" behavior. Together, these examples demonstrate how disinhibitory tendencies has been an important part of the psychopathy construct ever since its early conceptualizations (for an overview, see Yildirim & Derksen, 2015).

1.4.2 The Psychopathy Checklist

In 1980, Robert Hare published the Psychopathy Checklist, and the subsequent revised version (PCL-R; Hare, 2003) and its derivatives remain the most common instruments for the assessment of psychopathy in adults (Hare & Neumann, 2009; Yildirim & Derksen, 2015). The PCL-R, a clinical rating scale, has been subject to a wealth of research investigating its internal structure, with a proposed two-factor, three-factor, four-factor, and two-factor, four-facet solution. In the latter, illustrated in Figure 1.2, the psychopathy construct is divided into two separate factors, the first often referred to as 'interpersonal-affective' and the second as 'impulsive-antisocial'. These two factors are further divided into four facets, which represent deviant interpersonal relations, affective deficiencies, an impulsive and irresponsible lifestyle, and antisocial behavior.

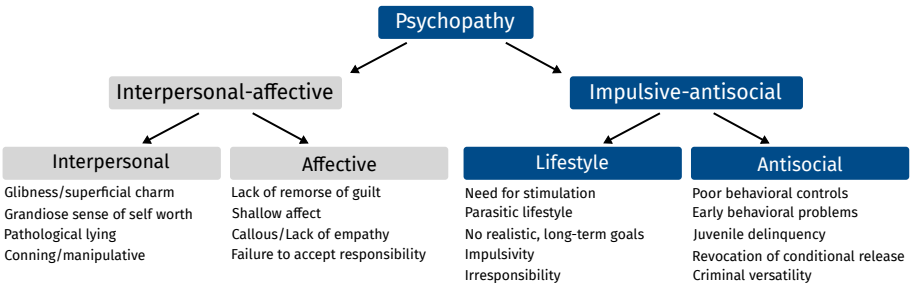

Figure 1.2: The four-facet structure of the Psychopathy Checklist-Revised (Hare, 2003). The right side highlights the two facets that are primarily associated with disinhibitory psychopathology.

Disinhibitory psychopathology as assessed by variants of the ESI has been robustly associated with the impulsive-antisocial factor of the PCL-R, and especially the facet representing an impulsive and irresponsible lifestyle, suggesting that the two instruments tap the same underlying construct (Patrick et al., 2009; Venables et al., 2014; Venables & Patrick, 2012).

1.4.3 Other conceptualizations of psychopathy

Despite decades of research, there is still no agreed upon definition of psychopathy at a conceptual level (e.g., Cooke et al., 2012). Furthermore, the immense popularity of the Psychopathy Checklist and its subsequent revisions and derivatives has made some researchers worried that the instrument has become equated with the theoretical psychopathy construct itself (Skeem & Cooke, 2010). It is not surprising, therefore, that several other conceptualizations of psychopathy have emerged, all of which highlight disinhibitory tendencies to varying degrees.

A relatively recent addition is the Triarchic Model of Psychopathy (TriPM; Patrick et al., 2009), which was developed as a complementary model to other approaches to psychopathy while also aiming to be more closely tied to psychological and neurobiological measures. The TriPM views psychopathy as the intersection of three separate but related components: Disinhibition, Boldness, and Meanness. Notably, the Disinhibition component of the TriPM is equivalent to General Disinhibition subfactor of the ESI-BF. For convenience, the TriPM Disinhibition component will henceforth be referred to as the General Disinhibition subfactor of the ESI-BF. The Boldness component reflects a capacity to remain calm and focused in pressuring or threatening situations, along with self-confidence and tolerance for risk and uncertainty, and the Meanness component reflects deficient empathy, exploitativeness, and empowerment through cruelty or destructiveness (Patrick et al., 2009, 2012).

Other examples include the Psychopathic Personality Inventory and its revised version (PPI-R; Lilienfeld & Widows, 2005), and the Comprehensive Assessment of Psychopathic Personality (CAPP; Cooke et al., 2012). The PPI-R consists of three subfactors: Fearless-Dominance, Self-centered Impulsivity, and Coldheartedness. The Self-centered Impulsivity subfactor is probably closest to the construct of disinhibitory psychopathology, and has been robustly associated with the ESI-BF General Disinhibition subfactor, both among mentally disordered offenders and in community volunteers (van Dongen et al., 2017). Similarly, the CAPP is organized into six thematic domains of personality: Attachment, Behavioral, Cognitive, Dominance, Emotional, and Self (Cooke et al., 2012). Recent research using a sample of self-reported offenders found that all CAPP domains were positively correlated with both the ESI-BF General Disinhibition subfactor, as well as the TriPM Meanness subfactor (Hanniball et al., 2019).

1.5 Disinhibitory psychopathology and crime

Disinhibitory psychopathology, especially when it leads to interpersonal violence and other forms of criminality, exacts a tremendous toll on society. The effects on victims of crime, in addition to possible personal injury and damage to personal property, include reduced quality of life as well as increased mental health problems (Hanson et al., 2010; Tan & Haining, 2016). Beyond individual suffering, estimates from Sweden suggests long-term societal costs of approximately €600 000 following a single assualt of "moderate" degree (Nilsson & Wadeskog, 2012, p. 97). With this in mind, research aimed at improving our understanding as to why some individuals are so deeply involved in criminal activity — and what we can do to prevent it — is an essential endeavor, and one in which disinhibitory psychopathology may play an important role.

1.5.1 Persistent offenders

Research has shown that a small minority of the population accounts for a large proportion of crime (e.g., Reidy et al., 2015), and that this small minority, to a large extent, is characterized by disinhibitory psychopathology. For instance, a study using a nationally representative sample of non-institutionalized participants from the US (N = 43 093, aged 18 years and older) identified a 'severe group' of persistent offenders, corresponding to ~5% of participants. The severe group had substantially higher rates of disinhibitory tendencies, including substance use and reckless, antagonistic, and violent behavior, both compared to the 'normative group' as well as to a 'high substance use/moderate antisocial behaviors' group and a 'low substance use/high antisocial behaviors' group (Vaughn et al., 2011). In a follow-up study, the authors used a large (N = 18 614) nationally representative sample of adolescents from the US and again identified a 'severe group', corresponding to ~5% of adolescents, with substantially higher rates of, primarily, severe expressions of disinhibitory psychopathology (Vaughn et al., 2014).

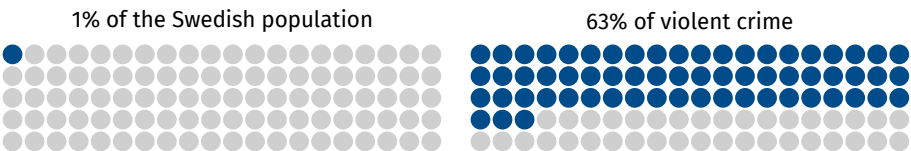

Figure 1.3: Illustrating the small minority responsible for a majority of violent crime convictions in Sweden (Falk et al., 2014).

Similar observations have been made in Sweden, as illustrated in Figure 1.3. A nationwide, multi-generation register study found that 1% of the total Swedish population born between 1958 and 1980 accounted for over 63% of all violent crime convictions between 1973 and 2004. Those involved in persistent violence displayed a high prevalence of disinhibitory psychopathology, including personality disorders, substance use disorders, previous drug-related offenses, and early-onset violent criminality (Falk et al., 2014).

All told, society has a lot to gain from preventing crime and recidivism, and the robust association between disinhibitory psychopathology and crime (de Carvalho et al., 2013; McReynolds et al., 2010; Wibbelink et al., 2017) puts into perspective the value of disinhibition-focused treatment. Specifically, developing intervention and prevention efforts targeting the core propensity towards impulse control problems in the small minority of persistently violent offenders may lead to substantial public health improvements, and ease the strain on the criminal justice system.

1.5.2 Risk assessment and recidivism prediction

Given the robust association between disinhibitory psychopathology and crime it is not surprising that research has pinpointed disinhibitory tendencies such as poor self-control, restlessness, and impulsivity as the most relevant clinical risk factors for recidivism (Bonta et al., 2014; Skeem et al., 2014). According to recent estimates, about a third of those released from prison in Sweden commit a new crime within three years (Kriminalvården, 2020), and about a third of those discharged from forensic psychiatric care commit a new crime within five years (RättspsyK, 2020). Since an essential objective of the criminal justice system is to prevent recidivism, whether following release from prison or discharge from forensic psychiatric care, there is a pressing need to reduce these recidivism rates. One way to do so is by appropriate risk management, which in turn requires accurate assessments of the risk of violence and recidivism.

There are well over two hundred different risk assessment tools in current use (Singh et al., 2014), although many are practically interchangeable, and most have at best moderate accuracy (Singh & Fazel, 2010; Yang et al., 2010). Research suggests that more than half of those predicted as high risk are incorrectly classified and will not recidivate, while around one in ten prediced as low risk will in fact commit a new crime (Fazel et al., 2012). As Fazel et al. (2012) point out, these limitations

lead to imporant ethical questions, such as whether some individuals are detained for longer than necessary, and whether some pose an increased — and unidentified — risk to the general public. Along these lines, the Swedish Prison and Probation Services has highlighted increased precision and methods that tailor assessments to individual needs as the primary challenges for risk assessment research (Kriminalvården, 2014b). Similarly, the Swedish Research Council recently concluded that there are large knowledge gaps in the field of forensic psychiatry, with research concerning the precision of risk assessments underscored as particularly lacking (Vetenskapsrådet, 2017).

Two recent developments have sparked a new generation of research aimed at improving the accuracy of risk assessments: increased recognition of the brain's role in antisocial and violent behavior, and pioneering advances in the field of artificial intelligence (Tortora et al., 2020). Specifically, a subset of methods from the domain of artifical intelligence called 'machine learning' are increasingly used in prediction models. The term 'artificial intelligence' generally refers to the science of making intelligent machines and computer programs, wheras 'machine learning' refers to computer programs that learn from data and improve with experience. In turn, a subset of machine learning called 'deep learning' may be regarded as the current forefront of artificial intelligence research (Kersting, 2018).

The prospect of incorporating neurobehavioral variables associated with disinhibitory psychopathology into prediction models — while fraught with both practical and ethical hurdles — has become increasingly feasible (e.g., Poldrack et al., 2018), with some promising preliminary findings. A prospective study by Aharoni et al. (2013) combined a neurocognitive task tapping disinhibitory tendencies with functional magnetic resonance imaging (functional MRI, or fMRI) to predict recidivism during a four year follow-up period after release from prison. They found that the odds of comitting a new crime were twice as high for offenders with relatively low neural activity in a region called the dorsal anterior cingulate cortex compared to offenders with relatively high activity in this region. Their results were later corroborated using additional statistical techniques, offering further support for the use of neuroimaging data in risk assessment models (Aharoni et al., 2014). Steele et al. (2015) then took a subset of the participants used in the study by Aharoni et al. (2013) and supplemented the data with measurements acquired using electroencephalography (EEG). Moreover, this

time they used both traditional statistical techniques (i.e., logistic and Cox proportional hazards regression) and a machine learning method called support vector machine; the latter proved superior in predicting recidivism.

While these studies demonstrate the promising potential of combining several neurobehavioral variables associated with disinhibitory psychopathology in prediction models, research using samples of mentally disordered offenders remains scarce. To paraphrase Nadelhoffer et al. (2012, p. 3), errors in the domain of 'neuroprediction' are both morally unacceptable and economically costly, and further research is required in order to assess the feasibility of utilizing neurobehavioral variables to improve the prediction of recidivism.

1.6 Disinhibitory psychopathology and neurocognition

There are six principal domains of neurocognitive function defined in the DSM-5 — complex attention, executive function, learning and memory, language, perceptual–motor function, and social cognition — and each contains several subdomains. Although often referred to in the context of neurocognitive disorders such as delirium and dementia (e.g., Sachdev et al., 2014), neurocognitive deficits, especially within the executive function (EF) subdomain, are also robustly associated with disinhibitory psychopathology.

1.6.1 Executive functions

Although EF impairments are believed to be involved in many forms of psychopathology (Snyder et al., 2015), there is also a robust association with disinhibitory psychopathology, with considerable evidence linking EF impairments to aggressive, antisocial, and violent behavior in both children (Granvald & Marciszko, 2016; Rohlf et al., 2018), adolescents (Barker et al., 2007; Zou et al., 2013), and adults (Meijers et al., 2015; Moffitt, 2018; Morgan & Lilienfeld, 2000; Ogilvie et al., 2011).

Barkley (2012) describes how the term 'executive function' arose from a loose description by Karl Pribram (1919-2015) of the functions performed by the prefrontal cortex, which he had labelled the 'executive of the brain' (Pribram, 1973, p. 293)[1]. In contemporary research, however, EFs may

[1] The interested reader may note that the referenced work was edited by Pribram and Alexander Luria, with contributions by, among several others, Evgenia Homskaya.

be defined as top-down, deliberate cognitive functions that support goal-directed behavior, although a more exact definition remains a matter of debate (e.g., Nigg, 2017). For instance, the DSM-5 lists six EF subdomains: decision-making, flexibility, inhibition, planning, responding to feedback, and working memory. Others instead promote a three-factor model consisting of three core EFs — flexibility (sometimes referred to as shifting), working memory (sometimes referred to as updating), and inhibition — from which more complex EFs, such as planning and problem-solving, are formed (Friedman & Miyake, 2017; Miyake et al., 2000).

In the three factor model, flexibility allows us to change perspectives and see things from others' point of view, to adjust our behavior to the current environment, and to switch between different tasks and rules. Working memory is the ability to mentally hold, manipulate, or reorder information, and is critical for temporal coherence, understanding language, and for planning and problem-solving. Inhibition enables us to control attention, behavior, thoughts, and emotions, and to override impulses, habits, and external stimuli in order to do what is appropriate or necessary (Diamond, 2013; Diamond & Ling, 2016). Together, these core EFs work together to influence and control lower level processes in order to regulate and maintain goal-directed behavior, and combine to form more complex EFs, such as planning and problem-solving and decision-making.

1.6.2 Response inhibition

As the name suggests, inhibition may be the most revelant EF for disinhibitory psychopathology. While Miyake et al. (2000) explicitly defined inhibition as the ability to suppress prepotent responses, Tiego et al. (2018) suggests that there are two types of inhibition, both associated with working memory capacity: response inhibition, reflecting the ability to suppress a prepotent motor response, and attentional inhibition, reflecting the ability to resist interference from distracting stimuli. Others argue that response inhibition may be further parsed into two different components: stopping and inhibiting. Specifically, some tasks, such as the Stop-Signal Task (SST), measure the ability to stop an already initiated response. In contrast, other tasks, such as the Go/NoGo, measure the ability to inhibit a prepotent response altogether (Littman & Takács, 2017). Finally, successfully inhibiting a prepotent response is associated with different neural processes, including both pre-motor

processes such as conflict monitoring as well as post-response outcome evaluation (Gajewski & Falkenstein, 2013; Huster et al., 2013).

Response inhibition has been proposed as a valuable endophenotype in studies on disinhibitory psychopathology (Young et al., 2009), and research has linked deficient response inhibition to impulsive and provoked aggression (Hecht & Latzman, 2018), impulsive-antisocial psychopathic traits (Feilhauer et al., 2012), and violent offending (Meijers et al., 2017). Still, the extent to which different forms of response inhibition, such as stopping and inhibiting, are associated with disinhibitory psychopathology is less clear. Furthermore, few studies have investigated how response inhibition relates to disinhibitory tendencies and measures of neurophysiological activity in mentally disordered offenders.

1.7 Disinhibitory psychopathology and neurobiology

Proponents of alternative frameworks such as the RDoC believe that we may have reached the limits of understanding mental disorders based solely on observable behaviors and self-reported thoughts and feelings (Clark et al., 2017). Indeed, although Raine (2018) argues that the failure to recognize antisocial personality disorder as a neurodevelopmental disorder may have hindered clinical progress in terms of treatment and intervention, neuroscientific discoveries have not mapped well onto traditional, categorical mental disorders (Kozak & Cuthbert, 2016). This shortcoming might explain why none of the mental disorders of the DSM-5 — not even those already termed 'neurodevelopmental disorders' — include neurobiological indicators for use in diagnosis (e.g., Patrick et al., 2019). Thus, a dimensional approach focused on lower-order constructs such as disinhibition combined with neurobiological measures may be the most promising way forward (Hyman, 2007; Nelson et al., 2016).

1.7.1 Structural overview of the human brain

The human brain, comprising just 2% of our body weight, consumes 20% of our oxygen supply, most of which is used for synaptic signalling (Harris et al., 2012). Around 80% of the brain's mass is taken up by the 2-4 mm thick outer layer called the cerebral cortex. The neocortex, which constitutes the major part of the cerebral cortex, contains around 20 billion neurons with an average of 7 000 synaptic connections each. These neurons,

the brain's so-called gray matter, are interconnected by upwards of 180 million meters of myelinated nerve fibers, or axons, which is the brain's so-called white matter (Azevedo et al., 2009; Drachman, 2005). To pack as many neurons as possible into the limited space available, the cerebral cortex is folded, creating a pattern of sulci, or groves, and gyri, or crests.

The brain's main part, the cerebrum, is divided into two hemispheres, and each hemisphere is further divided into four lobes — the frontal, parietal, occipital, and temporal — named after the skull bones they cover. The brain may be further divided into smaller regions of interest; the Desikan-Killiany atlas (Desikan et al., 2006), for instance, parcellates the brain into 34 cortical regions of interest in each hemisphere based on the brain's gyral structure.

1.7.2 Brain regions implicated in disinhibition

A wealth of research has found evidence of structural and functional alterations associated with various forms of criminal and antisocial behavior (Glenn & Raine, 2014; Ling et al., 2019; Raine, 2019). Given the prominent role of the brain's frontal regions in the complex interplay between cognition and behavior (Miller, 2000), it is not surprising that much research has been dedicated to probing frontal lobe structure and function in aggressive, antisocial, and violent individuals (Brower & Price, 2001; Séguin, 2009). Structural and functional brain alterations are usually mapped to different cognitive processes and observable behaviors, such as EFs, thus offering a more complete understanding of why individuals act a certain way. As an example, Raine & Yang (2006) proposed a 'neuromoral theory' of antisocial behavior, based on the overlap between brain regions implicated in antisocial, violent, and psychopathic behavior and those implicated in moral decision-making.

Yang & Raine (2009) conducted the first meta-analysis of structural and functional abnormalities in antisocial individuals. Based on 43 studies, three frontal regions, shown in Figure 1.4, emerged as particularly important: the anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), and dorsolateral prefrontal cortex (DLPFC). Although subsequent research has identified additional regions, including the insula, striatum, and amygdala, the ACC, OFC, and DLPFC remain key regions in the latest revision of the neuromoral theory (Raine, 2019). Still, it is possible that different forms of disinhibitory psychopathology are associated with different brain abnormalities, which in turn may be associated with different neurocognitive processes. For instance, Rodman et al. (2016)

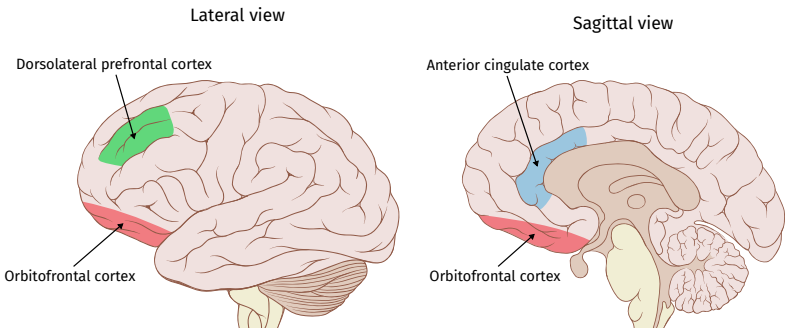

Figure 1.4: Three brain regions implicated in disinhibitory psychopathology. Figure adapted from images by Patrick J. Lynch and C. Carl Jaffe, licensed under CC BY 2.5 (https://creativecommons.org/licenses/by/2.5/).

found that variance unique to disinhibitory tendencies was associated with diminished DLPFC activity during response inhibition, whereas variance unique to psychopathic traits instead was related to increased fronto-parietal activity during both response inhibition and interference suppression. Thus, although there are multiple regions involved in aggressive, antisocial, and violent behavior, not all may be associated with the disinhibitory aspects of those behaviors.

1.7.2.1 The anterior cingulate cortex

The ACC is a complex region, not in the least due to its inconsistent nomenclature. The dorsal-anterior part of the cingulate cortex (i.e, the dorsal ACC) is referred to as the caudal ACC in the Desikan-Killiany atlas, and corresponds to a region defined as the anterior midcingulate cortex by Vogt (2016). Furthermore, the rostral part of the ACC corresponds to a region that Vogt (2016) refers to as the pregenual ACC. Notwithstanding these differences in terminology, parcellation of the cingulate cortex into four separate regions, illustrated in Figure 1.5, is corroborated by research showing different receptor binding patterns in each region (Palomero-Gallagher et al., 2009). In light of this, researchers should take care to specify which part of the cingulate cortex they refer to when studying the ACC.

The ACC — and the dorsal ACC in particular — is an important region in research on disinhibitory psychopathology. The precise functions of the dorsal ACC remain an area of intense research (Ebitz & Hayden, 2016), but it does play a fundamental role in monitoring and control. In short, the dorsal ACC is involved in the monitoring and evaluation of

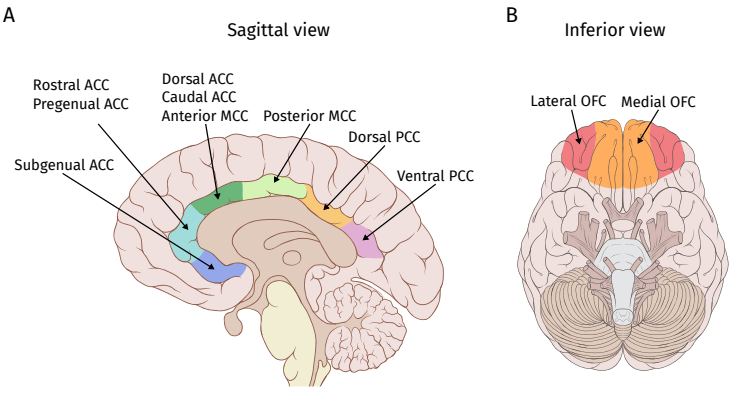

Figure 1.5: (A) The proposed four-region parcellation of the cingulate cortex by Vogt (2016). ACC, anterior cingulate cortex; MCC, midcingulate cortex; PCC, posterior cingulate cortex. Note that the retrosplenial cortex is not depicted. (B) The lateral and medial subdivisions of the orbitofrontal cortex (OFC). Figure adapted from images by Patrick J. Lynch and C. Carl Jaffe, licensed under CC BY 2.5 (https://creativecommons.org/licenses/by/2.5/).

errors, rewards, and conflicts, and it uses this information to control and regulate behavior (Heilbronner & Hayden, 2016). For instance, monitoring and evaluative functions of the dorsal ACC are believed to be activated during response inhibition tasks, both during successful and failed inhibitions (Whelan et al., 2012), and abnormal dorsal ACC structure has been associated with reduced inhibitory control capacity (Holmes et al., 2016).

The dorsal ACC has been proposed as a primary neural generator (along with the OFC) of the NoGo P3 event-related potential (Fallgatter et al., 2002; Hong et al., 2017; Schmajuk et al., 2006). Event-related potentials (ERPs) are recorded using EEG and reflect the concurrent firing of thousands of primarily pyramidal neurons in deep cortical layers (Kirschstein & Köhling, 2009). As such, ERPs provide a direct measurement of the brain's post-synaptic neurotransmission, with high temporal resolution. The NoGo P3 ERP, a positive voltage deflection occurring around 300 ms after a successful inhibition, is believed to represent processes related to monitoring or response evaluation (Huster et al., 2013), thus linking it to the dorsal ACC. Although research remains scarce, especially in violent and antisocial samples, reduced NoGo P3 amplitude and delayed NoGo P3 latency — a measure of the peak of the NoGo P3 waveform — are promising neurobehavioral correlates of disinhibitory psychopathology (e.g., Guan et al., 2015; Kaiser et al., 2020; Verona & Bresin, 2015).

1.7.2.2 The orbitofrontal cortex

The OFC has long been recognized for its role in emotion, decision-making, and reward processing (Rolls, 2019), as well as in the regulation of fear and anxiety (Hiser & Koenigs, 2018; Milad & Rauch, 2007). The reward processing capabilities of the OFC are vital to our day-to-day functioning since, according to Bechara et al. (2000), damage to the OFC results in pathological impairments in the decision-making process, in turn leading to adverse effects on the decisions that has to be made in daily life. Wallis (2007) presents a more detailed account of the OFC's role in decision-making: the OFC receives sensory, affective, and motivational information from multiple regions, including the temporal cortex, amygdala, and hypothalamus. It then integrates this information in order to derive the value of potential reward outcomes. The derived value is sent to the DLPFC, which constructs a plan for obtaining the reward outcome, and to the medial prefrontal cortex, which evalutes how much effort is involved in the plan. Finally, the DLPFC and the medial prefrontal cortex orchestrate a behavioral response. Importantly, the calculations carried out by the OFC, DLPFC, and medial prefrontal cortex in this context are all assumed to take place in working memory (see Wallis, 2007, Figure 4, p. 48, for an overview). Furthermore, as with the ACC, different subdivisions of the OFC may have distinct functions (Figure 1.5). The medial part of the OFC is believed to be involved in processing reward-related information in anticipation of a reward, while the lateral part of the OFC appears to primarily process information related to punishment and unobtained rewards (Rolls, 2019).

Individuals with OFC damage have been called disinhibited, socially inappropriate, and impulsive (Séguin, 2004), and it is no wonder — given the ubiquitous role of the OFC in processes essential to everyday life — that both structural (De Brito et al., 2009; Gao et al., 2020; Jiang et al., 2016) and functional (Murray et al., 2018) abnormalities in this region have been associated with antisocial behavior. However, the role of the OFC in different expressions of disinhibitory psychopathology is less clear, as is the role of its subdivisions.

1.7.2.3 The dorsolateral prefrontal cortex

The DLPFC is a multifaceted region involved in several executive functions. It has a central role in working memory and shifting, and has been identified as a key region involved in planning and problem-solving (Nitschke et al., 2017). Damage to the DLPFC has been associated with im-

pairments in arithmetic and reasoning, planning and problem-solving, and the ability to manipulate verbal and spatial knowledge in working memory (Barbey et al., 2013; Manes et al., 2002). Furthermore, since inhibition requires remembering when and what to inhibit, the DLPFC is also activated in response inhibition tasks with high working memory load (Courtney, 2004; Simmonds et al., 2008).

Structural abnormalities in the DLPFC have been associated with aggressive behavior (Fairchild et al., 2013), and a study by Sadeh & Verona (2008) found that impulsive-antisocial psychopathic traits were associated with impaired working memory. The authors speculate that this may lead to difficulties in maintaining cognitive control in complex situations (i.e., when working memory load is high), which then could result in increased risk-taking, frustration, and anger. The DLPFC's role in working memory thus suggests an important link between brain function, impaired EFs, and adverse behavioral outcomes. Similarly, Raine & Yang (2006) suggest that a dysfunctional DLPFC may predispose to response perseveration (i.e., a failure of shifting), increasing the risk of life-course persistent antisocial behavior, as well as poor planning and problem-solving. This could then result in a dysfunctional lifestyle due to problems faced at work and in social relationships (Raine & Yang, 2006). In light of these studies, and since the multifaceted DLPFC appears related to all of the EFs described in Section 1.6.1, the DLPFC is an important region to consider in research on disinhibitory psychopathology.

2. Aims

2.1 General aim

The overarching aim of the thesis is to further our understanding of how different neurobehavioral variables relate to the construct of disinhibitory psychopathology.

2.2 Specific aims

1. Quantify the prevalence of different expressions of disinhibitory psychopathology in offenders and the general population (*Studies I-IV*)
2. Examine associations between neurocognitive function and different expressions of disinhibitory psychopathology (*Studies I, III* and *IV*)
3. Examine associations between neurobiological structure and function and different expressions of disinhibitory psychopathology (*Studies II, III* and *IV*)
4. Explore how neurobehavioral variables associated with disinhibitory psychopathology may be used in the prediction of recidivism (*Study II*)

3. Methods

An overview of the methods used in each study is presented in Table 3.1. Since three out of the four studies use data from offenders, a brief introduction to relevant areas of the Swedish criminal justice system is offered in Appendix A. For readers not familiar with the Bayesian approach to statistical inference, a comparison of frequentist and Bayesian approaches is available in Appendix B, with a brief introduction to Bayesian statistical modeling presented in Appendix C. A more detailed overview of the Bayesian models used in *Studies III* and *IV* is provided in Appendix D.

Table 3.1: Methodological overview.

	Study I	Study II	Study III	Study IV
Thesis aims	Aims 1 and 2	Aims 1, 3, and 4	Aims 1-3	Aims 1-3
Study population	Violent offenders	Mentally disordered offenders	Violent mentally disordered offenders and healthy controls	Young adults
Sample characteristics	N = 213 (males only), mean age = 22 years, range = 19-25	N = 44 (4 females), mean age = 38 years, range = 20-79	N = 47 (males only; 27 patients), mean age = 35 years, range = 20-58	N = 59 (39 females, 19 males, 1 non-binary), mean age = 23 years, range = 18-32
Measures of disinhibitory psychopathology	Psychopathic traits	Cluster B personality disorder, substance use disorders, psychopathic traits, criminal history, recidivism	Self-reported disinhibitory traits and behaviors, DSM-5 diagnoses, criminal history	Self-reported externalizing traits and behaviors
Neurocognitive assessments	Flexibility, spatial working memory, response inhibition, planning and problem-solving	N/A	Response inhibition	Response inhibition
Neuroimaging technique	N/A	Single-photon emission computed tomography	Electroencephalography	Magnetic resonance imaging
Analytic approach	Frequentist and Bayesian statistics	Frequentist statistics and machine learning	Bayesian statistics	Frequentist and Bayesian statistics

With the exception of EEG data preprocessing in *Study III* and whole-brain analyses in *Study IV*, all data analysis was carried out using the R statistical programming language (R Core Team, 2020). In order to facilitate transparency and increase reproducibility (e.g., Allen & Mehler, 2019), the code used in each study (e.g., code used for data analysis, data preprocessing, and the Go/NoGo task) is freely and publicly available online for anyone to examine, use, and modify:

I. https://github.com/carldelfin/EF-psychopathy

II. https://github.com/carldelfin/neuroprediction

III. https://osf.io/yscdh

IV. https://osf.io/m4v9d

3.1 Ethics

With the exception of *Study II*, where informed consent was not deemed necessary, all participation was voluntary and based on informed, written consent. All studies were approved by a regional ethics review board:

Study I was approved by the regional ethics review board in Lund (2009/405), *Study II* was approved by the regional ethics review board in Lund (2007/64, 2014/911), *Study III* was approved by the regional ethics review board in Linköping (2017/56-31, 2018/7-32, 2018/321-32), and *Study IV* was approved by the regional ethics review board in Göteborg (538-18).

3.2 Study I

3.2.1 Participants and procedures

Study I is a cross-sectional study with participants recruited from the Development of Aggressive Antisocial Behavior Study; a nationally representative cohort of young violent offenders recruited between 2010 and 2012. Detailed descriptions are available in Wallinius et al. (2016) and Billstedt et al. (2017).[1]

Participants were assessed based on file reviews, structured clinical interviews, self-report assessments, observations, and neuropsychological testing. A clinical psychologist, with special training in the methods used, administered all interviews, observations, and neuropsychological testing during a full day of participation. Out of a total cohort of 269 offenders, 54 did not participate in or complete all the neuropsychological assessments used in *Study I*, and assessments of psychopathic traits were unavailable for two offenders. Thus, *Study I* consisted of 213 male, violent offenders, aged 19 to 25 years.

[1] One participant managed to participate twice, and *Study I* was published before this was discovered. The second participation was removed, which resulted in minor numerical differences, none of which changed the overall results or conclusions. A corrigendum is included following *Study I*.

3.2.2 Psychopathic traits

The PCL-R, consisting of 20 items scored on a 3-point scale (0 = does not apply, 1 = applies somewhat, 2 = does apply), was used to assess psychopathic traits. Ratings were performed using all information available from interviews, observations, and files. To ensure inter-rater reliability, training sessions with consensus ratings on participants were performed. Internal consistency was good, with Cronbach's alpha = 0.65, 0.81, 0.78, and 0.77 for the interpersonal, affective, lifestyle, and antisocial facets, respectively. The mean corrected item-total correlation was 0.56, 0.70, 0.64, and 0.61 for the interpersonal, affective, lifestyle, and antisocial facets, respectively. Note that item N ranged from 211 to 213 for the facets.

3.2.3 Executive functions

Four tests from the Cambridge Neuropsychological Test Automated Battery (CANTAB; Cambridge Cognition Ltd., Cambridge, UK) were used to assess flexibility, spatial working memory, response inhibition, and planning and problem-solving.

The Intra/Extra Dimensional Shift task (IED), a version of the Wisconsin Card Sorting Test (Heaton et al., 1993), was used to assess flexibility. A total of nine stages with increasing difficulty were completed by reaching a certain criterion at each stage. Outcome measures were the number of stages completed and the number of errors made.

The Spatial Working Memory task (SWM) was used to assess spatial working memory. Outcome measures were strategy score, a measure of optimal search strategy (Owen et al., 1990) with possible scores ranging from 0 to 40 (higher scores indicate a less efficient search strategy) and the number of errors made.

The SST was used to assess the ability to inhibit a prepotent response (Verbruggen & Logan, 2008). Outcome measures were stop-signal reaction time (the average time at which the participant is able to successfully inhibit the prepotent motor response) and the mean correct response time.

The Stockings of Cambridge task (SOC), a version of the Tower of London task (Shallice, 1982), was used to assess planning and problem-solving skills. Outcome measures were mean initial thinking time (the average time taken before attempting to solve a five-move problem) and the number of problems solved in the minimum number of moves.

3.2.4 Data analysis

3.2.4.1 Frequentist statistical tests

Zero-order Pearson's *r* correlations were used to assess the association between psychopathic traits and executive functions, with threshold for statistical significance set to $p < 0.05$.

3.2.4.2 Bayesian statistical models

Bayesian models were estimated using the R package `BayesMed` (Nuijten et al., 2015), which implements the Just Another Gibbs Sampler (JAGS; Plummer, 2003) along with a Jeffreys-Zellner-Siow prior in order to sample from the posterior distribution using Markov Chain Monte Carlo (MCMC; van Ravenzwaaij et al., 2018). The Jeffreys-Zellner-Siow prior has been suggested as a suitable prior that conveys little information (i.e., is "weakly informative") while also having the desired characteristics (Wetzels & Wagenmakers, 2012). A total of 20 000 iterations with 2000 burn-in samples were used, and results from Bayesian analyses are presented as the posterior probability of the observed correlation, given the data and model, and the corresponding Bayes factor (BF). A BF greater than one indicates evidence in favor of estimated correlation, and a value smaller than one indicates evidence against the estimated correlation. In line with Wetzels & Wagenmakers (2012), we interpreted BFs < 1/3 as indicating substantial evidence in favor of $\mathcal{H}_0$, 1/3 < BF < 1 as anecdotal evidence in favor of $\mathcal{H}_0$, 1 < BF < 3 as anecdotal evidence in favor of $\mathcal{H}_1$, and BF > 3 as indicating substantial evidence in favor of $\mathcal{H}_1$.

3.3 Study II

3.3.1 Participants and procedures

Study II is a retrospective follow-up study with participants recruited from the Forensic Psychiatric Follow-up Studies–the Malmö Cohort. The cohort is a nationally representative, total cohort consisting of participants within the Malmö University Hospital catchment area who, after being arrested for a crime, underwent a major forensic psychiatric investigation (FPI; N = 97) or a section-seven investigation (N = 28), and who were subsequently sentenced to involuntary forensic psychiatric inpatient care. In total, the cohort consists of 101 men and 24 women recruited between 1999 and 2005, aged between 17 and 79 years old at the

time of inclusion. Detailed descriptions are available in Andreasson et al. (2014) and Krona et al. (2017).

When recruitment began in 1999, FPI investigees were routinely referred for a neuroimaging assessment, which consisted of resting-state regional cerebral blood flow (rCBF) measurements acquired using single-photon emission computed tomography (SPECT). Due to changes in local FPI procedures during the recruitment period, however, only 50 participants underwent this assessment.[2] Two participants were omitted from *Study II* due to missing data on educational attainment, three were omitted due to missing data on age at first crime, and one was omitted due to incomplete data on psychopathic traits. Thus, *Study II* consisted of 44 participants, aged 20 to 79 years at the time of inclusion.

3.3.2 Baseline measures

Demographic (sex, age, educational attainment), forensic (previous criminality, age at first crime, mental disorders in first-degree relatives), and clinical (dates of admittance and discharge, mental disorders, pharmacological treatment) data was obtained from FPI protocols and patient records. Mental disorders were structured according to the DSM, 4th Edition (DSM-IV; American Psychiatric Association, 1994). The Psychopathy Checklist: Screening Version (PCL:SV; Hart et al., 1995), consisting of 12 items scored on a 3-point scale (0 = does not apply, 1 = applies to a certain extent, 2 = does apply), was used to assess psychopathic traits. Most PCL:SV assessments were carried out during the FPI based on clinical evaluation and extensive file and register reviews, although five assessments were performed retrospectively, based on file reviews, by the second author of *Study II*.

3.3.3 Follow-up data

Follow-up data was obtained from the National Council of Crime Prevention. In cases where the participant had deceased during follow-up, dates of death were retrieved from the Cause of Death Register at the National Board of Health and Welfare. Time at risk was defined as beginning at each patient's intake date and lasting until reconviction, death, deportation, or until the follow-up ended on the December 31, 2013. Recidivism was defined as any criminal conviction during the time at risk.

[2] Note that the neuroimaging assessment was clinically motivated and not originally intended for research purposes, as further discussed in Section 5.2.4.

3.3.4 Neuroimaging data

Resting-state rCBF was assessed using SPECT measurements with 900 MBq of ^{99mTc}-exametazime delivered through a pre-set cannula in a cubital vein (Ceretec^{TM}, Nycomed-Amersham/GE Healthcare), and recorded using a Ceraspect SPECT camera (Digital Scintigraphics Inc., Waltham, Massachusetts). The imaging procedure began about 15 minutes after ^{99mTc}-exametazime had been administered and continued for about 30 minutes, providing a snapshot of resting-state rCBF. The radioactive ^{99mTc}-exametazime decay was recorded in 180° to allow for three-dimensional reconstruction of activity proportional to rCBF, with a resolution of 9 mm full-width at half-maximum, corrected for scatter and attenuation. The recorded, three-dimensional activity was saved into a 128 x 128 x 64 voxel matrix, subdivided into 10 slices with 1 cm thickness parallel to the orbitomeatal line, and then parsed into a region of interest set. Activity in four regions of interest in each hemisphere (the frontal lobe, parietal lobe, temporal lobe, and cerebellum), quantified as percent of mean ^{99mTc}-exametazime concentration across the whole brain using Amersham software (GE Healthcare, Buckinghamshire, UK), was selected for analysis.

3.3.5 Data analysis

3.3.5.1 Frequentist statistical tests

Barnard's test (Barnard, 1945), a more powerful alternative to Fisher's exact test when sample sizes are small (Lydersen et al., 2009), was used to examine differences between recidivists and non-recidivists for all dichotomous variables. Welch's t-test, which performs better than Student's t-test when sample size and variance differs between groups, was used to examine group differences for all continuous variables. Finally, Spearman's ρ was used to examine the monotonic relationship between rCBF and age. Threshold for statistical significance was set to $p < 0.05$.

3.3.5.2 Machine learning

The random forest machine learning algorithm (Breiman, 2001), which uses an ensemble approach to aggregate predictions made by a large collection of decision trees, was used to predict recidivism. The ensemble approach makes random forests accurate and relatively robust to outliers and noise, and they work well with both small sample sizes and high-dimensional data (Biau & Scornet, 2016; Breiman, 2001).

A further advantage is that the random forest algorithm provides internal estimates of both model error and predictor importance, by creating each decision tree using a random bootstrap sample (with replacement) corresponding to approximately two thirds of the data. The remaining so-called out-of-bag data is used to estimate model error and predictor importance. Specifically, only a random subset of predictors is selected at each node in the decision tree, and each predictor is split to optimize tree performance. The predictor split that produces the highest tree performance is then selected for that node. Then, each observation in the out-of-bag data is passed down the decision tree and classified, and the final classification of each observation is based on a majority vote from all trees where that observation was in the out-of-bag data. The importance of each predictor is estimated during the out-of-bag phase using random permutation. For predictors important for classification, permutation results in a large decrease in classification accuracy, while unimportant variables are more or less unaffected.

Three random forest models were created: (1) a baseline model, containing traditional risk factors for recidivism, (2) an extended model, containing both traditional risk factors as well as regional rCBF measurements, and (3) a supplementary pharmacological model, containing traditional risk factors, regional rCBF measurements, and pharmacological data. All models were created using 10 000 trees and $\sqrt{p}$ predictors at each node, with p being the total number of predictors available. To ameliorate issues with class imbalance, the majority class (i.e., non-recidivists) was down-sampled so that each bootstrap sample contained the same number of non-recidivists as recidivists (Kuhn & Johnson, 2013). The predictive performance of each model was assessed using several metrics:

- Area under the (receiver operating characteristic) curve: the probability that a randomly selected recidivist will have been predicted to have a higher probability of recidivism than a randomly selected non-recidivist.
- Accuracy: the proportion of correct classifications.
- Sensitivity: the proportion of correctly classified recidivists.
- Specificity: the proportion of correctly classified non-recidivists.
- Positive predictive value: the proportion of predicted recidivists that did recidivate.
- Negative predictive value: the proportion of predicted non-recidivists that did not recidivate.

In addition, Scaled Brier scores were used to determine which model predicted an outcome with a probability closest to the true outcome. Regular, unscaled Brier scores are defined as the squared difference between the actual, binary outcome Y (0 or 1) and the continuous, predicted probability P (ranging from 0 to 1). Scaling the Brier score so that it no longer depends on the prevalence of Y results in a scaled Brier score that ranges between 0 and 1 (Steyerberg et al., 2010; Wu & Lee, 2014). Finally, the scaled mean decrease in accuracy for each predictor (the estimated decrease in model accuracy, should that predictor be omitted) was used as a measure of predictor importance, and partial dependence plots were used to visualize the direction and size of effect, after averaging out the effect of all other predictors (Friedman, 2001).

3.4 Study III

3.4.1 Participants and procedures

Study III is a cross-sectional study consisting of mentally disordered offenders sentenced to involuntary forensic psychiatric inpatient care at a maximum security forensic psychiatric hospital in Sweden, as well as healthy controls, recruited between 2017 and 2019.

Male mentally disordered offenders that had, at any point, been sentenced for a violent crime were recruited following completion of a parallel, ongoing study at the same facility. Out of the 65 offenders that had completed the parallel study during the recruitment window for *Study III*, 10 had left the facility before being approached, 9 were female, 4 had a history of brain damage with lasting effects, 5 were deemed unsuitable due to current psychiatric status or safety concerns, and 2 were not possible to contact during the recruitment window. The remaining 35 offenders were asked to participate after receiving oral and written information, out of which a total of 29 agreed to participate. All participation was carried out in a secure area within the facility with a research assistant always present, and was completed while participants were on their usual medication and treatment plan. A voucher for use in the facility's kiosk or at a local mall, worth ~$10, was given as compensation after completed participation.

Male volunteers, recruited from staff at two hospitals and students at a university using posters as well as oral and written advertisement, served as a control group. Exclusion criteria were having completed higher ed-

ucation, a history of brain damage with lasting effects, a current major mental disorder, and illegal substance use within the last 6 months. After receiving oral and written information, 25 eligible control group participants were recruited, recieving either a voucher (~$10) for use at a local mall or a movie ticket (~$10) as compensation after completed participation.

Participation consisted of several self-report questionnaires, a Go/NoGo response inhibition task with concurrent EEG acquisition, and an additional resting-state EEG task, and took approximately 60 minutes per participant. Upon completion, all participants were given the opportunity to ask questions and view a live feed of their EEG recording. Five control group participants were excluded after participation, but prior to data analysis, due to illegal substance use within the last 6 months and/or a history of brain damage with lasting effects being reported in the questionnaires[3]. One mentally disordered offender was removed from data analysis due to unsufficient EEG data, and one mentally disordered offender was excluded due to missing all self-report data. Furthermore, one mentally disordered offender lost approximately one third of the EEG data due to technical issues, although the remaining data was retained. Thus, *Study III* consists of 27 mentally disordered offenders and 20 healthy controls, aged 20 to 58 years.

3.4.2 Clinical data and self-report assessments

Clinical data (mental disorders, pharmacological treatment) was obtained from medical records, with mental disorders converted from ICD-10 to DSM-5. The ESI-BF was co-translated into Swedish by the author of this thesis and by the third author of *Study III*. The General Disinhibition subfactor (ESI-BF_{DIS}), which consists of 20 out of the 160 items in the full ESI-BF, was used as a self-report assessment of disinhibitory psychopathology. Items are scored on a 4-point scale, from 0 (Not true at all) to 3 (Completely true), with possible scores on the ESI-BF_{DIS} ranging from 0 to 60. The ESI-BF_{DIS} showed good internal consistency, with Cronbach's alpha = .91 and McDonald's omega total = .93.

[3]Note that this was despite being informed, verbally and in writing, about exclusion criteria prior to participation.

3.4.3 Response inhibition

A Go/NoGo task adapted from Kiehl et al. (2000) was used to assess the ability to inhibit a prepotent response (Figure 3.1). The task was implemented using Presentation (Neurobehavioral Systems Inc.) and presented on a 22 inch widescreen LCD TFT monitor (resolution = 1680 x 1050 @ 60 Hz) during EEG acquisition. A total of 326 trials were used, of which 274 (84%) were Go trials and 52 (16%) were NoGo trials, divided into two blocks of 163 trials each, with rest in between. Outcome measures used were percent of successful inhibitions (NoGo accuracy) and the median response time on correct trials (median NoGo response time).

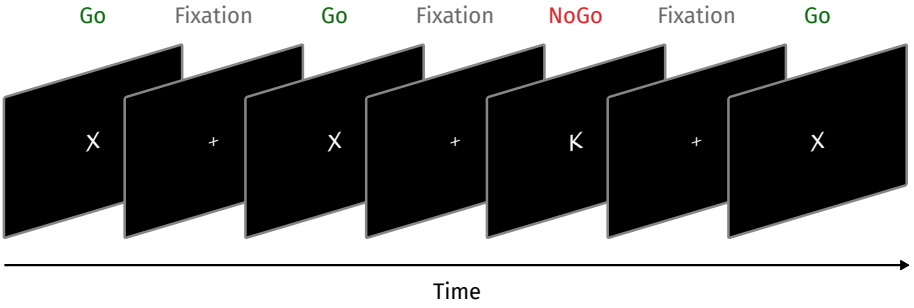

Figure 3.1: Overview of the Go/NoGo task used in Studies III and IV.

3.4.4 Event-related potentials

A high-impedance NetStation NA400 amplifier (Electrical Geodesics Inc.) and 128 Ag/AgCl electrodes positioned in a Hydrocel Geodesic sensor net was used to record EEG during the Go/NoGo task. Preprocessing of EEG data was carried out using version 0.20.4 of the MNE-Python module (Gramfort, 2013), running on Python 3.8.2. The preprocessing pipeline largely adhered to current recommendations in order to ensure quality and reproducibility (Jas et al., 2018), and was — apart from the removal and interpolation of bad channels, independent components, and artifacts — fully automated. All ERPs were quantified using non-subtracted, correct NoGo trials (Gajewski & Falkenstein, 2013) and were averaged across a region of interest consisting of nine frontocentral electrodes in order to increase reliability (Ribes-Guardiola et al., 2020).

Two ERP components — the NoGo N2 and the NoGo P3 — were chosen for further analysis. NoGo N2 amplitude was quantified as the mean amplitude 225 to 325 ms post-stimuli (NoGo $N2_{WIN}$), while NoGo P3 amplitude was quantified as the mean amplitude 325 to 625 ms post-stimuli

(NoGo $P3_{WIN}$). Latency was quantified as the 50% fractional area latency, defined as the time point before which 50% of the negative (for N2) or positive (for P3) area of the waveform is observed (Luck, 2014). Finally, moving window amplitudes, termed NoGo $N2_{MOV}$ and NoGo $P3_{MOV}$, were quantified as the mean amplitude 50 ms before and 50 ms after the 50% positive fractional area latency was observed in order to account for possible latency effects.

3.4.5 Data analysis

A fully Bayesian approach was used, and all models were specified manually using the R package brms (Bürkner, 2017). Importantly, brms interfaces R with the state-of-the-art Stan probabilistic programming language. Stan is more flexible than JAGS (which was used in *Study I*) and implements the more efficient and robust Hamiltonian MCMC sampler (Carpenter et al., 2017). Group differences were modeled using a robust linear regression approach, allowing unequal variances between groups, with a Student's t distribution (Lange et al., 1989). Correlations were, in similar fashion, modeled using a robust linear regression approach with a multivariate Student's t distribution and an LKJ(2) prior (Lewandowski et al., 2009). The reliability of the Go/NoGo paradigm and all ERP measurements was determined by estimating correlations (ρ_{SB}) between the averages of odd and even trials — a so-called split-half approach — corrected using the Spearman-Brown formula (de Vet et al., 2017). All numerical variables were standardized prior to modeling, and all priors were chosen to be weakly informative in order to have negligible impact on posterior estimates, while still providing moderate regularization of potential outliers (Gelman et al., 2017).

Results are presented as the median estimated group difference, the median estimated bias-corrected standardized group difference ($\widehat{\delta}$), and the median estimated correlation coefficient (ρ). Note that $\widehat{\delta}$ corresponds to what is often called Cohen's *d* with Hedges's *g* correction (McGrath & Meyer, 2006). Median estimates are presented along with 90% highest density intervals (HDIs) within square brackets (66% HDIs are included in figures). For instance, a 90% HDI has a 90% probability of containing the parameter of interest (see Appendix C). The probability of direction (P_D), ranging from 50% to 100%, was used to denote the probability that an effect is different from zero.

3.5 Study IV

3.5.1 Participants and procedures

Study IV is a cross-sectional study consisting of community volunteers recruited through ads on social media and university campuses between 2018 and 2019. A total of 59 individuals (39 females, 19 males, 1 non-binary) were eligible (i.e., were between 18 and 32 years old and did not meet exclusion criteria for MRI) and available for scanning at the proposed dates, and thus participated in *Study IV*. When participants arrived at the MRI facility, a research assistant provided oral and written information about the project, including health and safety guidelines for MRI scanners, as well as a consent form. In addition to MRI, participation also included a Go/NoGo task, which was performed either immediately before or after MRI. After participants had completed both MRI and the Go/NoGo task, they were debriefed and given the opportunity to ask further questions. If the participant had not yet completed all online surveys, they were asked to do so at the earliest possible opportunity. Participation, including online questionnaires, took approximately two hours and was compensated with four movie tickets (~$40).

3.5.2 Clinical data and self-report assessments

All clinical and self-report data was obtained using an online platform. Clinical data (mental disorders, pharmacological treatment) was obtained from structured self-report protocols. Mental disorders were assessed using the DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure-Adult (e.g., Mahoney et al., 2020). The ESI-BF_{DIS} subfactor (described above), the 19-item Callous-Aggression subfactor (ESI-BF_{AGG}), with possible scores ranging from 0 to 57, and the 18-item Substance Abuse subfactor (ESI-BF_{SUB}), with possible scores ranging from 0 to 54, were used to assess different expressions of disinhibitory psychopathology. All subfactors showed good internal consistency, with Cronbach's alpha = .87, .92, .89 and McDonald's omega total = .90, .94, .92, for the ESI-BF_{DIS}, ESI-BF_{AGG}, and ESI-BF_{SUB}, respectively.

3.5.3 Response inhibition

A Go/NoGo task identical to the one used in *Study III*, implemented on a 14 inch laptop, was used to assess response inhibition. Outcome measures used were percent of successful inhibitions (NoGo accuracy).

3.5.4 Magnetic resonance imaging

Regional cortical thickness (Figure 3.2) was estimated using MRI. While much previous research on individuals characterized by disinhibitory psychopathology has focused on cortical volume (e.g., Wallace et al., 2014; Yang & Raine, 2009), researchers are increasingly incorporating alternative measures of brain structure, such as cortical thickness (e.g., Ameis et al., 2014; Yang et al., 2015). Since volume is the product of both cortical thickness and cortical surface area — both of which are influenced by distinct genetic mechanisms (Panizzon et al., 2009) — it may be more beneficial to study them separately, rather than focusing solely on volume (Winkler et al., 2010).

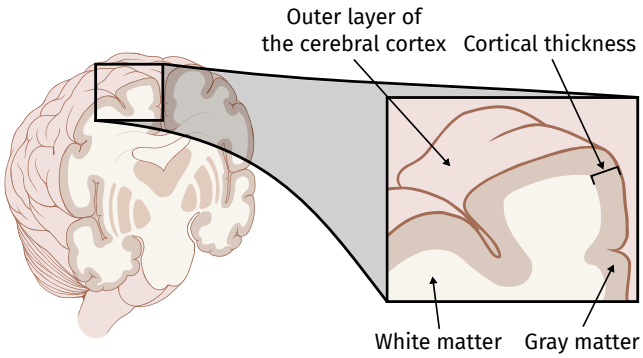

Figure 3.2: Illustration of the thin layer of cerebral cortex that covers the brain. Figure adapted from image by Patrick J. Lynch, licensed under CC BY 2.5 (https://creativecommons.org/licenses/by/2.5/).

A Philips Gyroscan 3T Achieva scanner, software release 3.2, with a 32 channel SENSE head coil (Philips, Eindhoven, the Netherlands), was used to obtain structural brain scans. T1-weighted scans (3D T1-TFE; 170 sagittal slices with scan resolution 1.0 × 1.0 × 1.0 mm^3) were acquired using flip angle = 8°, TE = 4.0 ms, TR = 8.4 ms, a SENSE factor of 2.7, and a TFE factor of 240. Structural brain scans were processed with the Computational Anatomy Toolbox, version r1615 (Gaser & Dahnke, 2016) for Statistical Parametric Mapping software, version 12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12), using MATLAB, version R2020a (Mathworks, Natick, MA, USA). Following automated procedures for surface-based morphometry, including denoising, bias correction, tissue segmentation, and spatial normalization to Montreal Neurological Institute coordinate space (Ashburner & Friston, 2011), all data was rated as having good (N = 58) or satisfactory (N = 1) image quality. The surface data was then resampled and smoothed using

the high-resolution Freesurfer mesh and a 12 mm full-width at half-maximum smoothing kernel. Finally, the projection-based thickness method (Dahnke et al., 2013) was used to estimate cortical thickness in thirty-four regions based on the Desikan-Killiany atlas (Desikan et al., 2006). Five regions of interest in each hemisphere (the caudal and rostral ACC, the medial and lateral OFC, and the rostral middle frontal gyrus, corresponding to the DLPFC) were selected for *a priori* region of interest analyses.

3.5.5 Data analysis

3.5.5.1 Frequentist statistical tests

Multivariate linear regression models were used for whole-brain, surface-based analysis of cortical thickness via SPM12, controlling for the effect of age, gender, and years of education. Peak family-wise error correction (FWE) was used to address multiple comparisons, and threshold for statistical significance was set to $p_{\text{FWE}} < 0.05$.

3.5.5.2 Bayesian statistical models

A robust linear regression approach, similar to *Study III*, was used for *a priori* analyses of the association between cortical thickness and different manifestations of the externalizing spectrum, controlling for age, gender, and years of education. Median estimates of standardized regression coefficients (β) and correlation coefficients (ρ) are presented along with 90% credible intervals (CrIs), presented within square brackets. Like a 90% HDI, a 90% CrI also has a 90% probability of containing the parameter of interest (the difference between a HDI and a CrI is explained in Appendix C). Finally, the probability of direction (P_D) is also reported, and 66% CrIs are included in figures.

4. Results

4.1 Prevalence of disinhibitory psychopathology

The first aim of this thesis, pursued in *Studies I-IV*, was to quantify the prevalence of different expressions of disinhibitory psychopathology in offenders and the general population.

4.1.1 Study I

In *Study I*, the expression of disinhibitory psychopathology was quantified using assessments of psychopathic traits in a large sample of male, young adult violent offenders (N = 213). The mean PCL-R total score was 17.52 (SD = 7.05), with an average score of 0.90 (SD = 1.34) for the interpersonal facet, 3.15 (SD = 2.26) for the affective facet, 6.45 (SD = 2.61) for the impulsive lifestyle facet, and 6.30 (SD = 2.88) for the antisocial behavior facet.

4.1.2 Study II

In *Study II*, the expression of disinhibitory psychopathology was quantified using assessments of Cluster B personality disorder and substance use disorders according to DSM-IV as well as assessments of psychopathic traits and information about criminal history and recidivism, in a sample of mentally disordered offenders (N = 44). Across the whole sample, 7 (16%) had a Cluster B personality disorder, 22 (50%) had a substance use disorder, and the mean PCL:SV score was 10.3 (SD = 5.97). Most (N = 28; 64%) had some form of previous criminality, with an average age at first crime of approximately 30 years (SD = 14).

Mentally disordered offenders who were reconvicted during follow-up were, in a broad sense, more distinctly characterized by disinhibitory psychopathology than those who were not reconvicted. Recidivists had a lower age at first crime and a higher frequency of both Cluster B personality disorder and substance use disorders compared to non-recidivists, although the difference was relatively non-robust for the latter (p = 0.238). Likewise, recidivists presented with a higher degree of psychopathic traits, although again, the difference was less robust (p = 0.142). Details are presented in Table 4.1.

Table 4.1: Baseline model risk factors in non-recidivists (N = 28) and recidivists (N = 16).

	Non-recidivists	Recidivists	Group difference	
	Mean ± SD or N (%)	Mean ± SD or N (%)	t or z	p
Age at FPI	42.29 ± 16.28	30.06 ± 6.95	3.46	0.001
Age at first crime	34.57 ± 15.69	22.94 ± 5.81	3.52	0.001
Cluster B PD	1 (4%)	6 (38%)	-2.96	0.008
Educational attainment	26 (93%)	14 (88%)	0.59	0.732
Male sex	25 (89%)	14 (88%)	0.18	0.967
Mental disorder in FDR	7 (25%)	6 (38%)	-0.87	0.459
PCL:SV total score	9.25 ± 5.6	12.12 ± 6.32	-1.51	0.142
Previous criminality	17 (61%)	11 (69%)	-0.53	0.608
Substance use disorder	12 (43%)	10 (62%)	-1.25	0.238

Note: FPI, (major) forensic psychiatric investigation; PD, personality disorder; FDR, first-degree relative; PCL:SV, Psychopathy Checklist: Screening Version.

4.1.3 Study III

In *Study III*, the expression of disinhibitory psychopathology was quantified using scores on the ESI-BF_{DIS} subscale, assessment of mental disorders according to DSM-5, and information about criminal history in a sample of male, violent mentally disordered offenders (N = 27). Although the most frequent primary diagnosis was within the schizophrenia spectrum, the total number of diagnoses associated with disinhibitory psychopathology (i.e., personality disorders, substance-related and addictive disorders, and ADHD) was double that of the number of schizophrenia spectrum diagnoses. The average age at first criminal sentence was approximately 21 years (SD = 7.30), with an average age of approximately 13 years (SD = 4.40) for first reported crime, and an average of 7.30 (SD = 7.10) previous sentences each. Eight (30%) mentally disordered offenders had committed acts of deadly violence, of which two (7%) on two or more occasions, and 23 (85%) had committed some form of assault, of which 18 (67%) on two or more occasions. Finally, the mean ESI-BF_{DIS} score was 26.62 (SD = 12.43), approximately 19 points higher than the control group (M = 7.50, SD = 6.91), with an estimated effect size of $\hat{\delta} = 1.83$ and a P_D of 100%.

4.1.4 Study IV

In *Study IV*, the expression of disinhibitory psychopathology was quantified using all three ESI-BF subfactors in a sample of community volunteers (N = 59). The average ESI-BF_{DIS} score was 14.20 (SD = 9.78), with average scores of 12.29 (SD = 10.25) for ESI-BF_{AGG} and 23.46 (SD = 11.17) for

ESI-BF_{SUB}. Scores on the ESI-BF_{DIS} were robustly associated with scores on both the ESI-BF_{AGG} (ρ = 0.51 [0.32, 0.66], P_D = 100%) and the ESI-BF_{SUB} (ρ = 0.52 [0.34, 0.66], P_D = 100%). Furthermore, ESI-BF_{AGG} scores were robustly associated with ESI-BF_{SUB} scores (ρ = 0.26 [0.04, 0.45], P_D = 98%).

4.2 Disinhibitory psychopathology and neurocognition

The second aim of this thesis, pursued in *Studies I, III*, and *IV*, was to examine associations between neurocognitive function and different expressions of disinhibitory psychopathology.

4.2.1 Study I

In *Study I*, neurocognitive function was assessed using four subtests from the CANTAB. A descriptive overview of performance on the four EF tasks is presented in Table 4.2. Correlation analyses showed that initial thinking time for a five-move problem in the SOC task was negatively associated with scores on all four facets of the PCL-R, but demonstrated the largest (r = -.22) and most robust association — corroborated by both frequentist (p < 0.01) and Bayesian analyses (posterior probability = 90%, BF = 8.6) — with psychopathic traits reflecting an impulsive lifestyle. In addition, a smaller (r = -.18) and less robust association (p = 0.01), with a posterior probability of 64% and a Bayes factor of 1.8 indicating anecdotal evidence, was found with psychopathic traits reflecting antisocial behaviors.

Table 4.2: Performance on executive function tasks in violent offenders (N = 213).

	Mean ± SD	Range
IED stages competed	8.11 ± 1.13	1 - 9
IED total errors	26.65 ± 12.51	7 - 63
SWM total errors	23.14 ± 17.28	0 - 90
SWM strategy score	32.5 ± 5.11	0 - 47
SST stop-signal response time (s)	0.19 ± 0.08	0.07 - 0.74
SST mean correct response time (s)	0.49 ± 0.14	0.3 - 1.27
SOC mean initial thinking time	6.03 ± 4.79	0 - 29.38
SOC problems solved	8.3 ± 1.75	4 - 12

Note: IED, Intra/Extra Dimensional Shift; SWM, Spatial Working Memory; SST, Stop-Signal Task; SOC, Stockings of Cambridge.

The number of problems solved in the SOC task was negatively associated with affective, impulsive, and antisocial psychopathic traits (rs -0.15 to -0.16, ps < 0.05), but these findings were not corroborated by Bayesian

analyses (posterior probabilities ranged between 39% and 47%, BFs < 1). Similarly, affective and antisocial psychopathic traits were positively associated with SWM strategy score (a high strategy score indicates poor use of strategy), with *r*s = 0.15, *p*s < 0.05, but these findings were, again, not corroborated by Bayesian analyses (posterior probabilities of around 40%, BFs < 1). Remaining associations of interest were small and non-robust.

4.2.2 Study III

In *Study III*, neurocognitive function was assessed using a Go/NoGo response inhibition task. The average NoGo accuracy in the whole sample was 64% (SD = 21%). Mentally disordered offenders (M = 66%, SD = 21%) had an estimated NoGo accuracy of approximately 5% [-7%, 15%] higher than controls (M = 62%, SD = 20%), corresponding to a small effect size ($\hat{\delta}$ = 0.22 [-0.31, 0.71], although the difference was relatively non-robust (P_D = 76%). Follow-up, exploratory analyses, based on data from 20 controls and 26 mentally disordered offenders, revealed that mentally disordered offenders had an estimated 9.15 ms [-11.04, 28.53] longer median NoGo response time than controls, corresponding to a small to moderate effect size ($\hat{\delta}$ = 0.29 [-0.33, 0.96]). Furthermore, median NoGo response time was positively associated with NoGo accuracy (ρ = 0.32 [0.09, 0.54], P_D = 98%) and NoGo P3 latency (ρ = 0.38 [0.14, 0.59], P_D = 99%).

Supplementary analyses not reported in *Study III* found no evidence of a meaningful association between ESI-BF_{DIS} scores and NoGo accuracy in the whole sample (ρ = -0.03 [-0.27, 0.20], P_D = 59%), nor in the control group (ρ = 0.00 [-0.34, 0.34], P_D = 50%). However, a relatively robust, negative association was observed among the mentally disordered offenders (ρ = -0.18 [-0.46, 0.12], P_D = 83%).

4.2.3 Study IV

In *Study IV*, neurocognitive function was assessed using the same Go/NoGo response inhibition task as in *Study III*. The average NoGo accuracy was 61% (SD = 20%). Scores on the ESI-BF_{DIS} subscale showed robust, negative associations with NoGo accuracy (ρ = -0.19 [-0.39, 0.03], P_D = 93%), while associations between NoGo accuracy and ESI-BF_{AGG} (ρ = 0.08 [-0.13, 0.29], P_D = 74%) and ESI-BF_{SUB} (ρ = –0.05 [-0.27, 0.17], P_D= 64%) were small and non-robust.

4.3 Disinhibitory psychopathology and neurobiology

The third aim of this thesis, pursued in *Studies II, III*, and *IV*, was to examine associations between brain structure and function and different expressions of disinhibitory psychopathology.

4.3.1 Study II

In *Study II*, brain function was quantified as resting-state rCBF acquired using SPECT. Mentally disordered offenders who were reconvicted during follow-up presented with lower bilateral parietal lobe, lower bilateral cerebellar, and higher bilateral temporal lobe rCBF than than non-recidivists, although the robustness varied, with *p*s ranging from < 0.001 to 0.168 (Table 4.3).

Table 4.3: Regional cerebral blood flow measurements, as percent of whole-brain perfusion, in non-recidivists (N = 28) and recidivists (N = 16).

	Non-recidivists	Recidivists	Group difference	
	Mean ± SD	Mean ± SD	t or z	p
Resting-state rCBF, left hemisphere				
Cerebellum	120.68 ± 4.92	118.31 ± 4.95	1.53	0.136
Frontal lobe	106.82 ± 4.34	106.31 ± 3.22	0.44	0.660
Parietal lobe	104.18 ± 4.11	102.19 ± 1.94	2.17	0.036
Temporal lobe	101.04 ± 2.85	102.38 ± 1.75	-1.93	0.060
Resting-state rCBF, right hemisphere				
Cerebellum	120.68 ± 4.6	117.81 ± 4.4	2.04	0.049
Frontal lobe	106.46 ± 4.52	106.88 ± 3.22	-0.35	0.728
Parietal lobe	106.11 ± 2.74	102.56 ± 2.71	4.16	0.000
Temporal lobe	102.07 ± 3.67	103.44 ± 2.73	-1.40	0.168

4.3.2 Study III

In *Study III*, brain function was quantified as NoGo N2 and NoGo P3 ERPs acquired using EEG. Overall, mentally disordered offenders had lower NoGo P3 amplitude and delayed NoGo P3 latency compared to healthy controls, whereas findings for the NoGo N2 ERP were small and less robust. The estimated median NoGo $P3_{WIN}$ amplitude was 0.73μV [-2.26, 0.78] lower among mentally disordered offenders than among controls, corresponding to a small effect size ($\widehat{\delta}$ = 0.28 [-0.83, 0.31]) with P_D = 79%. Similarly, the estimated median NoGo $P3_{MOV}$ amplitude was 0.82μV [-2.62, 0.91] lower among mentally disordered offenders than among controls, corresponding to a small effect size ($\widehat{\delta}$ = 0.24 [-0.74,

0.29]) with P_D = 78%. The estimated median NoGo P3 latency was 17.41 ms [2.23, 32.53] longer among mentally disordered offenders compared to controls, corresponding to a moderate effect size ($\hat{\delta}$ = 0.59 [0.06, 1.14]) with P_D = 97%. The grand average waveform is shown in Figure 4.1.

Across the whole sample, a robust, small to moderate association between NoGo P3 latency and NoGo accuracy was observed (ρ = 0.34 [0.12, 0.54], P_D = 99%), as well as a smaller association between NoGo P3 latency and ESI-BF_{DIS} score (ρ = 0.19 [-0.05, 0.42], P_D = 90%). No robust association between ESI-BF_{DIS} score and NoGo $P3_{WIN}$ (ρ = -0.04 [-0.28, 0.22], P_D = 59%) or NoGo $P3_{MOV}$ (ρ = 0.02 [-0.22, 0.25], P_D = 55%) was observed.

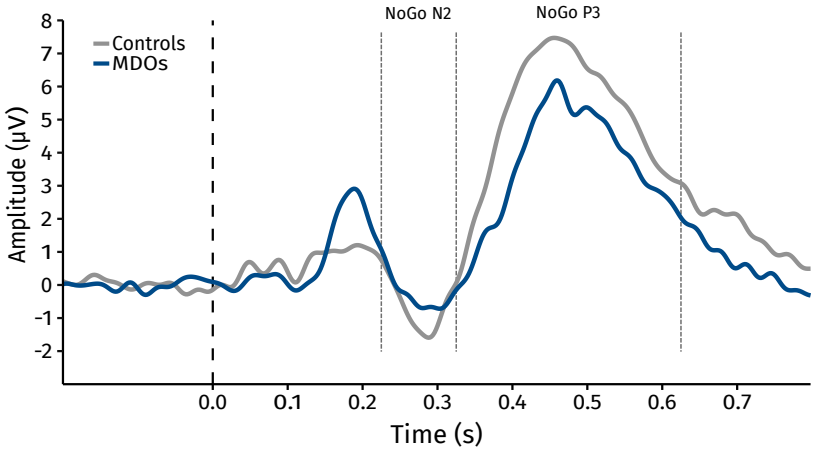

Figure 4.1: The grand average NoGo waveform. Dashed lines demarcate the two windowed ERPs.

4.3.3 Study IV

In *Study IV*, brain structure was quantified as cortical thickness acquired using MRI. A descriptive overview of participants' estimated regional cortical thickness is presented in Table 4.4.

Scores on the ESI-BF_{DIS} were associated with increased cortical thickness in the right lateral OFC (β = 0.20 [-0.07, 0.47], P_D = 89%) and, although with a smaller effect, with decreased cortical thickness in the DLPFC (β = -0.16 [-0.43, 0.11], P_D = 84%). ESI-BF_{AGG} scores were associated with increased cortical thickness in the left (β = 0.13 [-0.11, 0.37], P_D = 83%) and right (β = 0.25 [0.00, 0.52], P_D = 95%) medial OFC, as well as in the right caudal ACC (β = 0.17 [-0.05, 0.38], P_D = 90%). Finally, ESI-BF_{SUB} scores

were associated with increased cortical tickness in the right DLPFC (β = 0.18 [-0.09, 0.46], P_D = 87%).

NoGo accuracy showed robust associations with decreased cortical thickness in the left (β = -0.34 [-0.61, -0.06], P_D = 98%) and right (β = -0.29 [-0.56, -0.01], P_D = 96%) lateral OFC, with increased cortical thickness in the right medial OFC (β = 0.34 [0.06, 0.63], P_D = 98%), and, although less robust, with increased cortical thickness in the right rostral ACC (β = 0.15 [-0.10, 0.39], P_D = 84%)

Table 4.4: Estimated regional cortical thickness in young adults (N = 59).

	Mean ± SD	Range
Cortical thickness (mm), left hemisphere		
Caudal ACC	2.47 ± 0.24	2.04 - 3.04
Rostral ACC	2.52 ± 0.19	2.11 - 2.94
Lateral OFC	2.72 ± 0.12	2.27 - 3.08
Medial OFC	2.34 ± 0.1	2.11 - 2.52
DLPFC	2.55 ± 0.09	2.32 - 2.79
Cortical thickness (mm), right hemisphere		
Caudal ACC	2.42 ± 0.18	1.95 - 2.88
Rostral ACC	2.53 ± 0.22	2 - 3.12
Lateral OFC	2.62 ± 0.12	2.39 - 2.92
Medial OFC	2.33 ± 0.11	2.1 - 2.61
DLPFC	2.57 ± 0.08	2.42 - 2.78

Note: ACC, anterior cingulate cortex; OFC, orbitofrontal cortex; DLPFC, dorsolateral prefrontal cortex.

4.4 Neurobehavioral variables in recidivism prediction

The fourth aim of this thesis, pursued in *Study II*, was to explore how neurobehavioral variables associated with disinhibitory psychopathology may be used to improve the prediction of recidivism. With an average time at risk of almost 11 years and with almost one third of the sample still under forensic psychiatric care when follow-up ended, sixteen patients (36%) received a new conviction during the follow-up period. Most crimes were non-violent, although seven patients (16% of the total sample) were convicted of violent crimes, including robbery, unlawful threat, and assault.

Overall, the predictive performance of the Baseline model was modest (Table 4.5), but four variables associated with disinhibitory psychopathology — a Cluster B personality disorder, lower age at first crime, substance use disorder, and a higher degree of psychopathic traits — increased the

Table 4.5: Predictive performance of Baseline and Extended models.

Measure	Baseline	Extended
Area under the curve	0.69	0.81
Scaled Brier score	0.08	0.25
Accuracy	0.64	0.82
Sensitivity	0.63	0.75
Specificity	0.64	0.86
Positive predictive value	0.50	0.73
Negative predictive value	0.76	0.86

probability of being classified as a recidivist (Figure 4.2). Adding resting-state rCBF measurements increased predictive performance across all metrics (Table 4.5), and the Extended model correctly classified two additional recidivists and six additional non-recidivists compared to the Baseline model, resulting in 12 out of 16 recidivists and 24 out of 28 non-recidivists being correctly classified. In the Extended model, a combination of neurobehavioral variables associated with disinhibitory psychopathology emerged as the most important (Figure 4.2), with lower right parietal lobe rCBF, a Cluster B personality disorder, lower age at first crime, and lower left parietal lobe rCBF increasing the probability of being classified as a recidivist.

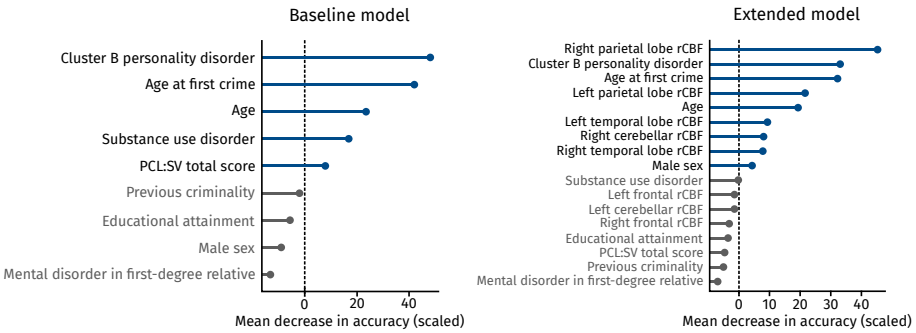

Figure 4.2: Individual variable importance in the Baseline and Extended models. Higher values represent a higher decrease in the accuracy of the model, should that variable be omitted. Variables with a positive impact on accuracy are highlighted.

5. Discussion

5.1 Comments on main findings

5.1.1 Prevalence of disinhibitory psychopathology

In *Study I*, the expression of disinhibitory psychopathology in violent offenders was quantified, based on the four-facet PCL-R structure, as impulsive lifestyle and antisocial psychopathic traits. These two facets correspond well with the construct of disinhibitory psychopathology; recent research on Swedish male offenders observed large associations between scores on the ESI-BF_{DIS} and the impulsive lifestyle (r = 0.68) and antisocial (r = 0.66) facets (Pauli et al., 2019)[1] . Similarly, previous research on US offenders, using a subset of ESI items to construct a scale similar to the ESI-BF_{DIS}, demonstrated moderate, positive associations (r = 0.40-0.46) with PCL-R impulsive lifestyle facet scores as well as with symptoms of adult antisocial behavior (Venables et al., 2014; Venables & Patrick, 2012).

Notably, the average scores on these facets were higher in *Study I* than previous estimates representative of Swedish male (primarily violent) offenders (Neumann et al., 2013; Pauli et al., 2019), and were similar to or even higher than previous reports from comparable international offender samples (e.g., Coid et al., 2009; Gray et al., 2019; Jeandarme et al., 2017; Lindberg et al., 2009; Mokros et al., 2011; Venables et al., 2015; Zwets et al., 2015). The high prevalence of disinhibitory psychopathology among male violent offenders in *Study I* is thus in line with previous findings, and may be linked to negative consequences. For instance, a study on mentally disordered offenders with a personality disorder found that both PCL-R impulsive lifestyle and antisocial facet scores were positively associated with several measures of physical aggression (Zwets et al., 2015).

Although an area of active research with sometimes contradictory findings, the impulsive lifestyle and antisocial facets may also be the most relevant aspects of psychopathy in terms of predicting both general and violent recidivism (Jeandarme et al., 2017; Kennealy et al., 2010; Olver & Wong, 2015). Unfortunately, high scores on the PCL-R impulsive lifestyle facet has also been associated with poor performance during treatment

[1] Readers are reminded that scores on the TriPM Disinhibition component are referred to as scores on the ESI-BF_{DIS}.

sessions (Olver, 2016), further highlighting the importance of the disinhibitory aspects of the psychopathy construct, both clinically in terms of treatment, and for the criminal justice system more broadly due to the association with recidivism.

In *Study II*, the expression of disinhibitory psychopathology among mentally disordered offenders was quantified using assessments of Cluster B personality disorders, PCL:SV total score, criminal history, and recidivism. Cluster B personality disorders were substantially more frequent among recidivists than among non-recidivists (38% vs. 4%). Since all Cluster B personality disorders are characterized by disinhibitory tendencies (Casillas & Clark, 2002; Taylor et al., 2006) and are positively associated (r = 0.23-0.54) with the impulsive lifestyle aspects of psychopathy (Blackburn, 2007), the large and robust difference in frequency of a Cluster B personality disorder signals a high prevalence of disinhibitory psychopathology in this group of mentally disordered offenders. In line with the higher frequency of Cluster B personality disorders, recidivists scored approximately three points higher than non-recidivists on the PCL:SV, and were also younger when committing their first crime (23 years compared to 35 years among non-recidivists); the latter in particular being indicative of a pattern of persistent offending characteristic of disinhibitory psychopathology (cf., Section 1.5.1). Thus, *Study II* suggests that a higher prevalence of disinhibitory psychopathology among mentally disordered offenders may be associated with an elevated risk of persistent offending (but see Section 5.2.3 and Figure 5.1), in line with research recognizing disinhibitory tendencies as among the most relevant risk factors for recidivism (Bonta et al., 2014; Skeem et al., 2014).

In *Study III*, the expression of disinhibitory psychopathology among violent mentally disordered offenders was quantified using scores on the ESI-BF_{DIS} in conjunction with assessment of DSM-5 diagnoses and criminal history. Unfortunately, since the ESI-BF is a relatively novel instrument, few studies have been conducted on mentally disordered offenders. Although differences in legislation concerning mentally disordered offenders render international comparisons difficult, the average ESI-BF_{DIS} score among violent mentally disordered offenders in *Study III* was higher (26.6 vs. 21.4) than a recent estimate based on Portuguese mentally disordered offenders (Pasion et al., 2018), and almost equivalent (26.6 vs. 26.7) to data from Dutch mentally disordered offenders (van Dongen et al., 2017). The ESI-BF_{DIS} scores were

accompanied by a high prevalence of diagnoses associated with disinhibitory psychopathology, including antisocial personality disorder, substance-related and addictive disorders, and ADHD. Together, these results corroborate findings from *Study II* in suggesting a relatively high prevalence of disinhibitory psychopathology among mentally disordered offenders. The findings of *Study III* also align well with reports of a robust association (r = 0.45) between symptoms of adult antisocial behavior and scores on a subset of ESI items resembling the ESI-BF_{DIS} scale (Venables & Patrick, 2012), suggesting that the diagnostic features of these disorders indeed tap a core disinhibitory tendency.

In addition to ESI-BF_{DIS} scores and DSM-5 diagnoses, most mentally disordered offenders in *Study III* had a history of being persistently violent, with an average age at first criminal sentence of 21 years (comparable to recidivists in *Study II*), a young average age at first reported crime, and with multiple previous criminal sentences. This pattern of persistent offending is, as discussed in relation to *Study II*, also characteristic of disinhibitory psychopathology, and is in broad agreement with recent research showing that disinhibitory tendencies among juvenile offenders are associated with elevated levels of aggression, procriminal sentiments, and a higher clinically rated risk of recidivism (Laurinavičius et al., 2020). Similarly, a recent prospective study from adolescence to young adulthood found that disinhibitory tendencies were associated with both physical and verbal aggression, elevated levels of conduct disorder and antisocial personality disorder symptoms, as well as more primitive strategies for coping with stress (Kyranides et al., 2017). Together, these recent studies, along with *Studies II* and *III*, demonstrate how early manifestations of disinhibitory psychopathology may increase the risk of long-term, persistent offending.

In *Study IV*, as well as in the control group of *Study III*, the expression of disinhibitory psychopathology was quantified using scores on the ESI-BF_{DIS}. As might be expected, research on participants from the general population (e.g., community and undergraduate samples) is more abundant than research on mentally disordered offenders. Therefore, to put the results from *Studies III* and *IV* into a broader context, previously reported ESI-BF_{DIS} scores, from community and undergraduate samples (Byrne et al., 2016; Esteller et al., 2016; Paiva et al., 2020; Pasion et al., 2018; Ribes-Guardiola et al., 2020; Ribes-Guardiola et al., 2020; Somma et al., 2019; Stanton et al., 2020; Tuvblad et al., 2019; van Dongen et al., 2017), is presented in Table 5.1.

Table 5.1: Disinhibitory psychopathology in community and undergraduate samples.

Source	N	Age (Mean ± SD)	ESI-BF_{DIS} (Mean ± SD)
Byrne et al., 2016	93	18-22[a]	15.39 ± 13.60
Esteller et al., 2016	180	20.62 ± 4.01	17.71 ± 8.23
van Dongen et al., 2017	496	27.70 ± 13.09[b]	11.46 ± 7.80
Pasion et al., 2018	48	32.0 ± 11.6	15.3 ± 7.9
Tuvblad et al., 2019	463[c]	19-20[a]	13.17 ± 9.25
Tuvblad et al., 2019	552[d]	19-20[a]	11.76 ± 8.08
Somma et al., 2019	1082	34.28 ± 13.10	17.3 ± 8.11
Paiva et al., 2020	1833	23.8 ± 7.64	15.1 ± 7.79
Ribes-Guardiola et al., 2020	161	20.55 ± 4.51	14.64 ± 7.67
Ribes-Guardiola et al., 2020	69[c]	20.57 ± 4.13[e]	16.99 ± 7.20
Ribes-Guardiola et al., 2020	131[d]	20.57 ± 4.13[e]	14.62 ± 7.13
Stanton et al., 2020	700	32.8 ± 10.1	21.1 ± 10.9
Stanton et al., 2020	527	19.2 ± 1.5	13.4 ± 7.6

Note: [a] Mean age missing [b] Age missing from 10 participants [c] Males only [d] Females only [e] Aged based on total sample

The average ESI-BF_{DIS} score of the studies presented in Table 5.1 is 15.23, which is similar to results from *Study IV* (M = 14.20), although more than double that observed among controls in *Study III* (M = 7.50). Furthermore, the estimates in Table 5.1 suggests a *positive* association ($r \sim 0.40$) between ESI-BF_{DIS} score and age, but control group participants in *Study III* were, on average, approximately ten years older than the young adults included in *Study IV*. Although *Study III* did include undergraduate students, who generally score lower than community participants, it remains unclear whether that can account for such a large difference. Another possibility is underreporting of sensitive information; control group participants in *Study III* answered questionnaires in the presence of a researcher, whereas participants in *Study IV* answered questionnaires online.

Mentally disordered offenders in *Study III* scored on average over 10 points higher on the ESI-BF_{DIS} than both young adults in *Study IV* and the average of previous undergraduate and community estimates presented in Table 5.1. While a higher prevalence of disinhibitory psychopathology among mentally disordered offenders was, of course, expected, it nevertheless signals a need for disinhibition-focused treatment, especially since *Study II* offers further evidence of its association with recidivism. Unfortunately, evidence addressing the treatment of mentally disordered offenders broadly (Howner et al., 2018), and antisocial behavior specifically (Brazil et al., 2018), is severely lacking. This, in conjunction with the fact that risk factors for persistent criminality — of which many

are associated with disinhibitory psychopathology — basically are the same whether an offender is mentally disordered or not (Bonta et al., 2014; Skeem et al., 2014), the question remains whether disinhibitory psychopathology receives enough attention in forensic mental health settings.

Notably, *Study IV* confirms that disinhibitory tendencies are present also in the general population. Indeed, it is imperative to keep in mind that disinhibitory psychopathology is a *dimensional* construct that is relevant even at subclinical levels. For instance, Drislane et al. (2014) found that higher ESI-BF_{DIS} scores were associated with normal-range personality traits such as increased aggression and increased reaction to stress among undergraduate students. In addition, Ljubin-Golub et al. (2019) found that higher ESI-BF_{DIS} scores were associated with academic cheating among college students, perhaps due to a lack of restraint in the face of potential reward. Disinhibitory psychopathology is a relevant construct in the workplace as well; Sutton et al. (2020, p. 12) found that elevated ESI-BF_{DIS} scores were "almost synonymous with destructive leadership", including abusive supervision and increased rates of burnout among managers.

In sum, *Studies I-IV* demonstrate a relatively high prevalence of disinhibitory psychopathology, both among offenders with and without a severe mental disorder and in the general population, in line with previous Swedish and international estimates. Together, these studies add to a broad literature documenting a considerable prevalence of disinhibitory psychopathology that — despite its association with adverse outcomes, including persistent offending and recidivism — may in many cases remain unidentified or untreated. Similarly, given its association with negative outcomes even at subclinical levels, and since research has demonstrated how early manifestations of disinhibitory psychopathology may increase the risk of persistent offending, there is much to gain from increased knowledge about the prevalence of disinhibitory psychopathology in the general population.

5.1.2 Disinhibitory psychopathology and neurocognition

In *Study I*, neurocognitive deficits were assessed using four different computer-based EF tasks. Although effects ranged between small and

medium, reduced mean initial thinking time in the SOC task was robustly associated with higher scores on the PCL-R impulsive lifestyle and antisocial facets. Interestingly, these facets have been positively associated with scores on two domains of the Barratt Impulsiveness Scale (Snowden & Gray, 2011): the Motor scale, containing items such as "I act on the spur of the moment" and "I do things without thinking", and the Nonplanning scale, containing items such as "I say things without thinking" and "I am more interested in the present than the future" (Patton et al., 1995). Thus, lower mean initial thinking times in the SoC task may manifest, more generally, as an impulsive approach to everyday tasks involving planning and problem-solving, which would be in line both with early descriptions of psychopathic individuals as highly impulsive (McCord & McCord, 1964, p. 8) and with the broader construct of disinhibitory psychopathology. Although not explicitly designed to target an impulsive approach to problem-solving, therapeutic interventions aimed at improving social problem-solving skills, such as the 'Stop & Think!' (Huband et al., 2007; McMurran et al., 2001) may nevertheless be helpful, and should be further explored. In addition, since intriguing findings from previous research suggest that performing a planning and problem-solving task leads to rapid, dynamic changes in frontal neural activity (Beauchamp et al., 2003), therapeutic interventions should also be evaluated using neuroimaging techniques.

While not corroborated by Bayesian analyses, less efficient strategic thinking during the SWM task was associated with increased antisocial psychopathic traits, in line with previous findings (Sadeh & Verona, 2008). Given the central role of the DLPFC in planning and problem-solving (e.g., Nitschke et al., 2017), and since damage to the DLPFC may result in impaired ability to manipulate verbal and spatial knowledge in working memory (Barbey et al., 2013; Manes et al., 2002), the results of *Study I* does hint at possible DLPFC dysfunction in individuals characterized by impulsive lifestyle and antisocial psychopathic traits. Furthermore, OFC lesioned patients demonstrate impulsive behavior that has been theorized to reflect a desire for immediate reward despite potential negative outcomes (Berlin et al., 2004), reminiscent of the early description of disinhibitory psychopathology by Gorenstein & Newman (1980). Although speculative, it is possible that a desire to quickly solve problems results in shorter mean initial thinking times in the SOC task, which thus could be indicative of OFC dysfunction. Since neuroimaging data was not available in *Study I*, these findings encouraged the inclusion of both the DLPFC and OFC as regions of interest in *Study IV*.

Surprisingly, no measure from the SST was robustly associated with psychopathic traits in *Study I*. This finding is in contrast to recent research on Swedish offenders, where negative associations between SST stop-signal reaction time and all aspects of psychopathy were observed (Pauli et al., 2019), indicative of *better* inhibitory capacity in psychopathy. Several possible explanations emerge: first, the SST measures the ability to inhibit an already initiated response, and it is possible that the impulsive-antisocial features of psychopathy — and disinhibitory psychopathology more generally — is more closely associated with the ability to inhibit a prepotent response altogether. A second and perhaps more likely explanation is that when performance on the SST is either too poor or to good, the model's assumptions are violated and no measures are available (e.g., Bø et al., 2016). Given the high rate of attrition (N = 53) in *Study I*, perhaps due to the fact that the SST was placed last in the battery of CANTAB tests, the SST — at least as configured in the CANTAB — may not be optimal in samples where either very high or very low levels of inhibition are expected. Thus, *Studies III* and *IV* instead used the simpler Go/NoGo task, which measures the ability to inhibit a prepotent response altogether.

Although the difference was relatively small and non-robust, mentally disordered offenders had approximately 5% better NoGo accuracy than controls in *Study III*. Follow-up analyses revealed that better NoGo accuracy was associated with longer median NoGo response times (i.e., the response time on *failed* NoGo trials), suggestive of *better* capacity for impulse control in mentally disordered offenders than controls. However, longer median NoGo response times were also associated with increased NoGo P3 latency, and since NoGo P3 latency is believed to be an index of neural efficiency (van Dinteren et al., 2014), less efficient neural information processing may have attenuated the prepotency to responding too quickly on NoGo trials. Thus, it is possible that a reduced neural capacity to respond within the permitted response window may have inadvertently resulted in a slightly higher NoGo accuracy among mentally disordered offenders. Future research should consider using longer response windows to possibly counteract this effect.

The average NoGo accuracy in *Study IV* was lower than the NoGo accuracy among both mentally disordered offenders and controls in *Study III*, in line with the robust negative association between ESI-BF_{DIS} scores and NoGo accuracy observed in *Study IV*. Interestingly, the estimated association between ESI-BF_{DIS} scores and NoGo accuracy was similar in

Study IV (ρ = -0.19) and among mentally disordered offenders in *Study III* (ρ = -0.18). Still, no meaningful association between ESI-BF_{DIS} scores and NoGo accuracy was found across the whole sample, nor among controls, in *Study III*. These partially divergent findings makes it difficult to fully evaluate the usefulness of Go/NoGo task performance as a neurobehavioral correlate of disinhibitory psychopathology. Indeed, some have argued that Go/NoGo task performance is different from real-world situations in which the inhibition of a response is required (Aron, 2011). On the other hand, a recent meta-analysis by Allom et al. (2016) found that inhibitory control training provides at least a short-term reduction of harmful behaviours, such as alcohol consumption. The authors made another, in the context of this thesis, compelling finding: studies employing Go/NoGo tasks observed larger effects than studies employing variants of the SST. Nevertheless, findings have been inconclusive (e.g., Jones et al., 2018), and knowledge about the neuroplasticity underlying improvements in inhibitory control is still at an early stage (Spierer et al., 2013). Future research should explore whether inhibitory control training has an effect on disinhibitory psychopathology, especially in the long-term, and whether such an effect is associated with altered brain structure and function. Since inhibition may be operationalized using dozens of different tasks (e.g., Schoemaker et al., 2013), future studies should also consider a multimethod approach (see Section 5.2.3).

In sum, *Studies I, III* and *IV* suggest that disinhibitory psychopathology may be associated with specific neurocognitive impairments in the form of an impulsive approach to planning and problem-solving and a reduced capacity for inhibitory control. Since situations that require planning and problem-solving are faced every day, a persistently impulsive approach could be detrimental for the ability to adapt to and interact with the environment. Likewise, inhibitory control is crucial for navigating an unpredictable and rapidly changing environment; a reduced capacity makes it more difficult to do what is appropriate or necessary, rather than relying on old habits or be tempted by external stimuli. Nevertheless, the inconsistent findings between *Study III* and *Study IV* combined with the lack of an effect of the SST task in *Study I* signal a need for further research into the role of different types of response inhibition, as well as possible therapeutic interventions and associated neurobiological alterations.

5.1.3 Disinhibitory psychopathology and neurobiology

Recidivists in *Study II* had lower bilateral parietal lobe and bilateral cerebellar rCBF, and higher bilateral temporal lobe rCBF, than non-recidivists, although results differed in robustness. Recidivists' reduced parietal rCBF is in line with previous research documenting lower parietal rCBF and lower parietal glucose metabolism in violent, impulsive, and aggressive individuals (Raine et al., 1997; Siever et al., 1999; Soderstrom et al., 2000). One parietal region in particular, the inferior parietal lobule, is involved in response inhibition (Steele et al., 2013), and it is possible that recidivists' reduced parietal rCBF reflects impaired response inhibition, which subsequently increased the risk of recidivism. Cautious interpretation is warranted, however; although this explanation is in line with recidivists' overall higher prevalence of disinhibitory psychopathology compared to non-recidivists, the parietal lobe is structurally and functionally diverse, consisting of multiple subregions, each with its own cortical and subcortical connections (Aversi-Ferreira et al., 2010). The coarse-grained resolution of SPECT measurements unfortunately makes it impossible to infer whether recidivists had reduced rCBF in the whole parietal lobe, in one specific subregion such as the inferior parietal lobule, or in several subregions.

The lack of spatial resolution may also explain why no difference in frontal lobe rCBF was observed, since previous research has associated criminal and antisocial behavior with structural and functional abnormalities in specific subregions, such as the ACC, DLPFC, and OFC (Raine, 2019; Yang & Raine, 2009). Recidivists demonstrated increased temporal lobe rCBF compared to non-recidivists, while previous research, for instance, has observed reduced rCBF in the right lateral temporal lobe in individuals with borderline or antisocial personality disorder compared to healthy controls (Goethals et al., 2005). Again, the lack of spatial resolution is an issue, since previous studies have associated disinhibitory psychopathology with increased reactivity to fearful facial expressions in deeper temporal structures, such as the amygdala (Coccaro et al., 2007; Cunha-Bang et al., 2019; Dotterer et al., 2017), suggesting that subcortical structures may offer further clues into the emotional aspects of disinhibitory psychopathology.

Recidivists also demonstrated reduced cerebellar rCBF compared to non-recidivists. Although it remains a relatively unexplored region, it is possible that reduced cerebellar perfusion is associated with neurocognitive deficits and disinhibited tendencies similar to those observed in

cerebellar cognitive affective syndrome (Schmahmann, 2010), which could increase the risk of recidivism. More recent research has found evidence of increased cerebellar volume in persistent violent offenders (Leutgeb et al., 2015; Tiihonen et al., 2008), providing further support for cerebellar structure and function as a potential neurobehavioral correlate of disinhibitory psychopathology that should be explored in future studies.

In *Study III*, mentally disordered offenders showed prolonged NoGo P3 latency and, although with a smaller effect and wider HDIs, reduced NoGo P3 amplitude compared to controls. These findings are in line with previous research demonstrating reduced NoGo P3 amplitude in antisocial individuals (e.g., Guan et al., 2015; Verona & Bresin, 2015), and the observed effects are similar to meta-analytic findings of reduced P3 amplitude and prolonged P3 latency in violent, impulsive, and aggressive samples (Gao & Raine, 2009). The NoGo P3 component is believed to be generated by several, concurrently activated brain networks (Huster et al., 2010), although perhaps primarily by the dorsal ACC (Hong et al., 2017). Since the dorsal ACC is a key region involved in the monitoring and evaluation of errors, rewards, and conflicts (Heilbronner & Hayden, 2016; Whelan et al., 2012), reduced NoGo P3 amplitude and delayed NoGo P3 latency may reflect aberrant post-synaptic neurotransmission associated with the monitoring and evaulation of one's own behavior. Importantly, such aberrant neurotransmission could have real-world consequences; the gyral regions of the dorsal ACC have been shown to be involved in predicting and monitoring outcomes of social decision-making (Apps et al., 2013). Furthermore, since increased NoGo P3 latency, which is believed to reflect neural inefficiency (van Dinteren et al., 2014), was associated with higher ESI-BF_{DIS} scores, NoGo P3 latency represents a promising neurobehavioral correlate of disinhibitory psychopathology that warrants further research.

Surprisingly, no robust association between NoGo P3 amplitude and ESI-BF_{DIS} was observed, despite the lower amplitude among mentally disordered offenders compared to controls, and despite recent findings of a negative association between P3 amplitude and ESI-BF_{DIS} scores (Ribes-Guardiola et al., 2020). Differences in sample characteristics between *Study III* and the study by Ribes-Guardiola et al. (2020) (i.e., mentally disordered offenders plus healthy controls vs. undergraduate students) and sample sizes (i.e., N = 47 vs. N = 142) may in part explain these divergent findings. Notwithstanding these differences,

further research — especially on mentally disordered offenders — is necessary to elucidate the role of NoGo P3 amplitude in disinhibitory psychopathology.

In *Study IV*, higher ESI-BF_{DIS} scores were moderately associated with increased thickness of the right lateral OFC. Although a diverse region, there is accumulating evidence of increased OFC activity during reward processing in impulsive-antisocial individuals (Murray et al., 2018), and the lateral OFC specifically appears involved in risky decision-making, possibly due to overriding a wish to abstain from unwanted or punished behavior (Elliott, 2000). Thus, increased thickness of the right lateral OFC may promote a kind of risky decision-making in the face of potential reward akin to the failure of self-control and disregard long-term goals in favour of instant gratification described by Gorenstein & Newman (1980).

Higher ESI-BF_{DIS} scores were also moderately associated with reduced thickness of the right DLPFC. Since the DLPFC is involved in most EFs, including working memory, shifting, and inhibition, as well as in planning and problem-solving (Nitschke et al., 2017; Zhang & Iwaki, 2019), there may be several possible explanations for the association between reduced DLPFC cortical thickness and higher prevalence of disinhibitory psychopathology. ESI-BF_{DIS} scores were negatively associated with NoGo accuracy in *Study IV*, and since the DLPFC is active during Go/NoGo task performance (Beeli et al., 2008; Menon et al., 2001), it is possible that reduced right DLPFC thickness resulted in impulse control problems that, in turn, were associated with the increased disinhibitory psychopathology tapped by the ESI-BF_{DIS}. Still, since no other neurocognitive functions were assessed, this conclusion cannot be firmly drawn, and future studies would need to include additional tasks, in line with the multimethod approach suggested in Section 5.2.3. Nevertheless, the results of *Study IV*, when viewed in light of the findings of *Study I* (which were indicative of DLPFC dysfunction in individuals characterized by impulsive lifestyle and antisocial psychopathic traits), suggest that the DLPFC may be a promising neurobehavioral correlate of disinhibitory psychopathology.

Higher scores on the ESI-BF_{AGG} subscale were robustly associated with a thicker bilateral medial OFC, as well as with a thicker right caudal ACC. Since right medial OFC thickness was robustly associated with higher NoGo accuracy, while NoGo accuracy was unrelated to ESI-BF_{AGG} scores, the positive association between ESI-BF_{AGG} scores and medial

OFC thickness may primarily reflect non-disinhibitory aspects of the externalizing spectrum, such as callousness and deficient empathy. Indeed, the ESI-BF_{AGG} subfactor is associated with both interpersonal-affective and impulsive-antisocial psychopathic traits, and thus serves as a link between the core impulse control problems of disinhibitory psychopathology and the interpersonal-affective features of psychopathy (Patrick et al., 2013; Venables & Patrick, 2012). Speculatively, increased thickness of the medial OFC — a region involved in reward processing — may increase the likelihood of finding it enjoyable to hurt other people (Glenn & Raine, 2009), or may facilitate an increased ability to control fear along with reduced anxiety (Kühn et al., 2011); both of which have been suggested as components of psychopathy (Neumann et al., 2013). Thus, a thicker medial OFC may be a neurobehavioral correlate both of disinhibitory psychopathology and of the broader construct of externalizing psychopathology. Interestingly, the HiTOP model does indeed distinguish between 'Disinhibitory externalizing' and 'Antagonistic externalizing', where the latter is characterized by callousness, deceitfulness, and egocentricity (e.g., Conway & Simms, 2020).

ESI-BF_{SUB} scores were associated with a thicker right DLPFC, contrary to previous research (e.g., Durazzo et al., 2011; Jacobus et al., 2016). A possible explanation is that the DLPFC is involved in the formation of drug-related working memories, in turn associated with increased self-reported substance use (Goldstein & Volkow, 2011). No association between ESI-BF_{SUB} scores and ACC or OFC thickness was observed, despite being key regions involved in substance abuse (Ersche et al., 2013; Goldstein & Volkow, 2011). Since *Study IV* used community volunteers, it is possible that individual variation in cortical morphology was too subtle to detect, compared to research on individuals at the extreme end of the substance abuse spectrum (i.e., as in Ersche et al., 2013). Future studies should have this in mind, and preferably use larger samples and additional measures of substance use.

No robust association between ACC thickness and ESI-BF_{DIS} scores were observed in *Study IV*, despite a positive association between right rostral ACC thickness and response inhibition performance, and despite the proposed role of the dorsal ACC in NoGo P3 latency in *Study III*. Although incorrect responses during the Go/NoGo task activates the rostral ACC (Kiehl et al., 2000), suggesting that the rostral ACC plays an important role in monitoring and evaluating behavior, and while cortical thickness has been positively associated with neural activity in the ACC during

the Go/NoGo task (Hegarty et al., 2012), it is important to acknowledge that there is no one-to-one mapping between the brain's structural and functional networks (Batista-García-Ramó & Fernández-Verdecia, 2018). Thus, structural findings in one region may not be directly tied to functional findings in that same region. One way to further investigate the relationship between dorsal ACC structure and function is concurrent (f)MRI and EEG acquisition. Furthermore, some of the regions investigated, such as the lateral OFC and the DLPFC, are functionally connected (Kahnt et al., 2012), and the findings in *Study IV* may therefore reflect disrupted communication between these regions, in turn caused by structural *or* functional aberrations. Future studies should consider the use of both structural and functional connectivity measures to explore this possibility further.

In sum, *Studies II, III* and *IV* suggest that disinhibitory psychopathology may be associated with distinct neurobiological alterations, including changes in brain metabolism in cerebellar, parietal, and temporal lobe regions, less efficient post-synaptic neurotransmission in the dorsal ACC, and variations in cortical thickness primarily in the OFC and DLPFC. These regions are involved in processes associated with the ability to control and regulate both emotions and behavior, including behavioral monitoring, decision-making, and reward evaluation, and therefore represent promising neurobehavioral correlates of disinhibitory psychopathology worthy of further exploration. Future research should expand the scope of regions under consideration to include subcortical structures, and should supplement analyses of regional structure and function with measures of connectivity.

5.1.4 Neurobehavioral variables in recidivism prediction

Study II investigated whether neurobehavioral variables associated with disinhibitory psychopathology could improve the prediction of recidivism, and found that a combination of neurobiological and behavioral data offered *incremental* predictive performance over using traditional risk factors based on behavior only. Importantly, since recidivists did not differ from non-recidivists in primary diagnosis, time at risk, average length of stay, or the number of patients still under forensic psychiatric care at the end of follow-up, any increase in predictive performance should not be attributable to these variables. An interesting pattern

of neurobehavioral variables that, to varying degrees, are associated with disinhibitory psychopathology emerged as the most important predictors in the Extended model: reduced bilateral parietal lobe rCBF, a Cluster B personality disorder, and a lower age at first crime. Although the emergence of these variables makes theoretical sense (see Section 5.1.3), the primary purpose of *Study II* was not to *explain* recidivism, but to *predict* recidivism. The difference between explanation and prediction is often overlooked in the social sciences (e.g., Yarkoni & Westfall, 2017), perhaps due to the field's unfamiliarity with machine learning methods, or due to its traditional focus on explaining behavior. Since the primary objective of predictive modeling is to generate accurate predictions, whether the method uncovers underlying causal mechanisms or not is of secondary importance (Shmueli, 2010).

In the best of worlds, all offenders, mentally disordered or not, would receive every form of treatment available. In reality, resources are constrained. Thus, rather than replacing well-trained staff members, machine learning models — able to incorporate vast amounts of data and uncover complex relationships — may be used as decision support systems. A potential use case of such a system would be to aid in directing resources to offenders with the highest risk of recidivism, similar to how artificial intelligence algorithms are employed in other clinical domains, such as in radioimaging (e.g., Hosny et al., 2018). Used this way, the decision support system would be in line with the risk, need, and responsivity model of offender rehabilitation, which posits that preventive efforts should target higher risk rather than lower risk offenders (e.g., Polaschek, 2012). Still, while improvements in the prediction of recidivism may be achievable if neuroimaging data is incorporated, the exploratory nature of *Study II* does come with several caveats. Due to the small and heterogenous sample, studies in larger samples of mentally disordered offenders are necessary before generalizations can be made. Other outcome measures, such as the number of adverse incidents during in-patient care or self-reported criminality, as well as different baseline risk factors, may be used to further establish whether neuroimaging data offers incremental predictive performance. It is possible that SPECT measurements are not available, or ethically defensible, in risk assessment situations, and other neuroimaging techniques should be explored. Finally, several ethical challenges (see Section 5.2.4), including biases, stigmatization, and privacy concerns, must be given careful consideration before clinical application is feasible (Gkotsi & Gasser, 2016; Tortora et al., 2020).

In sum, the use of machine learning algorithms and neuroimaging data represents a potent combination that may have important clinical applications. A conceivable use case is as a decision support system that helps clinicians direct preventive efforts to high risk offenders. Before that, however, further studies in larger samples are required, preferably using different outcome measures, and several ethical challenges must be carefully considered.

5.2 General discussion

5.2.1 From basic research to clinical application

This thesis has primarily been concerned with basic research questions along the lines of "how does this work?" rather than applied research questions along the lines of "is this useful?". Since the goal of basic medical research should be to provide a foundation for clinical application, it is important to consider ways in which the results of this thesis provide added value from a clinician's point of view. Two directions emerge as particularly promising: refining the nosology of psychopathology, and guiding individualized treatment.

Categorical diagnoses have been the foundation of both mental health research and clinical practice for over a century, but this hegemony seems, slowly but surely, to be declining (Conway & Simms, 2020). For instance, as mentioned in Section 1.3.1, antisocial personality disorder represents a categorical clinical diagnosis that is in close proximity to the concept of disinhibitory psychopathology (McKinley et al., 2018), yet personality disorders are, according to current research, likely not categorical in nature (Hopwood, 2018). Furthermore, some argue that we may have reached the point where a deeper understanding of mental disorders is not possible if based solely on behaviors and self-reported thoughts and feelings (Clark et al., 2017), and that research based on multiple approaches (or 'levels of analysis') is required for a full — and maximally useful — understanding of psychopathology (Anderson, 1998; Perkins, Joyner, et al., 2020).

In contrast to consensus based nosologies such as the DSM, initiatives such as the RDoC and the HiTOP represent the cutting-edge of a new generation of nosological frameworks that may transform how we approach diagnosis and treatment. While emerging neuroscientific findings have

mapped poorly to traditional, categorical mental disorders (Kozak & Cuthbert, 2016), these new frameworks are constructed and refined "from the ground up" based on empirical research that incorporates several levels of analysis, including measures of brain structure and function as well as behavioral assessments (Hyman, 2007; Nelson et al., 2016). However, while promising, these frameworks are in need of further refinement before being suitable for clinical application, and this thesis has taken two key steps to that end: the use of dimensional assessments, and interfacing behavioral assessments with neurobiological measures (i.e., a 'multilevel approach').

Dimensional assessments are substantially more reliable and valid than categorical diagnoses (Markon et al., 2011), and thus pave the way for a more accurate assessment of disinhibitory psychopathology. This thesis has involved both the translation and application of a dimensional self-report instrument — the ESI-BF — which is recommended for use in the HiTOP framework (Kotov et al., 2018). Pending further validation studies, the ESI-BF could be used, for instance, to record and track the level of disinhibitory psychopathology in offenders across time, or as a screening tool to indentify individuals in the general population in need of further intervention.

In addition to dimensional assessments, the RDoC and HiTOP frameworks advocate the incorporation of a broad array of neurobehavioral measures, including P3 ERPs, structural brain scans, and response inhibition tasks, in order to bridge the gap between brain and behavior (Patrick et al., 2019; Perkins, Latzman, et al., 2020). This thesis provides further support for the potential clinical utilization of such neurobehavioral measures, both among offenders and in the general population. As an example, and as discussed in Section 5.1.3, the HiTOP model distinguishes between disinhibited externalizing (i.e., what this thesis refers to as disinhibitory psychopathology), and antagonistic externalizing (i.e., externalizing characterized by, for instance, callousness and deceitfulness). In the HiTOP model, both of these spectra lead to antisocial behavior, but they may also have distinct neurobiological correlates, which could affect the effectiveness of different treatment approaches. As this thesis has shown, disinhibitory psychopathology may be associated with functional aberrations in the dorsal ACC; a region involved in monitoring and evaluating behavior, and with structural alterations in the lateral OFC; a region involved in risky decision-making. Antagonistic externalizing, on the other hand, may be associated with structural

alterations in the medial OFC; a region involved in controlling fear and anxiety. Thus, while both disinhibited and antagonistic externalizing are associated with antisocial behavior, their distinct neurobiological correlates may signal a need for different approaches to treatment. By mapping observable patterns of behavior to neurobiological referents, it might be possible to yield insights that could be used for individualized treatment (e.g., Perkins, Joyner, et al., 2020), and recent research has taken important steps towards intervention guided by knowledge of neurobehavioral correlates.

A recent study by Sergiou et al. (2020) was the first to demonstrate reduced aggression in mentally disordered offenders following transcranial direct current stimulation (tDCS) to the ventromedial prefrontal cortex (a region bordering the OFC). Similarly, Campanella et al. (2017) used tDCS focused on the right inferior frontal cortex, another region involved in response inhibition, with concurrent EEG recordings and a Go/NoGo task. They found that boosting right inferior frontal cortex activity enhanced inhibitory control capacity by decreasing the amount of neural power (i.e., reducing the P3 amplitude) required to correctly inhibit a response. Since *Study III* found evidence of reduced NoGo P3 amplitude elicited using a Go/NoGo task in mentally disordered offenders — albeit putatively associated with dorsal ACC function — these recent findings set the stage for the use of neuromodulation in forensic settings, and highlight the importance of a neurobehavioral approach to disinhibitory psychopathology.

Importantly, the findings of this thesis are not limited to the severe end of the disinhibitory psychopathology spectrum. It has long been suspected that some neurobehavioral correlates of disinhibitory psychopathology may represent biomarkers indicating a risk of later, more severe manifestations. For instance, Iacono et al. (2002) showed that blunted P3 amplitudes in adolescent males were predictive of substance use problems three years later. Although the cross-sectional design limits interpretation, *Study IV* provides a starting point for further research into whether altered cortical thickness also represents a biomarker of a risk for later, more severe forms of disinhibitory psychopathology.

5.2.2 Implications for future research

Before the findings of this thesis can be used to guide clinical practice, the specificity and generalizibility of neurobehavioral correlates of disinhibitory psychopathology must be further tested, especially in sam-

ples of mentally disordered offenders, where research remains scarce. Future research should examine a broader range of neurobiological regions, including subcortical structures, and consider using additional neuroimaging methods and neurocognitive assessments. It may also be advantageous to focus on structural and functional neural connectivity rather than the structure and function of individual brain regions in isolation. Multicenter studies and international collaboration is recommended in order to increase sample sizes and enable the use a broader range of modeling techniques. Additional challenges to overcome include time-consuming assessments, relatively expensive equipment, and costs of training and maintaining staff to carry out measurements.

Constructs from nearby fields should also be acknowledged. For instance, the construct of 'self-control', common in the criminological literature, bears striking resemblance to the definition of 'disinhibition' used in the current thesis. Individuals low in self-control, as described by Gottfredson & Hirschi (1990, p. 89) in their seminal work *A General Theory of Crime*, seek immediate gratification of desires, tend to be adventuresome and lacking in the cognitive capacity to plan ahead, and have little concern for long-term negative consequences. Indeed, the description by Gottfredson & Hirschi (1990) is reminiscent both of the early work by Gorenstein & Newman (1980) and of the findings presented in the current thesis. Interestingly, recent work has begun to incorporate the two terms into the same model of disinhibitory psychopathology (Venables, Foell, Yancey, Beaver, et al., 2018), paving the way for future research to continue along the same path.

Future research may also want to consider biological and psychosocial influences on the development of disinhibitory psychopathology. For instance, while some neurobiological correlates of disinhibitory psychopathology may have a strong genetic component (Yang et al., 2012), both Swedish and international studies have demonstrated considerable gene-environment interaction, such that shared environmental influences on antisocial behavior are exacerbated in disadvantaged neighborhoods (Burt et al., 2020; Tuvblad et al., 2006). In addition, environmental influences such as lead exposure have been associated with volume decrements in regions associated with disinhibitory psychopathology, including the ACC (Cecil et al., 2008). Likewise, the accumulation of maternal health risks during pregnancy has been associated with a sharp increase in the risk of later behavioral problems among male offspring (Jackson & Vaughn, 2018). Unravelling the link between

familial and environmental influence, neighborhood disadvantage, and associated neurobehavioral correlates of disinhibitory psychopathology therefore remains an important avenue for further research.

5.2.3 Strengths and limitations

This thesis has employed a wide array of methods, with several accompanying strengths and limitations. First, the computational resources used in each study are publicly and freely available for anyone to review and reuse, which should help facilitate reproducibility (Allen & Mehler, 2019). Second, a notable strength is the use of robust, Bayesian statistical models, which — assuming one agrees with the assumptions inherent in Bayesian statistical inference; see Appendix B — offers several advantages over frequentist statistics, including the ability to make genuine probability statements that remain equally valid regardless of sample size (Wagenmakers et al., 2018). Third, when using frequentist statistics, currently recommended methods were employed. For instance, *Study II* used Barnard's test instead of the more common Fisher's exact test. Since Fisher's exact test assumes fixed margins, some argue that it is rarely applicable in practice (Lydersen et al., 2009), and the more computationally intensive Barnard's test has been recommended as a viable alternative (for details, see Fagerland et al., 2017). Likewise, as with the fixed-margins assumption in Fisher's exact test, two groups of participants seldom — if ever — have the same variance (Delacre et al., 2017), which is why the recommended Welch's t-test was used in *Study II*. Fourth, interpretation of *p*-values and confidence intervals solely based on a frequentist ideas of statistical significance has been avoided as much as possible (Wasserstein et al., 2019).

Notwithstanding these strengths, there are two methodological limitations that, arguably more than any other, limit the robustness and clinical application of the findings reported in this thesis: the small sample sizes of *Studies II-IV*, and use of a bivariate mapping approach. Although Bayesian inferences are valid and often reasonable even with small sample sizes (wheras frequentist statistics rely on the central limit theorem), Bayesian inferences do improve as the sample size increases (Brutti et al., 2014; Depaoli & van de Schoot, 2017). Thus, larger samples are required in order to achieve increased precision of estimated parameters. Likewise, although the random forest algorithm used in *Study II* does perform well with small sample sizes, generalizability is a concern, especially with heterogenous samples.

In the bivariate mapping approach, different indicators (i.e., neurobehavioral variables) are, one by one, associated with a construct of interest (i.e., disinhibitory psychopathology). A multimethod approach, which combines several tasks, assessments, and instruments into latent variables of the construct of interest, increases reliability and construct validity while reducing measurement error and facilitating greater chances of replicability (Venables, Foell, Yancey, Kane, et al., 2018). Interestingly, Venables, Foell, Yancey, Kane, et al. (2018) recently developed a multimethod model of disinhibitory psychopathology with data from a sample of undergraduate volunteers. Further validating this model using additional variables (such as measures of brain structure) in different samples (such as offender populations) is a promising direction for future research that may bring neurobehavioral correlates of disinhibitory psychopathology even closer to clinical practice (e.g., Patrick et al., 2019). The multimethod approach requires large sample sizes, however, and given the challenges inherent in recruiting participants in offender populations, especially if neuroimaging methods are involved, multicenter studies and broad collaborative efforts may be required. In light of this, it is important to acknowledge that small sample, bivariate mapping studies — such as *Studies II-IV* — are of critical importance for continued validation of novel frameworks, and for unravelling potential neurobehavioral correlates that may be further explored in subsequent research (Perkins, Latzman, et al., 2020).

Figure 5.1: Visualizing the decrease in precision among measures associated with disinhibitory psychpathology.

On a more theoretical level, an important consideration is related to the precision with which individual variables capture the construct of disinhibitory psychopathology. Presumably, the further away a variable is from the core, target construct (i.e., a general propensity towards impulse control problems), the less valid and/or reliable it is (Figure

5.1). Thus, it is important to acknowledge that in *Study II*, for instance, recidivism was likely the outcome of several, interacting factors, and not solely attributable to disinhibitory psychopathology. This limitation may be ameliorated in future research by adopting the multimethod approach described on the previous page.

5.2.4 Ethical considerations

First and foremost, research on mentally disordered offenders presents a challenging ethical dilemma. On the one hand, research conducted on this population may be beneficial in terms of increased understanding of their mental disorders and level of functioning, and may result in new methods for treatment and intervention. On the other hand, their psychiatric status and the legal framework that stipulates what they can and cannot do may impede with their ability to provide informed consent (Munthe et al., 2010). For *Study III*, these issues were carefully discussed and taken into consideration when designing inclusion criteria and when preparing consent forms as well as verbal and written information about the study.

The ethical challenges were even more pronounced in *Study II*. First, participants underwent SPECT imaging using a radioactive compound with an effective dose equivalent of approximately 10 mSv (Huda & Sandison, 1989). Although the administered dose was well within clinical guidelines, it corresponded to approximately five to ten times the annual dose received from background radiation in Sweden. Second, the SPECT procedure was guided by clinical incentives at the time, and thus was not based on informed consent for research purposes. Third, during data collection for follow-up studies, active consent was not deemed necessary due to the inherent difficulties in contacting the participants, and since it was considered that contact could pose a risk to vulnerable individuals with mental health and/or legal problems. Nevertheless, as the data was already collected, it could be argued that it would be wasteful and ethically inadmissible to *not* utilize the data for research purposes, if it was believed that such research could have a beneficial impact on the treatment of mentally disordered offenders. Indeed, *Study II* did generate important knowledge that may form the basis for further studies, but research should always, to the extent possible, be based on informed consent.

More generally, research using neurobiological data to predict behavior (*Study II*) or to classify or characterize individuals (*Studies III* and *IV*) could

be accused of bolstering stigmatization of individuals based on brain alterations or of advocating reductionism and "biologization" (Jurjako et al., 2020). Acknowledging the non-deterministic nature of neurobiological findings (Jurjako et al., 2020) as well incorporating findings of environmental influce on antisocial behavior (such as those discussed in Section 5.2.2) is therefore important. It must also be emphasized that neurobiological data should not at this stage (and perhaps never) be used in the courtroom to, for instance, determine the type of sanction. Relatedly, the notion of research across different 'levels of analysis', as discussed in Section 5.2.1, has been critizised on the grounds of the fuzziness and ill-defined nature of such levels (Miller, 2010). Nevertheless, the intrinsic complexity of psychopathology — especially its neurobiological aspects — means that a levels-approach can still be heuristically useful (Eronen, 2019).

5.3 Summary and conclusions

In sum, this thesis has demonstrated a relatively high prevalence of disinhibitory psychopathology among both offenders and young adults of the general population. Due to its association with adverse outcomes, including persistent offending and recidivism, efforts to identify signs of disinhibitory psychopathology at an early stage are recommended. Disinhibitory psychopathology was associated with an impulsive approach to planning and problem-solving as well as with a reduced capacity for inhibitory control, both of which may result in difficulties in successfully adapting to and interacting with a rapidly changing environment. Still, this thesis also highlights a need for further evaluating the association between disinhibitory psychopathology and inhibitory control. Associations with altered brain structure and function in regions involved monitoring and evaluation of behavior, decision-making, and reward evaluation were also demonstrated. These findings are important for further refinement of novel nosological frameworks, and encourage research into targeted interventions guided by neurobiological findings.

To conclude, this thesis confirms the importance of disinhibitory psychopathology as a clinical construct. It adds to a scarce literature, especially on mentally disordered offenders, and provides much needed evidence of specific neurobehavioral correlates of disinhibitory psychopathology.

A. The Swedish criminal justice system

The Swedish criminal justice system is relatively unique in the sense that offenders who are *non compos mentis* — that is, considered unaccountable when committing a crime — can be held legally responsible. This includes mentally disordered offenders, who are considered capable of having criminal intent and thus eligible for prosecution and sentencing. Depending on if the court judges the prosecuted individual to suffer from a 'severe mental disorder' (SMD; a medicolegal concept explained further below), it decides on sanction, such as prison or forensic psychiatric care (for a thorough overview, see Svennerlind et al., 2010).

Around 1 800 individuals were under forensic psychiatric care in 2019. Over 80% were male, the median age was 40 years, the most common index crime was assault, most had a schizophrenia spectrum diagnosis, and over 90% had received some form of psychiatric treatment before being sentenced to forensic psychiatric care (RättspsyK, 2020). In contrast, around 4 400 individuals served a prison sentence in Sweden in 2019, with an additional 11 000 on probation. The vast majority of prisoners (94%) were males between 25 and 29 years old, and most had committed a violent crime (Kriminalvården, 2020).

The concept of 'severe mental disorder'

The medicolegal concept of SMD, introduced in 1992, is a cornerstone of Swedish forensic psychiatry. Unfortunately, the concept is not explicitly defined, but rather explained through a list of diagnoses that may constitute a SMD. First and foremost on that list are psychotic disorders, or any disorder with a markedly distorted view of reality, such as severe dementia. Other examples include severe depression with suicidal thoughts, and severe personality disorders (Proposition (Government Legislative Bill) 1990/91:58, 1990, p. 86). In practice, however, personality disorders are generally not considered SMDs, although this has varied over the years (Svennerlind et al., 2010).

While the presence of a SMD may result in forensic psychiatric care, there are cases when the degree, severity, or kind of mental disorder

does not qualify for SMD, and the court may decide on a prison sentence instead. In general, if there is no apparent need for forensic psychiatric care, then the sanction is typically a prison sentence. Still, a report from 2014 showed that almost half (44%) of all prisoners had a psychiatric diagnosis, and furthermore that 70% had substance use problems (Kriminalvården, 2014a). Although some prisons have special sections devoted to prisoners with mental health problems, prisoners may also be transferred to a forensic psychiatric hospital, either voluntarily or by recommendation from the prison's treating physician, for a period of psychiatric inpatient care.

Pre-conviction mental health investigations

Three kinds of investigations are used to aid the courts in deciding whether a prosecuted individual should be sentenced to serve time in prison, or be handed over to forensic psychiatric care:

1. A *pre-sentence personal case study*, where an investigator from the local probation authority, a branch of the Swedish Prison and Probation Services, gathers information about the individual's lifestyle and social circumstances. Based on that information, the investigator may recommend further forensic psychiatric investigation.

2. A so-called *section-seven investigation*, where a court-appointed psychiatrist licenced by the National Board of Forensic Medicine reviews the opinion of the local probation authority, police reports, and medical files, and conducts an interview with the individual. The section-seven investigation, sometimes referred to as a *minor forensic psychiatric investigation*, may or may not recommend a major forensic psychiatric investigation (FPI).

3. A *major forensic psychiatric investigation*, which takes around four weeks to complete, is carried out by a team consisting of a specialist in forensic psychiatry, a psychologist, a forensic social investigator, and ward staff. Omitting the finer details, the primary objectives of an FPI are to (1) investigate the presence of a SMD, (2) investigate whether the individual is in need of forensic psychiatric care, and (3) to conduct a risk assessment in order to evaluate whether the presence of a SMD consitutes a risk factor for relapse into serious criminality.

The majority of subjects are examined on remand and are therefore inpatients at the investigative unit. The recommendations following a FPI are almost universally followed by the court, but a complementary statement may be requested from the Committee for Forensic Psychiatry, Social and Medical Legal Questions within the National Board of Health and Welfare, which happens in around 5% of FPIs (Svennerlind et al., 2010).

A total of 1 242 section-seven investigations as well as 573 FPIs — the highest number in fifteen years — were conducted in 2019. Around 60% of those that underwent a FPI were considered to suffer from a SMD (Rättsmedicinalverket, 2020). When the court decides to sentence an individual to forensic psychiatric care, they may also include a condition called 'special court supervision', based on the risk assessment carried out during the FPI. Special court supervision, which is included in approximately 80% of sentences for females and 90% of sentences for males (RättspsyK, 2020), means that any changes in privileges, outpatient care, and discharge must be approved by an administrative court. The administrative court hearing is held every six months, and consults the prosecutor from the initial trial as well as an independent expert on psychiatry.

B. Frequentist vs. Bayesian inference

Several fields, including psychology, medicine, and biology, have been accused of delving into a "mindless statistical ritual" that, according to Gigerenzer (2004, p. 588), consists of three steps:

1. Set up a statistical null hypothesis of "no difference" or "zero correlation".
2. Use 5% as a convention for rejecting the null hypothesis (and if significant, accepting the research hypothesis while reporting results as $p < 0.05$, $p < 0.01$, or $p < 0.001$, whichever comes closest).
3. Always perform this procedure.

Although often left out of statistics textbooks, this procedure, commonly referred to as null hypothesis significance testing (NHST), is in fact an inconsistent hybrid of the work by British statistician Sir Ronald Fisher (1890-1962), Polish mathematician Jerzy Neyman (1894-1981), and British statistician Egon Pearson (1895-1980).

In Fisher's original theory of null hypothesis testing, there were no such things as 'statistical power', 'confidence interval', 'effect size', or a 'Type II error'; those were borrowed from Neyman-Pearson's decision theory. Furthermore, within the Neyman-Pearson framework, terms such as 'highly significant' or 'marginally significant' are meaningless, since hypotheses are either accepted or rejected (Dienes, 2008; Gigerenzer, 2004)[1]. In their quest for a "one size fits all" statistics toolbox, however — and without approval of either Fisher, Neyman, or Pearson — textbook writers soon began fusing their theories together, and NHST became institutionalized within psychology research around the mid 1950s, from where it spread to other disciplines (Gigerenzer, 2018, 2004).

As an example of how firmly rooted the NHST ritual is, consider the use of $p < 0.05$ as a threshold for 'statistical significance', which remains a cornerstone of much empirical research even today (*Studies I*, *II*, and *IV* in this thesis are no exceptions). Although Fisher was not first in his use of a *p*-value — that award goes to John Arbuthnot in 1710 — he was among

[1] Neyman also opposed to the term 'significance', instead preferring 'size of the test' (Dienes, 2008, p. 61).

the first to formalize it, along with the accompanying $p < 0.05$ convention, in the mid 1920s (Kennedy-Shaffer, 2019). Fisher was not too enthusiastic, however; he suggested $p < 0.05$ merely as a matter of convenience (Fisher, 1950, p. 44). Furthermore, the choice of $p < 0.05$ (and of $p < 0.01$) was just a mathematical coincidence. It just so happens that it is easy to calculate the 95% or 99% probabilities for any parameter approximated or modeled by a normal distribution. Thus, before the dawn of computers and pocket calculators, these thresholds were used as a means to avoid weeks of manual calculation (Hacking, 2001, p. 217). Still, in the end, even Fisher himself dismissed the idea of a conventional level of significance, calling it "absurdely academic" (Fisher, 1956, p. 42), further stating that:

> No scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of evidence and his ideas.

Over 60 years have passed since Fisher recoiled on his idea of a fixed level of signifiance, yet *p*-values, confidence intervals, and associated terms such as 'statistical significance' continue to be misunderstood, misinterpreted, and misused (McShane & Gal, 2017; Wasserstein & Lazar, 2016), even in top journals such as *Science* and *Nature* (Nieuwenhuis et al., 2011), and even by those assigned to teach them (Haller & Krauss, 2002).

In recent years, the issues inherent in NHST have been increasingly scrutinized, particularly in light of the so-called 'replication crisis' (Ioannidis, 2005). Researchers are now urged to "avoid using statistical significance or p values; simply omit any mention of null-hypothesis significance testing", and to "move beyond NHST and use the most appropriate methods, whether estimation or other approaches" (Cumming, 2014, p. 8). With the rising popularity of statistical programming languages such as R and Stan, the Bayesian approach to statistical modeling and parameter estimation is rapidly gaining foothold as a viable alternative to NHST (for thorough introductions, see Kruschke, 2015; McElreath, 2020). The following sections will briefly outline a few common issues with the NHST approach and show how a Bayesian approach may be more intuitive.

What does it mean when $p < 0.05$?

Imagine that we want to compare NoGo P3 amplitude between two groups: mentally disordered offenders and healthy controls. We collect some data, run a t-test, and the result is that mentally disordered offenders (M = 3.2 μV) have a lower amplitude than controls (M = 4.9 μV), $p = 0.02$. Assuming a threshold for statistical significance of $p < 0.05$, which while completely arbitrary is the *de facto* standard, how can we interpret this result? Could we simply state that "mentally disordered offenders have a lower NoGo P3 amplitude than controls, $p = 0.02$"? Not quite. The statement is correct for this sample, and perhaps even for the population, but it has nothing to do with the *p*-value. Instead, an accurate statement would be along the lines of:

> We observed evidence against the null hypothesis of no difference in NoGo P3 amplitude ($p = 0.02$), and the observed difference was in favour of a lower NoGo P3 amplitude in mentally disordered offenders compared to controls. If this study would be *exactly* replicated *indefinitely*, in 2% of such replications a result equal to or more extreme than that of the current study would be observed, if the null hypothesis is true.[2]

Thus, *p*-values concern the *long-run relative frequency* of obtaining a result, or one more extreme, if the null hypothesis ($\mathcal{H}_0$) is true, but they do not allow interpreting the probability of the *alternative* hypothesis ($\mathcal{H}_1$). Since it is a useful albeit erroneous heuristic, however, the pseudo-Bayesian interpretation of *p*-values as representing the probability of $\mathcal{H}_1$ is quite common (Wagenmakers et al., 2018). In contrast, the Bayesian approach allows you to quantify the probability of both $\mathcal{H}_0$ and $\mathcal{H}_1$ in light of the data at hand. There is no need for long-run frequencies; one could just state that "according to our model, the probability that mentally disordered offenders have a lower NoGo P3 amplitude than controls is 92%", and that would be correct. Note that there is no inherent need to dichotomize probabilities in the Bayesian framework, for instance suggesting that a probability of 90% or higher means something different than a probability of 89%. Such interpretations are best left to the reader, although in practice, editors and reviewers may request such heuristics.

[2] This and subsequent examples are paraphrased from Frank Harrell at: https://www.fharrell.com/post/bayes-freq-stmts/ and https://discourse.datamethods.org/t/language-for-communicating-frequentist-results-about-treatment-effects/934.

What does it mean when $p > 0.05$?

Using the above example, imagine that the result instead was that mentally disordered offenders (M = 4.3 μV) have a lower amplitude than controls (M = 5.1 μV), $p = 0.21$. Again assuming a threshold for statistical significance of $p < 0.05$, how can we interpret this finding? While findings where $p > 0.05$ are often labelled "negative", it is important to keep in mind that absence of evidence is not evidence of absence (Altman & Bland, 1995). Thus, we cannot simply state that "mentally disordered offenders did not have lower amplitude than controls, $p = 0.21$". A more accurate statement would be along the lines of:

> If mentally disordered offenders had *exactly* the same mean NoGo P3 amplitude as controls, and if this study would be *exactly* replicated *indefinitely*, in 21% of such replications a result equal to or more extreme than that of the current study would be observed.

Again, the Bayesian approach on the other hand allows us make genuine probabilistic statements, such as "according to our model, the probability that mentally disordered offenders have a lower NoGo P3 amplitude than controls is 79%".

What does a confidence interval mean?

Confidence intervals (CIs) are recommended by some as a viable alternative to *p*-values (e.g., Cumming, 2014), but while certainly more nuanced than just reporting a *p*-value, they too are limited in terms of interpretation. Again, using the above example, imagine that we run a t-test and obtain a 95% CI for the difference in means of [0.2, 2.8]. Does this mean that there is a 95% probability that the true mean difference in NoGo P3 amplitude between mentally disordered offenders and controls is between 0.2 and 2.8 μV? No, unfortunately it does not. The specific CI we calculated either does (100%) or does not (0%) contain the true mean difference; we do not know which one it is. Thus, a correct statement would be along the lines of:

> If this study could be *exactly* replicated *indefinitely*, using the same calculation of a confidence interval each time, then 95% of these calculated confidence intervals would contain the *unknown* true difference in mean NoGo P3 amplitude.

The long-run relative frequency assumption of CIs and p-values is easy to visualize using simulation. Figure B.1 shows the result of 50 simulated datasets containing two groups of 25 data points each. For each dataset, a t-test was calculated and the resulting 95% CI saved. Note that 28 out of the 50 simulations were not statistically significant at $p < 0.05$, but the p-value says nothing of the probability of $\mathcal{H}_0$ in a single simulation. Likewise, if we were to continue simulating datasets and calculate CIs for all eternity, then 95% of those would indeed contain the true difference in means, but each single CI either does contain the true difference, or does not.

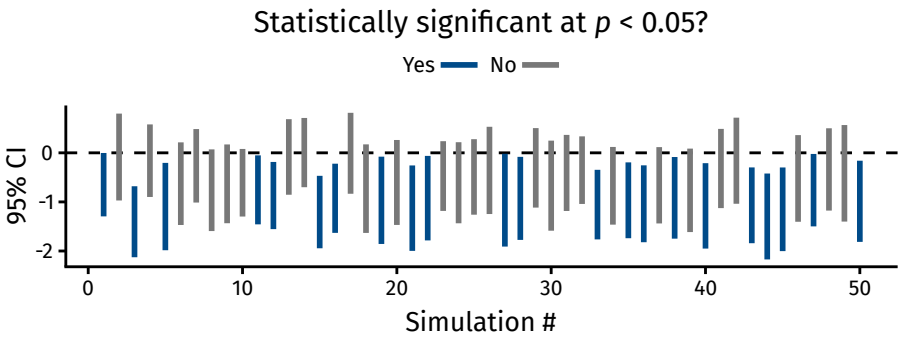

Figure B.1: 50 simulated 95% confidence intervals for a t-test of difference in means.

The Bayesian alternative to CIs, termed credible intervals (CrIs), does, however, summarize the posterior belief about the estimated parameter. In other words, in the Bayesian approach, you could simply state that "according to our model, there is a 95% probability that the difference in NoGo P3 amplitude is between 0.2 and 2.8 μV".

Choosing one approach over the other?

It is important to understand that statistics proper is *not* equal to NHST. Fisher, for instance, was strictly opposed to a universal theory of statistics, and he fought hard against the "cookbook approach" to statistics (Hacking, 2001, p. 226). Rather, it is the frequent misuse, misinterpretation, and "mindless statistical rituals" inherent in the NHST approach that is increasingly being questioned; not the long-run relative frequency interpretation of probability *per se*.

If one does endorse the long-run relative frequency type of probability — and many do — then one must also accept that p-values and confidence intervals are probabilities in the context of this long-run relative

frequency, and that they do not apply to individual events (i.e., individual studies or experiments). Thus, there is nothing wrong with favouring the long-run frequency interpretation, as long as one understands its assumptions and consequences.

On the other hand, as British economist John Maynard Keynes (1883-1946), himself a dogmatic Bayesian, remarked, "in the long run we are all dead" (Hacking, 2001, p. 149). If one instead prefers the subjective, Bayesian interpretation of probability, with all *its* inherent assumptions, then one is awarded with, for instance, a continous degree of posterior probability that is valid an often reasonable even for small sample sizes, that can be updated in light of new data, and that can quantify evidence in favour of both $\mathcal{H}_0$ and $\mathcal{H}_1$ (Depaoli & van de Schoot, 2017; Wagenmakers et al., 2018). Of course, there is ample critique against the Bayesian approach as well, especially its reliance on priors, and some advocate a purely likelihood-based approach without any priors or posteriors at all (Dienes, 2008, p. 123).

In the end, what it all comes down to is deciding which approach makes most sense or provides the best answer to a research question. Both approaches have made, and continue to make, important contributions to statistics; they may even facilitate the continued development of each other (Berger & Bayarri, 2004). As George Box so famously has stated on several occasions, "all models are wrong, but some are useful" (e.g., Box & Draper, 1987, p. 424).

C. Introduction to Bayesian statistical modeling

For readers not familiar with the Bayesian approach to statistical modeling, the following sections will provide a (very) brief, conceptual introduction. Thorough accounts are offered in Gelman et al. (2013), Kruschke (2015), and McElreath (2020).[1]

Bayes' theorem

Bayesian statistical modeling is based on Thomas Bayes' famous theorem, which provides the probability of an event given prior knowledge of conditions that might be related to that event:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$P(A|B)$ is the probability of event A given that event B is true, $P(B|A)$ is the probability of event B given that event A is true, and $P(A)$ and $P(B)$ are the probabilities of observing A and B, respectively. If we use $\mathcal{H}$ to denote the hypothesis and $\mathcal{D}$ to denote the data, the theorem can be stated as:

$$P(\mathcal{H}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{H})P(\mathcal{H})}{P(\mathcal{D})}$$

Thus, the objective of Bayesian modeling is to calculate $P(\mathcal{H}|\mathcal{D})$, the probability of the hypothesis $\mathcal{H}$ given the data $\mathcal{D}$. In contrast, a *p*-value is the inverse, $P(\mathcal{D}|\mathcal{H})$, which is the probability (in the long run) of $\mathcal{D}$ given $\mathcal{H}$, where $\mathcal{H} = \mathcal{H}_0$. When estimating a parameter, $\mathcal{H}$ can be replaced with the parameter of interest, usually denoted θ. For instance, θ could represent the difference in NoGo P3 amplitude between mentally disordered offenders and controls:

[1] A free, online version of Gelman et al. (2013) is available at: http://www.stat.columbia.edu/~gelman/book/BDA3.pdf.

$$\overbrace{P(\theta|\mathcal{D})}^{\text{Posterior}} = \frac{\overbrace{P(\mathcal{D}|\theta)}^{\text{Likelihood}}\overbrace{P(\theta)}^{\text{Prior}}}{\underbrace{P(\mathcal{D})}_{\text{Marginal likelihood}}}$$

The goal is to obtain $P(\theta|\mathcal{D})$, or the ***posterior distribution*** of θ after taking the observed data into account. The ***likelihood***, $P(\mathcal{D}|\theta)$, represents how likely the observed data is, given a distribution of possible values of θ. The ***prior***, $P(\theta)$, specifies our belief about the distribution of θ before taking $\mathcal{D}$ into account. For instance, one can use a subjective prior that incorporates previous research findings, or one could use a "weakly informative" prior that perhaps is centered around zero while also ruling out impossible or improbable values of θ. Finally, $P(\mathcal{D})$ represents the ***marginal likelihood***, which is the likelihood of $\mathcal{D}$ given $P(\theta)$ evaluated at every possible value of θ:

$$P(\mathcal{D}) = \int P(\mathcal{D}|\theta) \times P(\theta) d\theta$$

The marginal likelihood is used to normalize the posterior to achieve a probability density (i.e., ranging from zero to one).

Bayesian priors

Priors are crucial to Bayesian inference, and the subjectiveness inherent in chosing an appropriate prior is often a point of objection among those who oppose the Bayesian approach (Gelman, 2008). To understand how priors work, imagine that we want to estimate the difference in NoGo P3 latency between mentally disordered offenders and a group of healthy controls. As shown in Figure C.1, our ***prior*** belief is represented by a normal probability distribution centered around zero. This prior could be considered skeptical, or weakly informative, since it suggests that the difference is most likely zero, but that values ranging all the way between around -30 ms and 30 ms are possible, albeit unprobable. A more subjective prior could be used, based on previous research, but in pratice it is best to remain skeptical.

The next step is to collect some data. As shown in Figure C.1, the data — modeled using a likelihood function — suggests a difference distributed

around a mean of approximately 10 ms, with possible values ranging from around -5 ms to 25 ms. Using Bayes' theorem, we update our *posterior* belief by multiplying the prior with the likelihood, and the result is represented by a posterior distribution centered around 7 ms, with possible values ranging from around -10 ms to 2 ms. Using the full posterior distribution, we can calculate, for instance, the *probability* that the difference is above or below zero, above or below a certain threshold, or within a certain range.

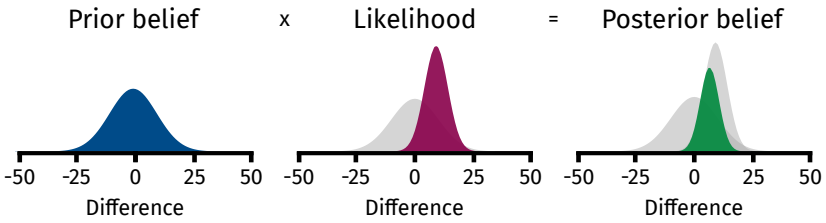

Figure C.1: In the Bayesian framework, prior beliefs are updated, using the likelihood, in light of new data.

Markov Chain Monte Carlo

Bayes' theorem is easily solved by hand for discrete data, but for more complex models and continous data, the marginal likelihood and thus the normalized posterior distribution is not directly computable. There are several numerical techniques that can solve this problem, either by approximating or drawing samples from the posterior distribution, but MCMC is probably the most common today.

The MCMC algorithm dates back to the development of the first atomic and hydrogen bombs at Los Alamos in the 1940s and 50s, where Stanislaw Ulam[2], John von Neumann, and others worked on thermodynamics and nuclear fission. The world's first computer, ENIAC (short for Electronic Numerical Integrator and Computer), was built at Los Alamos in 1946 and was used for regular Monte Carlo calculations by John von Neuman. The MCMC algorithm was first implemented on its successor, the MANIAC[3], built at Los Alamos in the mid 1950s (Robert & Casella, 2011).

[2]The Stan programming language is named in his honour.

[3]The MANIAC was built by a group at Los Alamos led by Nicholas Metropolis, who also came up with the name Monte Carlo, inspired by the casino in Monaco. He named the computer MANIAC, short for Mathematical Analyzer Numerical Integrator and Computer, hoping to put an end to the use of such ridiculous acronyms, with questionable success (Metropolis, 1987).

The key idea behind using MCMC for Bayesian statisical modeling is that the algorithm makes it possible to draw samples from a distribution knowing just how to calculate its likelihood (van Ravenzwaaij et al., 2018). As its history suggests, however, MCMC is computationally intensive, and the resurgence of Bayesian statistics in the 1990s owes a great deal to the development of faster computers (McElreath, 2020, p. 45).

Reporting Bayesian results

The outcome of a Bayesian model is the posterior distribution, but in some cases is may not be feasible to report the entire posterior distribution, or some other kind of Bayesian inference may be preferred.

BFs, used in *Study II*, are perhaps the most well-known form of Bayesian inference, and may be especially appealing to those who convert from a frequentist to a Bayesian approach. In essence, BFs indicate how credible one hypothesis is in relation to another, with a key advantage being that they allow for testing null hypotheses within a Bayesian framework (Williams et al., 2017). One can test other hypotheses as well, but then it is perhaps better to just focus on the posterior distribution.

The P_D, used in *Studies III* and *IV*, ranges between 50% and 100% and represents the probability that a parameter is different from zero (in either a positive or negative direction). Thus, the P_D is reminiscent of the original, Fisherian idea of a *p*-value as an index of the *existence* of an effect, rather than the *significance* (in the literal sense of the word) of an effect (Makowski et al., 2019). Although somewhat contrary to the purpose of a Bayesian approach, it is possible to derive a number corresponding to the frequentist *p*-value from the P_D, which may be helpful for readers who are not familiar with Bayesian inference.

CrIs are used in a similar fashion to frequentist CIs (but mind the difference in interpretation), and may, like CIs, be calulated in different ways. For instance, the posterior distribution can be summarized using a quantile interval (QI; used in *Study IV*), also known as the percentile or equal-tailed interval. Another option is the HDI (used in *Study III*), also known as the highest posterior density interval, which summarizes the posterior distribution such that all values within the HDI have a higher probability density than all values outside the HDI. For more or less symmetric posterior distributions, the difference between a QI and a HDI is hardly noticable. The width of the CrI varies according to the pref-

erence of the researcher who presents them. For instance, McElreath (2020) uses 67%, 89%, and 97% intervals — all prime numbers — to remind readers that conventions such as 95% are completely arbitrary, and that being prime is no worse a justification than reporting a 95% interval just because Fisher remarked that it is convenient almost 100 years ago. Nonetheless, many are not aware of its arbitrariness, or may not find it a problem, which can cause difficulties, for instance, during the publication process. To circumwent this, *Studies III* and *IV* borrowed examples of how to describe probabilities in everyday language from the Intergovernmental Panel on Climate Change (Mastrandrea et al., 2011).

D. Bayesian models used in Study III and IV

In *Study III*, differences between controls and mentally disordered offenders were modeled using a robust linear regression approach, allowing unequal variance between the two groups, with a Student's t distribution for the response (or dependent) variable[1] (Lange et al., 1989). The Student's t distribution has a so-called 'normality' parameter, ν, that when set to large values actually approaches a normal distribution. When ν is small, however, the Student's t distribution has heavier tails, and is therefore more robust to outliers (see Figure D.1).

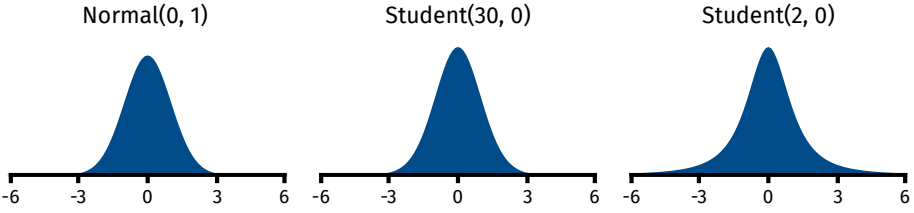

Figure D.1: A comparison between the standard normal distribution and two Student's t distributions with varying normality parameters. Notice the similarity between the Normal(0, 1) and Student(30, 0) distributions, as well as the wider tails of the Student(2, 0) distribution.

The distribution of a response variable (e.g., NoGo P3 amplitude) can therefore be described as:

$$y_{ik} \sim T(\nu, \mu_{ik}, \sigma_{ik})$$

where y (i.e., the response variable) is a random draw from a t distribution with normality parameter ν, mean μ_{ik}, and standard deviation σ_{ik}.[2] The i index represents each row (i.e., participant) in the data and the k index represents the dummy coded group (e.g., 0 or 1 for control or mentally disordered offender, respectively). Both μ and σ are modeled as linear regressions:

[1] In other words, this is the likelihood function used to estimate $P(\mathcal{D}|\theta)$.

[2] Technically, μ and σ are called location and scale, respectively, in the context of a t distribution.

$$\mu_{ik} = \beta_0 + \beta_1 x_{ik}$$
$$\sigma_{ik} = \gamma_0 + \gamma_1 x_{ik}$$

where β_0 represents the control group's mean, β_1 represents the effect (or difference) of being a mentally disordered offender, and x represents the dummy coded group variable. Thus, when x is zero, β_1 is also zero, and so μ is the mean of the control group. When x is one, μ is β_0 plus β_1 (the group difference), which gives the mean of the mentally disordered offenders. The standard deviation is modeled in a similar fashion, with γ_0 and γ_1 representing the control group's standard deviation and the difference in standard deviation between groups, respectively.[3]

Study IV used an identical approach, only with additional variables, to model the association between cortical thickness and different expressions of disinhibitory psychopathology, while controlling for age, gender, and years of education. In *Study IV*, the outcome of interest was not a difference between groups but the association between two variables estimated by the standardized beta coefficient (β).

When regressing a multivariate model with two response variables on zero (i.e., without predictors), the residual correlation estimated by brms is actually the correlation between the two variables. Thus, correlations between two variables x_1 and x_2 were modeled, in both *Study III* and *Study IV*, using a linear regression approach with a multivariate Student's t distribution:

$$\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \sim MVT(\nu, \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix}, \Sigma)$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \Omega \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$$

where Σ is the covariance matrix and Ω is the correlation matrix.

Both *Study III* and *Study IV* used a Normal(0, 10) prior for μ, allowing the occasional extreme value, but with the majority of mass centered around zero. A Gamma(2, 0.1) prior was used for the ν parameter, which pushes

[3] In practice, one can instruct brms to suppress the intercept, so that the means of the two groups are modeled simultaneously.

ν towards low values — and thus increasing robustness — while still allowing higher values due to the sloping right tail. A Cauchy(0, 1) prior was used for the standard deviation σ, with most of its mass centered around zero, but with wide tails allowing more extreme values. Finally, an LKJ(2) prior (Lewandowski et al., 2009) was used for the correlation matrices. A graphical depiction of these priors is presented in Figure D.2.

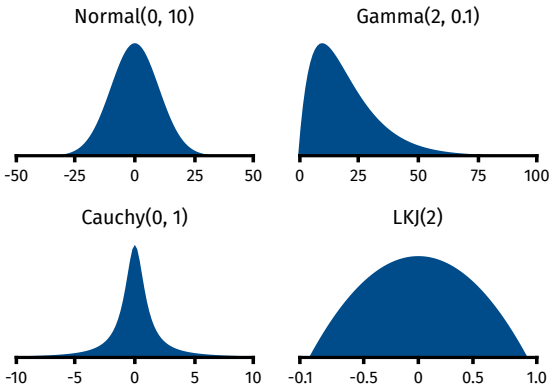

Figure D.2: Graphical depiction of the Bayesian priors used in Study III and IV.

References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs: General and Applied*, *80*(7), 1–37.

Adams, H. E., Luscher, K. A., & Bernat, J. A. (2002). The classification of abnormal behavior: An overview. In P. B. Sutker & H. E. Adams (Eds.), *Comprehensive handbook of psychopathology* (Third, pp. 3–28). Kluwer Academic Publishers.

Aharoni, E., Mallett, J., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W., Gazzaniga, M. S., & Kiehl, K. A. (2014). Predictive accuracy in the neuroprediction of rearrest. *Social Neuroscience*, *9*(4), 332–336. https://doi.org/10.1080/17470919.2014.907201

Aharoni, E., Vincent, G. M., Harenski, C. L., Calhoun, V. D., Sinnott-Armstrong, W., Gazzaniga, M. S., & Kiehl, K. A. (2013). Neuroprediction of future rearrest. *Proceedings of the National Academy of Sciences*, *110*(15), 6223–6228. https://doi.org/10.1073/pnas.1219302110

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, *17*(5), e3000246. https://doi.org/10.1371/journal.pbio.3000246

Allom, V., Mullan, B., & Hagger, M. (2016). Does inhibitory control training improve health behaviour? A meta-analysis. *Health Psychology Review*, *10*(2), 168–186. https://doi.org/10.1080/17437199.2015.1051078

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ*, *311*(7003), 485. https://doi.org/10.1136/bmj.311.7003.485

Ameis, S. H., Ducharme, S., Albaugh, M. D., Hudziak, J. J., Botteron, K. N., Lepage, C., Zhao, L., Khundrakpam, B., Collins, D. L., Lerch, J. P., Wheeler, A., Schachar, R., Evans, A. C., & Karama, S. (2014). Cortical thickness, cortico-amygdalar networks, and externalizing behaviors in healthy children. *Biological Psychiatry*, *75*(1), 65–72. https://doi.org/10.1016/j.biopsych.2013.06.008

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*. American Psychiatric Publishing, Inc.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Publishing, Inc.

Anderson, N. B. (1998). Levels of analysis in health science: A framework for integrating sociobehavioral and biomedical research. *Annals of the*

New York Academy of Sciences, *840*, 563–576. https://doi.org/10.1111/j.1749-6632.1998.tb09595.x

Andreasson, H., Nyman, M., Krona, H., Meyer, L., Anckarsäter, H., Nilsson, T., & Hofvander, B. (2014). Predictors of length of stay in forensic psychiatry: The influence of perceived risk of violence. *International Journal of Law and Psychiatry*, *37*(6), 635–642. https://doi.org/10.1016/j.ijlp.2014.02.038

Apps, M. A. J., Lockwood, P. L., & Balsters, J. H. (2013). The role of the midcingulate cortex in monitoring others' decisions. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00251

Aron, A. R. (2011). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biological Psychiatry*, *69*(12), e55–e68. https://doi.org/10.1016/j.biopsych.2010.07.024

Ashburner, J., & Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage*, *55*(3), 954–967. https://doi.org/10.1016/j.neuroimage.2010.12.049

Aversi-Ferreira, T. A., Araújo, M. F. P. de, Lopes, D. B., & Nishijo, H. (2010). History, citoarchitecture and neurophysiology of human and non human primates' parietal lobe: A review. *Dementia & Neuropsychologia*, *4*(3), 173–180. https://doi.org/10.1590/S1980-57642010DN40300005

Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Filho, W. J., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, *513*(5), 532–541. https://doi.org/10.1002/cne.21974

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, *49*(5), 1195–1205. https://doi.org/10.1016/j.cortex.2012.05.022

Barker, E. D., Séguin, J. R., White, H. R., Bates, M. E., Lacourse, É., Carbonneau, R., & Tremblay, R. E. (2007). Developmental trajectories of male physical violence and theft: Relations to neurocognitive performance. *Archives of General Psychiatry*, *64*(5), 592. https://doi.org/10.1001/archpsyc.64.5.592

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. The Guilford Press.

Barnard, G. A. (1945). A new test for 2 x 2 tables. *Nature*, *156*(3954), 177–177. https://doi.org/10.1038/156177a0

Batista-García-Ramó, K., & Fernández-Verdecia, C. I. (2018). What we know about the brain structure-function relationship. *Behavioral Sciences*, *8*(4). https://doi.org/10.3390/bs8040039

Beauchaine, T. P., Zisner, A. R., & Sauder, C. L. (2017). Trait impulsivity and the externalizing spectrum. *Annual Review of Clinical Psychology*, *13*(1), 343–368. https://doi.org/10.1146/annurev-clinpsy-021815-093253

Beauchamp, M. H., Dagher, A., Aston, J. A. D., & Doyon, J. (2003). Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. *NeuroImage*, *20*(3), 1649–1660. https://doi.org/10.1016/j.neuroimage.2003.07.003

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, *10*(3), 295–307. https://doi.org/10.1093/cercor/10.3.295

Beeli, G., Casutt, G., Baumgartner, T., & Jäncke, L. (2008). Modulating presence and impulsiveness by external stimulation of the brain. *Behavioral and Brain Functions*, *4*(1), 33. https://doi.org/10.1186/1744-9081-4-33

Berger, J. O., & Bayarri, M. J. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, *19*(1), 58–80. https://doi.org/10.1214/088342304000000116

Berlin, H. A., Rolls, E. T., & Kischka, U. (2004). Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. *Brain*, *127*(5), 1108–1126. https://doi.org/10.1093/brain/awh135

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Billstedt, E., Anckarsäter, H., Wallinius, M., & Hofvander, B. (2017). Neurodevelopmental disorders in young violent offenders: Overlap and background characteristics. *Psychiatry Research*, *252*, 234–241. https://doi.org/10.1016/j.psychres.2017.03.004

Blackburn, R. (2007). Personality disorder and antisocial deviance: Comments on the debate on the structure of the psychopathy checklist-revised. *Journal of Personality Disorders*, *21*(2), 142–159. https://doi.org/10.1521/pedi.2007.21.2.142

Blais, J., Solodukhin, E., & Forth, A. E. (2014). A meta-analysis exploring the relationship between psychopathy and instrumental versus reactive violence. *Criminal Justice and Behavior*, *41*(7), 797–821. https://doi.org/10.1177/0093854813519629

Bonta, J., Blais, J., & Wilson, H. A. (2014). A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior*, *19*(3), 278–287. https://doi.org/10.1016/j.avb.2014.04.014

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. Wiley.

Brazil, I. A., van Dongen, J. D. M., Maes, J. H. R., Mars, R. B., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews*, *91*, 259–277. https://doi.org/10.1016/j.neubiorev.2016.10.010

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brower, M. C., & Price, B. H. (2001). Neuropsychiatry of frontal lobe dysfunction in violent and criminal behaviour: A critical review. *Journal of Neurology, Neurosurgery & Psychiatry*, *71*(6), 720–726.

Brutti, P., De Santis, F., & Gubbiotti, S. (2014). Bayesian-frequentist sample size determination: A game of two priors. *METRON*, *72*(2), 133–151. https://doi.org/10.1007/s40300-014-0043-2

Burt, S. A., Pearson, A. L., Carroll, S., Klump, K. L., & Neiderhiser, J. M. (2020). Child antisocial behavior is more environmental in origin in disadvantaged neighborhoods: Evidence across residents' perceptions and geographic scales in two samples. *Journal of Abnormal Child Psychology*, *48*(2), 265–276. https://doi.org/10.1007/s10802-019-00587-6

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Byrne, K. A., Patrick, C. J., & Worthy, D. A. (2016). Striatal dopamine, externalizing proneness, and substance abuse: Effects on wanting and learning during reward-based decision making. *Clinical Psychological Science*, *4*(5), 760–774. https://doi.org/10.1177/2167702615618163

Bø, R., Aker, M., Billieux, J., & Landrø, N. I. (2016). Binge drinkers are fast, able to stop - but they fail to adjust. *Journal of the International Neuropsychological Society: JINS*, *22*(1), 38–46. https://doi.org/10.1017/S1355617715001204

Campanella, S., Schroder, E., Monnart, A., Vanderhasselt, M.-A., Duprat, R., Rabijns, M., Kornreich, C., Verbanck, P., & Baeken, C. (2017). Transcranial direct current stimulation over the right frontal inferior cortex decreases neural activity needed to achieve inhibition: A double-blind ERP study in a male population. *Clinical EEG and Neuroscience*, *48*(3), 176–188. https://doi.org/10.1177/1550059416645977

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Casillas, A., & Clark, L. A. (2002). Dependency, impulsivity and self-harm: Traits hypothesized to underlie the association between Cluster B personality and substance use disorders. *Journal of Personality Disorders*, *16*(5), 424–436. https://doi.org/10.1521/pedi.16.5.424.22124

Cecil, K. M., Brubaker, C. J., Adler, C. M., Dietrich, K. N., Altaye, M., Egelhoff, J. C., Wessel, S., Elangovan, I., Hornung, R., Jarvis, K., & Lanphear, B. P. (2008). Decreased brain volume in adults with childhood lead exposure. *PLoS Medicine*, *5*(5), e112. https://doi.org/10.1371/journal.pmed.0050112

Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, *18*(2), 72–145. https://doi.org/10.1177/1529100617727266

Cleckley, H. M. (1988). *The mask of sanity: An attempt to clarify some issues about the so-called psychopathic personality* (Fifth edition). Emily S. Cleckley.

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and orbitofrontal reactivity to social threat in individuals with impulsive aggression. *Biological Psychiatry*, *62*(2), 168–178. https://doi.org/10.1016/j.biopsych.2006.08.024

Coid, J., Yang, M., Ullrich, S., Roberts, A., Moran, P., Bebbington, P., Brugha, T., Jenkins, R., Farrell, M., Lewis, G., Singleton, N., & Hare, R. (2009). Psychopathy among prisoners in England and Wales. *International Journal of Law and Psychiatry*, *32*(3), 134–141. https://doi.org/10.1016/j.ijlp.2009.02.008

Conway, C. C., & Simms, L. J. (2020). Maximizing the applied value of structural models of psychopathology: Introduction to a special issue of Personality and Mental Health. *Personality and Mental Health*, *14*(1), 3–8. https://doi.org/10.1002/pmh.1474

Cooke, D. J., Hart, S. D., Logan, C., & Michie, C. (2012). Explicating the construct of psychopathy: Development and validation of a conceptual model, the Comprehensive Assessment of Psychopathic Personality (CAPP). *International Journal of Forensic Mental Health*, *11*(4), 242–252. https://doi.org/10.1080/14999013.2012.746759

Courtney, S. M. (2004). Attention and cognitive control as emergent properties of information representation in working memory. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(4), 501–516. https://doi.org/10.3758/CABN.4.4.501

Craft, M. J. (Ed.). (1966). *Psychopathic disorders and their assessment*. Pergamon Press.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Cunha-Bang, S. da, Fisher, P. M., Hjordt, L. V., Holst, K., & Knudsen, G. M. (2019). Amygdala reactivity to fearful faces correlates positively with impulsive aggression. *Social Neuroscience*, *14*(2), 162–172. https://doi.org/10.1080/17470919.2017.1421262

Dahnke, R., Yotter, R. A., & Gaser, C. (2013). Cortical thickness and central surface estimation. *NeuroImage*, *65*, 336–348. https://doi.org/10.1016/j.neuroimage.2012.09.050

De Brito, S. A., Mechelli, A., Wilke, M., Laurens, K. R., Jones, A. P., Barker, G. J., Hodgins, S., & Viding, E. (2009). Size matters: Increased grey matter in boys with conduct problems and callousUnemotional traits. *Brain*, *132*(4),

843–852. https://doi.org/10.1093/brain/awp011

de Carvalho, H. W., Andreoli, S. B., Vaidyanathan, U., Patrick, C. J., Quintana, M. I., & Jorge, M. R. (2013). The structure of common mental disorders in incarcerated offenders. *Comprehensive Psychiatry*, *54*(2), 111–116. https://doi.org/10.1016/j.comppsych.2012.07.006

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's *t*-test instead of Student's *t*-test. *International Review of Social Psychology*, *30*(1), 92. https://doi.org/10.5334/irsp.82

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*(2), 240–261. https://doi.org/10.1037/met0000065

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

de Vet, H. C. W., Mokkink, L. B., Mosmuller, D. G., & Terwee, C. B. (2017). Spearman-Brown prophecy formula and Cronbach's alpha: Different faces of reliability and opportunities for new applications. *Journal of Clinical Epidemiology*, *85*, 45–49. https://doi.org/10.1016/j.jclinepi.2017.01.013

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48. https://doi.org/10.1016/j.dcn.2015.11.005

Dienes, Z. (2008). *Understanding psychology as a science:* Palgrave Macmillan.

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. https://doi.org/10.1016/j.dcn.2017.02.008

Drachman, D. A. (2005). Do we have brain to spare? *Neurology*, *64*(12), 2004–2005. https://doi.org/10.1212/01.WNL.0000166914.38327.BB

Drislane, L. E., Patrick, C. J., & Arsal, G. (2014). Clarifying the content coverage of differing psychopathy inventories through reference to the Triarchic Psychopathy Measure. *Psychological Assessment*, *26*(2), 350–362. https://doi.org/10.1037/a0035152

Durazzo, T. C., Tosun, D., Buckley, S., Gazdzinski, S., Mon, A., Fryer, S. L., & Meyerhoff, D. J. (2011). Cortical thickness, surface area, and volume of the brain reward system in alcohol dependence: Relationships to relapse and extended abstinence. *Alcoholism: Clinical and Experimental Research*, *35*(6),

1187–1200. https://doi.org/10.1111/j.1530-0277.2011.01452.x

Ebitz, R. B., & Hayden, B. Y. (2016). Dorsal anterior cingulate: A Rorschach test for cognitive neuroscience. *Nature Neuroscience*, *19*(10), 1278–1279. https://doi.org/10.1038/nn.4387

Elliott, R. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex*, *10*(3), 308–317. https://doi.org/10.1093/cercor/10.3.308

Eronen, M. I. (2019). The levels problem in psychopathology. *Psychological Medicine*, 1–7. https://doi.org/10.1017/S0033291719002514

Ersche, K. D., Williams, G. B., Robbins, T. W., & Bullmore, E. T. (2013). Meta-analysis of structural brain abnormalities associated with stimulant drug dependence and neuroimaging of addiction vulnerability and resilience. *Current Opinion in Neurobiology*, *23*(4), 615–624. https://doi.org/10.1016/j.conb.2013.02.017

Esteller, À., Poy, R., & Moltó, J. (2016). Deficient aversive-potentiated startle and the triarchic model of psychopathy: The role of boldness. *Biological Psychology*, *117*, 131–140. https://doi.org/10.1016/j.biopsycho.2016.03.012

Fagerland, M. W., Lydersen, S., & Laake, P. (2017). *Statistical analysis of contingency tables*. CRC Press.

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. https://doi.org/10.1111/j.1469-7610.2012.02617.x

Falk, Ö., Wallinius, M., Lundström, S., Frisell, T., Anckarsäter, H., & Kerekes, N. (2014). The 1 % of the population accountable for 63 % of all violent crime convictions. *Social Psychiatry and Psychiatric Epidemiology*, *49*(4), 559–571. https://doi.org/10.1007/s00127-013-0783-y

Fallgatter, A. J., Bartsch, A. J., & Herrmann, M. J. (2002). Electrophysiological measurements of anterior cingulate function. *Journal of Neural Transmission*, *109*(5-6), 977–988. https://doi.org/10.1007/s007020200080

Fazel, S., Singh, J. P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *BMJ*, *345*(jul24 2), e4692–e4692. https://doi.org/10.1136/bmj.e4692

Feilhauer, J., Cima, M., Korebrits, A., & Kunert, H.-J. (2012). Differential associations between psychopathy dimensions, types of aggression, and response inhibition. *Aggressive Behavior*, *38*(1), 77–88. https://doi.org/10.1002/ab.20415

Fisher, R. A. (1950). *Statistical methods for research workers* (Eleventh edition). Oliver and Boyd.

Fisher, R. A. (1956). *Statistical methods and scientific inference*. Oliver and Boyd.

Fox, B., & DeLisi, M. (2019). Psychopathic killers: A meta-analytic review of the psychopathy-homicide nexus. *Aggression and Violent Behavior*, *44*, 67–79. https://doi.org/10.1016/j.avb.2018.11.005

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Gajewski, P. D., & Falkenstein, M. (2013). Effects of task complexity on ERP components in Go/NoGo tasks. *International Journal of Psychophysiology*, *87*(3), 273–278. https://doi.org/10.1016/j.ijpsycho.2012.08.007

Gao, Y., Jiang, Y., Ming, Q., Zhang, J., Ma, R., Wu, Q., Dong, D., Guo, X., Liu, M., Wang, X., Situ, W., Pauli, R., & Yao, S. (2020). Gray matter changes in the orbitofrontal-paralimbic cortex in male youths with non-comorbid conduct disorder. *Frontiers in Psychology*, *11*, 843. https://doi.org/10.3389/fpsyg.2020.00843

Gao, Y., & Raine, A. (2009). P3 event-related potential impairments in antisocial and psychopathic individuals: A meta-analysis. *Biological Psychology*, *82*(3), 199–210. https://doi.org/10.1016/j.biopsycho.2009.06.006

Gaser, C., & Dahnke, R. (2016). CAT - A Computational Anatomy Toolbox for the analysis of structural MRI data. *OHBM Annual Meeting*.

Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, *3*(3), 445–449. https://doi.org/10.1214/08-BA318

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (Third edition). Chapman & Hall/CRC.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555. https://doi.org/10.3390/e19100555

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218. https://doi.org/10.1177/2515245918771329

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*(5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033

Gkotsi, G. M., & Gasser, J. (2016). Neuroscience in forensic psychiatry: From responsibility to dangerousness. Ethical and legal implications of using neuroscience for dangerousness assessments. *International Journal of Law and Psychiatry*, *46*, 58–67. https://doi.org/10.1016/j.ijlp.2016.02.030

Glenn, A. L., & Raine, A. (2014). Neurocriminology: Implications for the punishment, prediction and prevention of criminal behaviour. *Nature Reviews Neuroscience*, *15*(1), 54–63. https://doi.org/10.1038/nrn3640

Glenn, A. L., & Raine, A. (2009). Psychopathy and instrumental aggression: Evolutionary, neurobiological, and legal perspectives. *International Journal of Law and Psychiatry*, *32*(4), 253–258. https://doi.org/10.1016/j.ijlp.2009.04.002

Goethals, I., Audenaert, K., Jacobs, F., Van den Eynde, F., Bernagie, K., Kolindou, A., Vervaet, M., Dierckx, R., & Van Heeringen, C. (2005). Brain perfusion SPECT in impulsivity-related personality disorders. *Behavioural Brain Research*, *157*(1), 187–192. https://doi.org/10.1016/j.bbr.2004.06.022

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, *12*(11), 652–669. https://doi.org/10.1038/nrn3119

Gorenstein, E. E., & Newman, J. P. (1980). Disinhibitory psychopathology: A new perspective and a model for research. *Psychological Review*, *87*(3), 301–315. https://doi.org/10.1037/0033-295X.87.3.301

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Granvald, V., & Marciszko, C. (2016). Relations between key executive functions and aggression in childhood. *Child Neuropsychology*, *22*(5), 537–555. https://doi.org/10.1080/09297049.2015.1018152

Gray, N. S., Weidacker, K., & Snowden, R. J. (2019). Psychopathy and impulsivity: The relationship of psychopathy to different aspects of UPPS-P impulsivity. *Psychiatry Research*, *272*, 474–482. https://doi.org/10.1016/j.psychres.2018.12.155

Guan, M., Liao, Y., Ren, H., Wang, X., Yang, Q., Liu, X., & Wang, W. (2015). Impaired response inhibition in juvenile delinquents with antisocial personality characteristics: A preliminary ERP study in a Go/Nogo task. *Neuroscience Letters*, *603*, 1–5. https://doi.org/10.1016/j.neulet.2015.06.062

Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge University Press.

Haller, H., & Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research Online*, *7*(1), 1–20.

Hanniball, K. B., Gatner, D. T., Douglas, K. S., Viljoen, J. L., & Aknin, L. B. (2019). Examining the Triarchic Psychopathy Measure and comprehensive assessment of psychopathic personality in self-identified offender populations. *Personality Disorders: Theory, Research, and Treatment*, *10*(4), 340–353.

https://doi.org/10.1037/per0000333

Hanson, R. F., Sawyer, G. K., Begle, A. M., & Hubel, G. S. (2010). The impact of crime victimization on quality of life. *Journal of Traumatic Stress*, *23*(2), 189–197. https://doi.org/10.1002/jts.20508

Hare, R. D. (2003). *The Psychopathy Checklist-Revised*. Multi-Health Systems.

Hare, R. D., & Neumann, C. S. (2009). Psychopathy: Assessment and forensic implications. *The Canadian Journal of Psychiatry*, *54*(12), 791–802.

Harris, J. J., Jolivet, R., & Attwell, D. (2012). Synaptic energy use and supply. *Neuron*, *75*(5), 762–777. https://doi.org/10.1016/j.neuron.2012.08.019

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Psychopathy Checklist: Screening Version (PCL:SV)*. Multi-Health Systems.

Heaton, S. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Psychological Assessment Resources.

Hecht, L. K., & Latzman, R. D. (2018). Exploring the differential associations between components of executive functioning and reactive and proactive aggression. *Journal of Clinical and Experimental Neuropsychology*, *40*(1), 62–74. https://doi.org/10.1080/13803395.2017.1314450

Hegarty, C. E., Foland-Ross, L. C., Narr, K. L., Townsend, J. D., Bookheimer, S. Y., Thompson, P. M., & Altshuler, L. L. (2012). Anterior cingulate activation relates to local cortical thickness: *NeuroReport*, *23*(7), 420–424. https://doi.org/10.1097/WNR.0b013e3283525a95

Heilbronner, S. R., & Hayden, B. Y. (2016). Dorsal anterior cingulate cortex: A bottom-up view. *Annual Review of Neuroscience*, *39*(1), 149–170. https://doi.org/10.1146/annurev-neuro-070815-013952

Hicks, B. M., Foster, K. T., Iacono, W. G., & McGue, M. (2013). Genetic and environmental influences on the familial transmission of externalizing disorders in adoptive and twin offspring. *JAMA Psychiatry*, *70*(10), 1076–1083. https://doi.org/10.1001/jamapsychiatry.2013.258

Hiser, J., & Koenigs, M. (2018). The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biological Psychiatry*, *83*(8), 638–647. https://doi.org/10.1016/j.biopsych.2017.10.030

Holmes, A. J., Hollinshead, M. O., Roffman, J. L., Smoller, J. W., & Buckner, R. L. (2016). Individual differences in cognitive control circuit anatomy link sensation seeking, impulsivity, and substance use. *The Journal of Neuroscience*, *36*(14), 4038–4049. https://doi.org/10.1523/JNEUROSCI.3206-15.2016

Hong, X., Wang, Y., Sun, J., Li, C., & Tong, S. (2017). Segregating top-down selective attention from response inhibition in a spatial cueing Go/NoGo task: An ERP and source localization study. *Scientific Reports*, *7*(1), 1–14.

https://doi.org/10.1038/s41598-017-08807-z

Hopwood, C. J. (2018). A framework for treating DSM-5 alternative model for personality disorder features: A framework for treating DSM-5 alternative model for personality disorder features. *Personality and Mental Health*, *12*(2), 107–125. https://doi.org/10.1002/pmh.1414

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews. Cancer*, *18*(8), 500–510. https://doi.org/10.1038/s41568-018-0016-5

Howner, K., Andiné, P., Bertilsson, G., Hultcrantz, M., Lindström, E., Mowafi, F., Snellman, A., & Hofvander, B. (2018). Mapping systematic reviews on forensic psychiatric care: A systematic review identifying knowledge gaps. *Frontiers in Psychiatry*, *9*, 452. https://doi.org/10.3389/fpsyt.2018.00452

Huband, N., McMurran, M., Evans, C., & Duggan, C. (2007). Social problem-solving plus psychoeducation for adults with personality disorder. *British Journal of Psychiatry*, *190*, 307–313.

Huda, W., & Sandison, G. A. (1989). The use of the effective dose equivalent, HE, for 99mTc labelled radiopharmaceuticals. *European Journal of Nuclear Medicine*, *15*(4), 174–179. https://doi.org/10.1007/BF00253789

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., & Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. *International Journal of Psychophysiology*, *87*(3), 217–233. https://doi.org/10.1016/j.ijpsycho.2012.08.001

Huster, R. J., Westerhausen, R., Pantev, C., & Konrad, C. (2010). The role of the cingulate cortex as neural generator of the N200 and P300 in a tactile response inhibition task. *Human Brain Mapping*, *31*(8), 1260–1271. https://doi.org/10.1002/hbm.20933

Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, *8*(9), 725–732. https://doi.org/10.1038/nrn2218

Iacono, W. G., Carlson, S. R., Malone, S. M., & McGue, M. (2002). P3 event-related potential amplitude and the risk for disinhibitory disorders in adolescent boys. *Archives of General Psychiatry*, *59*(8), 750–757. https://doi.org/10.1001/archpsyc.59.8.750

Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & Mcgue, M. (1999). Behavioral disinhibition and the development of substance-use disorders: Findings from the Minnesota Twin Family Study. *Development and Psychopathology*, *11*(4), 869–900. https://doi.org/10.1017/S0954579499002369

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Jackson, D. B., & Vaughn, M. G. (2018). Maternal medical risks during pregnancy and childhood externalizing behavior. *Social Science & Medicine*, *207*,

19–24. https://doi.org/10.1016/j.socscimed.2018.04.032

Jacobus, J., Castro, N., Squeglia, L. M., Meloy, M. J., Brumback, T., Huestis, M. A., & Tapert, S. F. (2016). Adolescent cortical thickness pre- and post marijuana and alcohol initiation. *Neurotoxicology and Teratology*, *57*, 20–29. https://doi.org/10.1016/j.ntt.2016.09.005

Jas, M., Larson, E., Engemann, D. A., Leppäkangas, J., Taulu, S., Hämäläinen, M., & Gramfort, A. (2018). A reproducible MEG/EEG group study with the MNE doftware: Recommendations, quality assessments, and good practices. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00530

Jeandarme, I., Edens, J. F., Habets, P., Bruckers, L., Oei, K., & Bogaerts, S. (2017). PCL-R field validity in prison and hospital settings. *Law and Human Behavior*, *41*(1), 29–43. https://doi.org/10.1037/lhb0000222

Jiang, W., Li, G., Liu, H., Shi, F., Wang, T., Shen, C., Shen, H., Lee, S.-W., Hu, D., Wang, W., & Shen, D. (2016). Reduced cortical thickness and increased surface area in antisocial personality disorder. *Neuroscience*, *337*, 143–152. https://doi.org/10.1016/j.neuroscience.2016.08.052

Jones, A., McGrath, E., Robinson, E., Houben, K., Nederkoorn, C., & Field, M. (2018). A randomized controlled trial of inhibitory control training for the reduction of alcohol consumption in problem drinkers. *Journal of Consulting and Clinical Psychology*, *86*(12), 991–1004. https://doi.org/10.1037/ccp0000312

Jurjako, M., Malatesti, L., & Brazil, I. A. (2020). Biocognitive classification of antisocial Iindividuals without explanatory reductionism. *Perspectives on Psychological Science*, *15*(4), 957–972. https://doi.org/10.1177/1745691620904160

Kahnt, T., Chang, L. J., Park, S. Q., Heinzle, J., & Haynes, J.-D. (2012). Connectivity-based parcellation of the human orbitofrontal cortex. *The Journal of Neuroscience*, *32*(18), 6240–6250. https://doi.org/10.1523/JNEUROSCI.0257-12.2012

Kaiser, A., Aggensteiner, P.-M., Baumeister, S., Holz, N. E., Banaschewski, T., & Brandeis, D. (2020). Earlier versus later cognitive event-related potentials (ERPs) in attention-deficit/hyperactivity disorder (ADHD): A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *112*, 117–134. https://doi.org/10.1016/j.neubiorev.2020.01.019

Kendler, K. S., Lönn, S. L., Maes, H. H., Lichtenstein, P., Sundquist, J., & Sundquist, K. (2016). A swedish population-based multivariate twin study of externalizing disorders. *Behavior Genetics*, *46*(2), 183–192. https://doi.org/10.1007/s10519-015-9741-7

Kennealy, P. J., Skeem, J. L., Walters, G. D., & Camp, J. (2010). Do core interpersonal and affective traits of PCL-R psychopathy interact with antisocial behavior and disinhibition to predict violence? *Psychological Assessment*, *22*(3), 569–580. https://doi.org/10.1037/a0019618

Kennedy-Shaffer, L. (2019). Before $p < 0.05$ to beyond $p < 0.05$: Using history to contextualize p -values and significance testing. *The American Statistician*, *73*(sup1), 82–90. https://doi.org/10.1080/00031305.2018.1537891

Kersting, K. (2018). Machine learning and artificial intelligence: Two fellow travelers on the quest for intelligent behavior in machines. *Frontiers in Big Data*, *1*. https://doi.org/10.3389/fdata.2018.00006

Kiehl, K. A., & Hoffman, M. B. (2011). The criminal psychopath: History, neuroscience, and economics. *Jurimetrics*, *51*, 355–397.

Kiehl, K. A., Liddle, P. F., & Hopfinger, J. B. (2000). Error processing and the rostral anterior cingulate: An event-related fMRI study. *Psychophysiology*, *37*(2), 216–223. https://doi.org/10.1111/1469-8986.3720216

Kirschstein, T., & Köhling, R. (2009). What is the source of the EEG? *Clinical EEG and Neuroscience*, *40*(3), 146–149. https://doi.org/10.1177/155005940904000305

Kotov, R., Krueger, R. F., & Watson, D. (2018). A paradigm shift in psychiatric classification: The Hierarchical Taxonomy Of Psychopathology (HiTOP). *World Psychiatry*, *17*(1), 24–25. https://doi.org/10.1002/wps.20478

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH Research Domain Criteria initiative: Background, issues, and pragmatics. *Psychophysiology*, *53*(3), 286–297. https://doi.org/10.1111/psyp.12518

Kriminalvården. (2014a). *Klientkartläggning 2013 - Ett regeringsuppdrag: En presentation av bakgrundsfaktorer hos Kriminalvårdens klienter*. Kriminalvården.

Kriminalvården. (2014b). *Riskbedömning i kriminalvård och rättspsykiatri: Sammanfattningsrapport*. Kriminalvården.

Kriminalvården. (2020). *Kriminalvård och statistik 2019*. Kriminalvården.

Krona, H., Nyman, M., Andreasson, H., Vicencio, N., Anckarsäter, H., Wallinius, M., Nilsson, T., & Hofvander, B. (2017). Mentally disordered offenders in Sweden: Differentiating recidivists from non-recidivists in a 10-year follow-up study. *Nordic Journal of Psychiatry*, *71*(2), 102–109. https://doi.org/10.1080/08039488.2016.1236400

Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., Simms, L. J., Widiger, T. A., Achenbach, T. M., Bach, B., Bagby, R. M., Bornovalova, M. A., Carpenter, W. T., Chmielewski, M., Cicero, D. C., Clark, L. A., Conway, C., DeClercq, B., DeYoung, C. G., … Zimmermann, J.

(2018). *Progress in achieving quantitative classification of psychopathology*. *17*(3), 282–293. https://doi.org/10.1002/wps.20566

Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, *116*(4), 645–666. https://doi.org/10.1037/0021-843X.116.4.645

Krueger, R. F., & South, S. C. (2009). Externalizing disorders: Cluster 5 of the proposed meta-structure for DSM-V and ICD-11. *Psychological Medicine*, *39*(12), 2061–2070. https://doi.org/10.1017/S0033291709990328

Krueger, R. F., & Tackett, J. L. (2015). The externalizing spectrum of personality and psychopathology: An empirical and quantitative alternative to discrete disorder approaches. In T. P. Beauchaine & S. P. Hinshaw (Eds.), *Oxford handbook of externalizing disorders* (pp. 79–89). Oxford University Press.

Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Second edition). Academic Press.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Kühn, S., Schubert, F., & Gallinat, J. (2011). Structural correlates of trait anxiety: Reduced thickness in medial orbitofrontal cortex accompanied by volume increase in nucleus accumbens. *Journal of Affective Disorders*, *134*(1-3), 315–319. https://doi.org/10.1016/j.jad.2011.06.003

Kyranides, M. N., Fanti, K. A., Sikki, M., & Patrick, C. J. (2017). Triarchic dimensions of psychopathy in young adulthood: Associations with clinical and physiological measures after accounting for adolescent psychopathic traits. *Personality Disorders: Theory, Research, and Treatment*, *8*(2), 140–149. https://doi.org/10.1037/per0000193

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, *84*(408), 881–896.

Laurinavičius, A., Sellbom, M., Klimukienė, V., Wygant, D. B., Laurinaitytė, I., Ustinavičiūtė, L., & Baltrūnas, M. (2020). Examination of Triarchic Psychopathy Measure in a sample of Lithuanian juvenile offenders. *Psychological Assessment*, *32*(4), 407–413. https://doi.org/10.1037/pas0000793

Leistico, A.-M. R., Salekin, R. T., DeCoster, J., & Rogers, R. (2008). A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior*, *32*(1), 28–45. https://doi.org/10.1007/s10979-007-9096-6

Leutgeb, V., Leitner, M., Wabnegger, A., Klug, D., Scharmüller, W., Zussner, T., & Schienle, A. (2015). Brain abnormalities in high-risk violent offenders and their association with psychopathic traits and criminal recidivism. *Neuroscience*, *308*, 194–201. https://doi.org/10.1016/j.neuroscience.2015.09.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Lilienfeld, S. O., & Widows, M. R. (2005). *Psychopathy Personality Inventory Revised (PPI-R). Professional manual*. Psychological Assessment Resources.

Lindberg, N., Laajasalo, T., Holi, M., Putkonen, H., Weizmann-Henelius, G., & Häkkänen-Nyholm, H. (2009). Psychopathic traits and offender characteristics a nationwide consecutive sample of homicidal male adolescents. *BMC Psychiatry*, *9*(1), 18. https://doi.org/10.1186/1471-244X-9-18

Ling, S., Umbach, R., & Raine, A. (2019). Biological explanations of criminal behavior. *Psychology, Crime & Law*, *25*(6), 626–640. https://doi.org/10.1080/1068316X.2019.1572753

Littman, R., & Takács, Á. (2017). Do all inhibitions act alike? A study of Go/No-Go and Stop-Signal paradigms. *PLOS ONE*, *12*(10), e0186774. https://doi.org/10.1371/journal.pone.0186774

Ljubin-Golub, T., Petričević, E., & Sokić, K. (2019). Predicting academic Cheating with Triarchic Psychopathy and Cheating Attitudes. *Journal of Academic Ethics*. https://doi.org/10.1007/s10805-019-09338-0

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luria, A. R. (1961). *The role of speech in the regulation of normal and abnormal behavior*. Liveright Publishing Corporation.

Lydersen, S., Fagerland, M. W., & Laake, P. (2009). Recommended tests for association in 2 tables. *Statistics in Medicine*, *28*(7), 1159–1175. https://doi.org/10.1002/sim.3531

Mahoney, M. R., Farmer, C., Sinclair, S., Sung, S., Dehaut, K., & Chung, J. Y. (2020). Utilization of the DSM-5 Self-Rated Level 1 Cross-Cutting Symptom Measure-Adult to screen healthy volunteers for research studies. *Psychiatry Research*, *286*, 112822. https://doi.org/10.1016/j.psychres.2020.112822

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767. https://doi.org/10.3389/fpsyg.2019.02767

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., & Robbins, T. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, *125*(3), 624–639. https://doi.org/10.1093/brain/awf049

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathol-

ogy: A quantitative review. *Psychological Bulletin*, *137*(5), 856–879. https://doi.org/10.1037/a0023678

Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Edenhofer, O., Stocker, T. F., Field, C. B., Ebi, K. L., & Matschoss, P. R. (2011). The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change*, *108*(4), 675–691. https://doi.org/10.1007/s10584-011-0178-6

McCord, W. M., & McCord, J. (1964). *The psychopath: An essay on the criminal mind*. Van Nostrand Reinhold Company.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (Second edition). Chapman & Hall/CRC.

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods*, *11*(4), 386–401. https://doi.org/10.1037/1082-989X.11.4.386

McKinley, S., Patrick, C. J., & Verona, E. (2018). Antisocial personality disorder: Neurophysiological mechanisms and distinct subtypes. *Current Behavioral Neuroscience Reports*, *5*(1), 72–80. https://doi.org/10.1007/s40473-018-0142-0

McMurran, M., Fyffe, S., McCarthy, L., Duggan, C., & Latham, A. (2001). "Stop & Think!": Social problem-solving therapy with personality-disordered offenders. *Criminal Behaviour and Mental Health*, *11*(4), 273–285. https://doi.org/10.1002/cbm.401

McReynolds, L. S., Schwalbe, C. S., & Wasserman, G. A. (2010). The contribution of psychiatric disorder to juvenile recidivism. *Criminal Justice and Behavior*, *37*(2), 204–216. https://doi.org/10.1177/0093854809354961

McShane, B. B., & Gal, D. (2017). Statistical significance and the dichotomization of evidence. *Journal of the American Statistical Association*, *112*(519), 885–895. https://doi.org/10.1080/01621459.2017.1289846

Meijers, J., Harte, J. M., Jonker, F. A., & Meynen, G. (2015). Prison brain? Executive dysfunction in prisoners. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00043

Meijers, J., Harte, J. M., Meynen, G., & Cuijpers, P. (2017). Differences in executive functioning between violent and non-violent offenders. *Psychological Medicine*, *47*(10), 1784–1793. https://doi.org/10.1017/S0033291717000241

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, *12*, 131–143.

Metropolis, N. (1987). The beginning of the Monte Carlo method. *Los Alamos Science*, *15*, 125–130.

Milad, M. R., & Rauch, S. L. (2007). The role of the orbitofrontal cortex in anxiety disorders. *Annals of the New York Academy of Sciences*, *1121*(1), 546–561. https://doi.org/10.1196/annals.1401.006

Miller, E. K. (2000). The prefontral cortex and cognitive control. *Nature Reviews Neuroscience*, *1*(1), 59–65. https://doi.org/10.1038/35036228

Miller, G. A. (2010). Mistreating Psychology in the Decades of the Brain. *Perspectives on Psychological Science*, *5*(6), 716–743. https://doi.org/10.1177/1745691610388774

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moffitt, T. E. (2018). Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, *2*(3), 177–186. https://doi.org/10.1038/s41562-018-0309-4

Mokros, A., Neumann, C. S., Stadtland, C., Osterheider, M., Nedopil, N., & Hare, R. D. (2011). Assessing measurement invariance of PCL-R assessments from file reviews of North American and German offenders. *International Journal of Law and Psychiatry*, *34*(1), 56–63. https://doi.org/10.1016/j.ijlp.2010.11.009

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical Psychology Review*, *20*(1), 113–136. https://doi.org/10.1016/S0272-7358(98)00096-8

Mullins-Sweatt, S. N., DeShong, H. L., Lengel, G. J., Helle, A. C., & Krueger, R. F. (2019). Disinhibition as a unifying construct in understanding how personality dispositions undergird psychopathology. *Journal of Research in Personality*, *80*, 55–61. https://doi.org/10.1016/j.jrp.2019.04.006

Munthe, C., Radovic, S., & Anckarsäter, H. (2010). Ethical issues in forensic psychiatric research on mentally disordered offenders. *Bioethics*, *24*(1), 35–44. https://doi.org/10.1111/j.1467-8519.2009.01773.x

Murray, L., Waller, R., & Hyde, L. W. (2018). A systematic review examining the link between psychopathic personality traits, antisocial behavior, and neural reactivity during reward and loss processing. *Personality Disorders: Theory, Research, and Treatment*, *9*(6), 497–509. https://doi.org/10.1037/per0000308

Nadelhoffer, T., Bibas, S., Grafton, S., Kiehl, K. A., Mansfield, A., Sinnott-Armstrong, W., & Gazzaniga, M. (2012). Neuroprediction, violence, and the law: Setting the stage. *Neuroethics*, *5*(1), 67–99. https://doi.org/10.1007/s12152-010-9095-z

Nelson, L. D., Strickland, C., Krueger, R. F., Arbisi, P. A., & Patrick, C. J. (2016). Neurobehavioral traits as transdiagnostic predictors of clinical problems.

Assessment, *23*(1), 75–85. https://doi.org/10.1177/1073191115570110

Neumann, C. S., Johansson, P. T., & Hare, R. D. (2013). The Psychopathy Checklist-Revised (PCL-R), low anxiety, and fearlessness: A structural equation modeling analysis. *Personality Disorders: Theory, Research, and Treatment*, *4*(2), 129–137. https://doi.org/10.1037/a0027886

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Nilsson, I., & Wadeskog, A. (2012). *Gatuvåldets ekonomi: Ett kommunalt perspektiv*. Skandia.

Nitschke, K., Köstering, L., Finkel, L., Weiller, C., & Kaller, C. P. (2017). A meta-analysis on the neural basis of planning: Activation likelihood estimation of functional brain imaging results in the Tower of London task. *Human Brain Mapping*, *38*(1), 396–413. https://doi.org/10.1002/hbm.23368

Nuijten, M. B., Wetzels, R., Matzke, D., Dolan, C. V., & Wagenmakers, E.-J. (2015). A default Bayesian hypothesis test for mediation. *Behavior Research Methods*, *47*(1), 85–97. https://doi.org/10.3758/s13428-014-0470-2

Ogilvie, J. M., Stewart, A. L., Chan, R. C. K., & Shum, D. H. K. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology*, *49*(4), 1063–1107. https://doi.org/10.1111/j.1745-9125.2011.00252.x

Olver, M. E. (2016). Treatment of psychopathic offenders: Evidence, issues, and controversies. *Journal of Community Safety and Well-Being*, *1*(3), 75. https://doi.org/10.35502/jcswb.25

Olver, M. E., & Wong, S. C. P. (2015). Short- and long-term recidivism prediction of the PCL-R and the effects of age: A 24-year follow-up. *Personality Disorders: Theory, Research, and Treatment*, *6*(1), 97–105. https://doi.org/10.1037/per0000095

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*(10), 1021–1034. https://doi.org/10.1016/0028-3932(90)90137-d

Paiva, T. O., Pasion, R., Patrick, C. J., Moreira, D., Almeida, P. R., & Barbosa, F. (2020). Further evaluation of the Triarchic Psychopathy Measure: Evidence from community adult and prisoner samples from Portugal. *Psychological Assessment*, *32*(3), e1–e14. https://doi.org/10.1037/pas0000797

Palomero-Gallagher, N., Vogt, B. A., Schleicher, A., Mayberg, H. S., & Zilles, K. (2009). Receptor architecture of human cingulate cortex: Evaluation of the four-region neurobiological model. *Human Brain Mapping*, *30*(8), 2336–2355. https://doi.org/10.1002/hbm.20667

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., Xian, H., Tsuang, M., Fischl, B., Seidman, L., Dale, A., & Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex (New York, N.Y.: 1991)*, *19*(11), 2728–2735. https://doi.org/10.1093/cercor/bhp026

Pasion, R., Cruz, A. R., & Barbosa, F. (2018). Dissociable effects of psychopathic traits on executive functioning: Insights from the Triarchic Model. *Frontiers in Psychology*, *9*, 1713. https://doi.org/10.3389/fpsyg.2018.01713

Patrick, C. J., Drislane, L. E., & Strickland, C. (2012). Conceptualizing psychopathy in triarchic terms: Implications for treatment. *International Journal of Forensic Mental Health*, *11*(4), 253–266. https://doi.org/10.1080/14999013.2012.746761

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: Developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology*, *21*(3), 913–938. https://doi.org/10.1017/S0954579409000492

Patrick, C. J., Iacono, W. G., & Venables, N. C. (2019). Incorporating neurophysiological measures into clinical assessments: Fundamental challenges and a strategy for addressing them. *Psychological Assessment*, *31*(12), 1512–1529. https://doi.org/10.1037/pas0000713

Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: Development of a brief form of the Externalizing Spectrum Inventory. *Psychological Assessment*, *25*(4), 1332–1348. https://doi.org/10.1037/a0034864

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:6%3C768::aid-jclp2270510607%3E3.0.co;2-1

Pauli, M., Liljeberg, J., Gustavsson, P., Kristiansson, M., & Howner, K. (2019). Assessing the relevance of self-reported ADHD symptoms and cognitive functioning for psychopathy using the PCL-R and the TriPM. *The Journal of Forensic Psychiatry & Psychology*, *30*(4), 642–657. https://doi.org/10.1080/14789949.2018.1560489

Pavlov, I. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford University Press.

Perkins, E. R., Joyner, K. J., Patrick, C. J., Bartholow, B. D., Latzman, R. D., DeYoung, C. G., Kotov, R., Reininghaus, U., Cooper, S. E., Afzali, M.

H., Docherty, A. R., Dretsch, M. N., Eaton, N. R., Goghari, V. M., Haltigan, J. D., Krueger, R. F., Martin, E. A., Michelini, G., Ruocco, A. C., … Zald, D. H. (2020). Neurobiology and the Hierarchical Taxonomy of Psychopathology: Progress toward ontogenetically informed and clinically useful nosology. *Dialogues in Clinical Neuroscience*, *22*(1), 51–63. https://doi.org/10.31887/DCNS.2020.22.1/eperkins

Perkins, E. R., Latzman, R. D., & Patrick, C. J. (2020). Interfacing neural constructs with the Hierarchical Taxonomy of Psychopathology: "Why" and "how". *Personality and Mental Health*, *14*(1), 106–122. https://doi.org/10.1002/pmh.1460

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshopon Distributed Statistical Computing (DSC 2003)*.

Polaschek, D. L. L. (2012). An appraisal of the risk-need-responsivity (RNR) model of offender rehabilitation and its application in correctional treatment. *Legal and Criminological Psychology*, *17*(1), 1–17. https://doi.org/10.1111/j.2044-8333.2011.02038.x

Poldrack, R. A., Monahan, J., Imrey, P. B., Reyna, V., Raichle, M. E., Faigman, D., & Buckholtz, J. W. (2018). Predicting violent behavior: What can neuroscience add? *Trends in Cognitive Sciences*, *22*(2), 111–123. https://doi.org/10.1016/j.tics.2017.11.003

Pribram, K. H. (1973). The primate frontal cortex - Executive of the brain. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 293–314). Academic Press.

Proposition (Government Legislative Bill) 1990/91:58. (1990). *Regeringens proposition 1990/91:58 om psykiatrisk tvångsvård, m.m.*

Raine, A. (2018). Antisocial personality as a neurodevelopmental disorder. *Annual Review of Clinical Psychology*, *14*, 259–289. https://doi.org/10.1146/annurev-clinpsy-050817-084819

Raine, A. (2019). The neuromoral theory of antisocial, violent, and psychopathic behavior. *Psychiatry Research*, *277*, 64–69. https://doi.org/10.1016/j.psychres.2018.11.025

Raine, A., Buchsbaum, M., & LaCasse, L. (1997). Brain abnormalities in murderers indicated by positron emission tomography. *Biological Psychiatry*, *42*(6), 495–508. https://doi.org/10.1016/S0006-3223(96)00362-9

Raine, A., & Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Social Cognitive and Affective Neuroscience*, *1*(3), 203–213. https://doi.org/10.1093/scan/nsl033

Rättsmedicinalverket. (2020). *Årsredovisning 2019*. Rättsmedicinalverket.

RättspsyK. (2020). *Årsrapport 2019*. Nationellt rättspsykiatriskt kvalitetsregister (RättspsyK).

R Core Team. (2020). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.

Reidy, D. E., Kearns, M. C., DeGue, S., Lilienfeld, S. O., Massetti, G., & Kiehl, K. A. (2015). Why psychopathy matters: Implications for public health and violence prevention. *Aggression and Violent Behavior*, *24*, 214–225. https://doi.org/10.1016/j.avb.2015.05.018

Ribes-Guardiola, P., Poy, R., Patrick, C. J., & Moltó, J. (2020). Electrocortical measures of performance monitoring from Go/No-Go and flanker tasks: Differential relations with trait dimensions of the triarchic model of psychopathy. *Psychophysiology*, *57*(6). https://doi.org/10.1111/psyp.13573

Ribes-Guardiola, P., Poy, R., Segarra, P., Branchadell, V., & Moltó, J. (2020). Response perseveration and the triarchic model of psychopathy in an undergraduate sample. *Personality Disorders*, *11*(1), 54–62. https://doi.org/10.1037/per0000371

Robert, C., & Casella, G. (2011). A short history of Markov Chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, *26*(1), 102–115. https://doi.org/10.1214/10-STS351

Rodman, A. M., Kastman, E. K., Dorfman, H. M., Baskin-Sommers, A. R., Kiehl, K. A., Newman, J. P., & Buckholtz, J. W. (2016). Selective mapping of psychopathy and externalizing to dissociable circuits for inhibitory self-control. *Clinical Psychological Science*, *4*(3), 559–571. https://doi.org/10.1177/2167702616631495

Rohlf, H. L., Holl, A. K., Kirsch, F., Krahé, B., & Elsner, B. (2018). Longitudinal links between executive function, anger, and aggression in middle childhood. *Frontiers in Behavioral Neuroscience*, *12*, 27. https://doi.org/10.3389/fnbeh.2018.00027

Rolls, E. T. (2019). The orbitofrontal cortex and emotion in health and disease, including depression. *Neuropsychologia*, *128*, 14–43. https://doi.org/10.1016/j.neuropsychologia.2017.09.021

Sachdev, P. S., Blacker, D., Blazer, D. G., Ganguli, M., Jeste, D. V., Paulsen, J. S., & Petersen, R. C. (2014). Classifying neurocognitive disorders: The DSM-5 approach. *Nature Reviews Neurology*, *10*(11), 634–642. https://doi.org/10.1038/nrneurol.2014.181

Sadeh, N., & Verona, E. (2008). Psychopathic personality traits associated with abnormal selective attention and impaired cognitive control. *Neuropsychology*, *22*(5), 669–680. https://doi.org/10.1037/a0012692

Schmahmann, J. D. (2010). The role of the cerebellum in cognition and emotion: Personal reflections since 1982 on the dysmetria of thought hypothesis, and its historical evolution from theory to therapy. *Neuropsychology Review*, *20*(3), 236–260. https://doi.org/10.1007/s11065-010-9142-x

Schmajuk, M., Liotti, M., Busse, L., & Woldorff, M. G. (2006). Electrophysiological activity underlying inhibitory control processes in normal adults. *Neu-*

ropsychologia, *44*(3), 384–395. https://doi.org/10.1016/j.neuropsychologia.2005.06.005

Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of Abnormal Child Psychology*, *41*(3), 457–471. https://doi.org/10.1007/s10802-012-9684-x

Sergiou, C. S., Santarnecchi, E., Franken, I. H. A., Rassin, E., van Beveren, N., & van Dongen, J. D. M. (2020). *tDCS targeting the vmPFC modulates reactive aggression and electrophysiological responses: An EEG-tDCS study in a forensic population* [PsyArXiv].

Séguin, J. R. (2009). The frontal lobe and aggression. *European Journal of Developmental Psychology*, *6*(1), 100–119. https://doi.org/10.1080/17405620701669871

Séguin, J. R. (2004). Neurocognitive elements of antisocial behavior: Relevance of an orbitofrontal cortex account. *Brain and Cognition*, *55*(1), 185–197. https://doi.org/10.1016/S0278-2626(03)00273-2

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*(1089), 199–209. https://doi.org/10.1098/rstb.1982.0082

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Siever, L. J., Buchsbaum, M. S., New, A. S., Spiegel-Cohen, J., Wei, T., Hazlett, E. A., Sevin, E., Nunn, M., & Mitropoulou, V. (1999). D,l-fenfluramine response in impulsive personality disorder assessed with [18F]Fluorodeoxyglucose positron emission tomography. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *20*(5), 413–423. https://doi.org/10.1016/S0893-133X(98)00111-0

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, *46*(1), 224–232. https://doi.org/10.1016/j.neuropsychologia.2007.07.015

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, K., & Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, *13*(3), 193–206. https://doi.org/10.1080/14999013.2014.922141

Singh, J. P., & Fazel, S. (2010). Forensic risk assessment: A metareview. *Criminal Justice and Behavior*, *37*(9), 965–988. https://doi.org/10.1177/0093854810374274

Skeem, J. L., & Cooke, D. J. (2010). Is criminal behavior a central component of psychopathy? Conceptual directions for resolving the debate. *Psychological Assessment*, *22*(2), 433–445. https://doi.org/10.1037/a0008512

Skeem, J. L., Winter, E., Kennealy, P. J., Louden, J. E., & Tatar, J. R. (2014). Offenders with mental illness have criminogenic needs, too: Toward recidivism reduction. *Law and Human Behavior*, *38*(3), 212–224. https://doi.org/10.1037/lhb0000054

Snowden, R. J., & Gray, N. S. (2011). Impulsivity and psychopathy: Associations between the Barrett Impulsivity Scale and the Psychopathy Checklist revised. *Psychiatry Research*, *187*(3), 414–417. https://doi.org/10.1016/j.psychres.2011.02.003

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00328

Soderstrom, H., Tullberg, M., Wikkelsö, C., Ekholm, S., & Forsman, A. (2000). Reduced regional cerebral blood flow in non-psychotic violent offenders. *Psychiatry Research*, *98*(1), 29–41. https://doi.org/10.1016/s0925-4927(99)00049-9

Somma, A., Borroni, S., Drislane, L. E., Patrick, C. J., & Fossati, A. (2019). Modeling the structure of the Triarchic Psychopathy Measure: Conceptual, empirical, and analytic considerations. *Journal of Personality Disorders*, *33*(4), 470–496. https://doi.org/10.1521/pedi_2018_32_354

Spierer, L., Chavan, C., & Manuel, A. L. (2013). Training-induced behavioral and brain plasticity in inhibitory control. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00427

Stanton, K., Brown, M. F. D., & Watson, D. (2020). Examining the item-level structure of the Triarchic Psychopathy Measure: Sharpening assessment of psychopathy constructs. *Assessment*, 1073191120927786. https://doi.org/10.1177/1073191120927786

Steele, V. R., Aharoni, E., Munro, G. E., Calhoun, V. D., Nyalakanti, P., Stevens, M. C., Pearlson, G., & Kiehl, K. A. (2013). A large scale (N=102) functional neuroimaging study of response inhibition in a Go/NoGo task. *Behavioural Brain Research*, *256*, 529–536. https://doi.org/10.1016/j.bbr.2013.06.001

Steele, V. R., Claus, E. D., Aharoni, E., Vincent, G. M., Calhoun, V. D., & Kiehl, K. A. (2015). Multimodal imaging measures predict rearrest. *Frontiers in Human Neuroscience*, *9*, 425. https://doi.org/10.3389/fnhum.2015.00425

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*, *21*(1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Sutton, A., Roche, M., Stapleton, M., & Roemer, A. (2020). Can psychopathy be adaptive at work? Development and application of a work focused self- and other-report measure of the Triarchic Psychopathy Model. *International Journal of Environmental Research and Public Health*, *17*(11), 3938. https://doi.org/10.3390/ijerph17113938

Svennerlind, C., Nilsson, T., Kerekes, N., Andiné, P., Lagerkvist, M., Forsman, A., Anckarsäter, H., & Malmgren, H. (2010). Mentally disordered criminal offenders in the Swedish criminal system. *International Journal of Law and Psychiatry*, *33*(4), 220–226. https://doi.org/10.1016/j.ijlp.2010.06.003

Tan, S.-Y., & Haining, R. (2016). Crime victimization and the implications for individual health and wellbeing: A Sheffield case study. *Social Science & Medicine*, *167*, 128–139. https://doi.org/10.1016/j.socscimed.2016.08.018

Taylor, J., Reeves, M., James, L., & Bobadilla, L. (2006). Disinhibitory trait profile and its relation to Cluster B personality disorder features and substance use problems. *European Journal of Personality*, *20*(4), 271–284. https://doi.org/10.1002/per.585

Tiego, J., Testa, R., Bellgrove, M. A., Pantelis, C., & Whittle, S. (2018). A hierarchical model of inhibitory control. *Frontiers in Psychology*, *9*, 1339. https://doi.org/10.3389/fpsyg.2018.01339

Tiihonen, J., Rossi, R., Laakso, M. P., Hodgins, S., Testa, C., Perez, J., Repo-Tiihonen, E., Vaurio, O., Soininen, H., Aronen, H. J., Könönen, M., Thompson, P. M., & Frisoni, G. B. (2008). Brain anatomy of persistent violent offenders: More rather than less. *Psychiatry Research: Neuroimaging*, *163*(3), 201–212. https://doi.org/10.1016/j.pscychresns.2007.08.012

Tortora, L., Meynen, G., Bijlsma, J., Tronci, E., & Ferracuti, S. (2020). Neuroprediction and AI in forensic psychiatry and criminal justice: A neurolaw perspective. *Frontiers in Psychology*, *11*, 220. https://doi.org/10.3389/fpsyg.2020.00220

Tuvblad, C., Grann, M., & Lichtenstein, P. (2006). Heritability for adolescent antisocial behavior differs with socioeconomic status: Gene-environment interaction. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(7), 734–743. https://doi.org/10.1111/j.1469-7610.2005.01552.x

Tuvblad, C., Wang, P., Patrick, C. J., Berntsen, L., Raine, A., & Baker, L. A. (2019). Genetic and environmental influences on disinhibition, boldness, and meanness as assessed by the triarchic psychopathy measure in 19-year-old twins. *Psychological Medicine*, *49*(09), 1500–1509. https://doi.org/10.1017/S0033291718002052

van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 development across the lifespan: A systematic review and meta-analysis. *PLoS ONE*, *9*(2), e87347. https://doi.org/10.1371/journal.pone.0087347

van Dongen, J. D. M., Drislane, L. E., Nijman, H., Soe-Agnie, S. E., & van Marle, H. J. C. (2017). Further evidence for reliability and validity of the

Triarchic Psychopathy Measure in a forensic sample and a community sample. *Journal of Psychopathology and Behavioral Assessment*, *39*(1), 58–66. https://doi.org/10.1007/s10862-016-9567-5

van Dongen, J. D. M., & Franken, I. H. A. (2019). Neuroscience in forensic psychiatry and psychology: An introduction to the special issue. *International Journal of Forensic Mental Health*, *18*(3), 179–186. https://doi.org/10.1080/14999013.2019.1652708

van Ravenzwaaij, D., Cassey, P., & Brown, S. D. (2018). A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic Bulletin & Review*, *25*(1), 143–154. https://doi.org/10.3758/s13423-016-1015-8

Vaughn, M. G., DeLisi, M., Gunter, T., Fu, Q., Beaver, K. M., Perron, B. E., & Howard, M. O. (2011). The severe 5%: A latent class analysis of the externalizing behavior spectrum in the united states. *Journal of Criminal Justice*, *39*(1), 75–80. https://doi.org/10.1016/j.jcrimjus.2010.12.001

Vaughn, M. G., Salas-Wright, C. P., DeLisi, M., & Maynard, B. R. (2014). Violence and externalizing behavior among youth in the United States: Is there a severe 5%? *Youth Violence and Juvenile Justice*, *12*(1), 3–21. https://doi.org/10.1177/1541204013478973

Venables, N. C., Foell, J., Yancey, J. R., Beaver, K. M., Iacono, W. G., & Patrick, C. J. (2018). Integrating criminological and mental health perspectives on low self-control: A multi-domain analysis. *Journal of Criminal Justice*, *56*, 2–10. https://doi.org/10.1016/j.jcrimjus.2017.10.004

Venables, N. C., Foell, J., Yancey, J. R., Kane, M. J., Engle, R. W., & Patrick, C. J. (2018). Quantifying inhibitory control as externalizing proneness: A cross-domain model. *Clinical Psychological Science*, *6*(4), 561–580. https://doi.org/10.1177/2167702618757690

Venables, N. C., Hall, J. R., & Patrick, C. J. (2014). Differentiating psychopathy from antisocial personality disorder: A triarchic model perspective. *Psychological Medicine*, *44*(5), 1005–1013. https://doi.org/10.1017/S003329171300161X

Venables, N. C., Hall, J. R., Yancey, J. R., & Patrick, C. J. (2015). Factors of psychopathy and electrocortical response to emotional pictures: Further evidence for a two-process theory. *Journal of Abnormal Psychology*, *124*(2), 319–328. https://doi.org/10.1037/abn0000032

Venables, N. C., & Patrick, C. J. (2012). Validity of the Externalizing Spectrum Inventory in a criminal offender sample: Relations with disinhibitory psychopathology, personality, and psychopathic features. *Psychological Assessment*, *24*(1), 88–100. https://doi.org/10.1037/a0024703

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418–424. https://doi.org/10.1016/j.tics.2008.07.005

Verona, E., & Bresin, K. (2015). Aggression proneness: Transdiagnostic processes involving negative valence and cognitive systems. *International Journal of Psychophysiology*, *98*(2), 321–329. https://doi.org/10.1016/j.ijpsycho.2015.03.008

Vetenskapsrådet. (2017). *Kartläggning av rättspsykiatrisk forskning*. Vetenskapsrådet.

Vogt, B. A. (2016). Midcingulate cortex: Structure, connections, homologies, functions and diseases. *Journal of Chemical Neuroanatomy*, *74*, 28–46. https://doi.org/10.1016/j.jchemneu.2016.01.010

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

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.e1. https://doi.org/10.1016/j.jaac.2013.12.008

Wallinius, M., Delfin, C., Billstedt, E., Nilsson, T., Anckarsäter, H., & Hofvander, B. (2016). Offenders in emerging adulthood: School maladjustment, childhood adversities, and prediction of aggressive antisocial behaviors. *Law and Human Behavior*, *40*(5), 551–563. https://doi.org/10.1037/lhb0000202

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annual Review of Neuroscience*, *30*(1), 31–56. https://doi.org/10.1146/annurev.neuro.30.051606.094334

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p $<$ 0.05". *The American Statistician*, *73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064. https://doi.org/10.3758/s13423-012-0295-x

Whelan, R., Conrod, P. J., Lourdusamy, A., Banaschewski, T., Barker, G. J., Bellgrove, M. A., Büchel, C., Byrne, M., Cummins, T. D. R., Fauth-Bühler, M., Flor, H., Gallinat, J., Heinz, A., Ittermann, B., Mann, K., Martinot, J.-L., Lalor, E. C., Lathrop, M., Loth, E., … The IMAGEN Consortium. (2012). Adolescent impulsivity phenotypes characterized by distinct brain networks. *Nature Neuroscience*, *15*(6), 920–925. https://doi.org/10.1038/nn.3092

Wibbelink, C. J. M., Hoeve, M., Stams, G. J. J. M., & Oort, F. J. (2017). A meta-analysis of the association between mental disorders and juvenile recidivism. *Aggression and Violent Behavior*, *33*, 78–90. https://doi.org/10.1016/j.avb.2017.01.005

Widiger, T. A., & Simonsen, E. (2005). Alternative dimensional models of personality disorder: Finding a common ground. *Journal of Personality Disorders*, *19*(2), 110–130. https://doi.org/10.1521/pedi.19.2.110.62628

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, *62*(2), 71–83. https://doi.org/10.1037/0003-066X.62.2.71

Williams, M. N., Bååth, R. A., & Philipp, M. C. (2017). Using Bayes Factors to test hypotheses in developmental research. *Research in Human Development*, *14*(4), 321–337. https://doi.org/10.1080/15427609.2017.1370964

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*(3), 1135–1146. https://doi.org/10.1016/j.neuroimage.2009.12.028

World Health Organization. (2020). *International statistical classification of diseases and related health problems (ICD-11)*. https://icd.who.int/.

Wu, Y.-C., & Lee, W.-C. (2014). Alternative performance measures for prediction models. *PLOS ONE*, *9*(3), e91249. https://doi.org/10.1371/journal.pone.0091249

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, *136*(5), 740–767. https://doi.org/10.1037/a0020473

Yang, Y., Joshi, A. A., Joshi, S. H., Baker, L. A., Narr, K. L., Raine, A., Thompson, P. M., & Damasio, H. (2012). Genetic and environmental influences on cortical thickness among 14-year-old twins. *Neuroreport*, *23*(12), 702–706. https://doi.org/10.1097/WNR.0b013e328355a62a

Yang, Y., & Raine, A. (2009). Prefrontal structural and functional brain imaging findings in antisocial, violent, and psychopathic individuals: A meta-analysis. *Psychiatry Research: Neuroimaging*, *174*(2), 81–88. https://doi.org/10.1016/j.pscychresns.2009.03.012

Yang, Y., Wang, P., Baker, L. A., Narr, K. L., Joshi, S. H., Hafzalla, G., Raine, A., & Thompson, P. M. (2015). Thicker temporal cortex associates with a developmental trajectory for psychopathic traits in adolescents. *PLOS ONE*, *10*(5), e0127025. https://doi.org/10.1371/journal.pone.0127025

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yildirim, B. O., & Derksen, J. J. L. (2015). Clarifying the heterogeneity in psychopathic samples: Towards a new continuum of primary and secondary psychopathy. *Aggression and Violent Behavior*, *24*, 9–41. https://doi.org/10.1016/j.avb.2015.05.001

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*(1), 117–130. https://doi.org/10.1037/a0014657

Young, S. E., Stallings, M. C., Corley, R. P., Krauter, K. S., & Hewitt, J. K. (2000). Genetic and environmental influences on behavioral disinhibition. *American Journal of Medical Genetics (Neuropsychiatric Genetics)*, *96*, 684–695.

Zhang, F., & Iwaki, S. (2019). Common neural network for different functions: An investigation of proactive and reactive inhibition. *Frontiers in Behavioral Neuroscience*, *13*, 1–13. https://doi.org/10.3389/fnbeh.2019.00124

Zou, Z., Meng, H., Ma, Z., Deng, W., Du, L., Wang, H., Chen, P., & Hu, H. (2013). Executive functioning deficits and childhood trauma in juvenile violent offenders in China. *Psychiatry Research*, *207*(3), 218–224. https://doi.org/10.1016/j.psychres.2012.09.013

Zwets, A. J., Hornsveld, R. H. J., Neumann, C., Muris, P., & van Marle, H. J. C. (2015). The four-factor model of the Psychopathy Checklist-Revised: Validation in a Dutch forensic inpatient sample. *International Journal of Law and Psychiatry*, *39*, 13–22. https://doi.org/10.1016/j.ijlp.2015.01.016