

Self-Regulation in Inmates with and without Antisocial Personality Disorder: Investigating Emotion Regulation, Aggression and Cognitive Inhibitory Control

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Elena Schreiner

aus Speyer

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:

30.04.2020

Dekan:

Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter:

Prof. Dr. Martin Hautzinger

2. Berichterstatter:

Prof. Dr. Rüdiger Wulf

Acknowledgments

An dieser Stelle möchte ich mich bei allen teilnehmenden Probanden und JVAs bedanken, ohne deren Mitwirken das vorliegende Projekt nicht hätte realisiert werden können. Vielen Dank an Andreas Alter aus der JVA München, der mich zum wiederholten Male bei einem meiner Forschungsvorhaben unterstützte. Frau Dr. Larissa Wolkenstein ebnete mir überhaupt erst den Weg in die Wissenschaft und ermöglichte mir die Durchführung meines eigenen Forschungsprojektes. Vielen Dank für deine zahlreichen, hilfreichen Anmerkungen, deinen fortwährenden Optimismus und dein stetes (übersteigertes) Zutrauen in meine Fertigkeiten. Danke auch an den sozialen Druck in der UB und den dortigen, noch nicht prokrastinationsbelasteten, Arbeitsplatz. Situationsauswahl funktioniert. Ein großer Dank gilt zudem meiner Frustrationstoleranz, auf die ich mich immer wieder verlassen konnte. Ich hoffe, sie künftig nicht mehr so zu beanspruchen. Meine Eltern sind klasse – auch das soll an dieser Stelle nicht unerwähnt bleiben. Schön, dass es euch gibt.

Und zu guter Letzt: Mein herzliches Dankeschön an einen ganz besonderen Menschen.

Table of Contents

List of Abbreviations.....	9
Table Directory	10
Table of Figures	11
Abstract	12
Zusammenfassung.....	14
1. General Introduction.....	16
1.1. Aggression and Aggression Theories.....	17
1.2. Self-Regulation and its Many Facets	21
1.2.1. The Umbrella Term Self-Regulation	21
1.2.2. Emotion and Emotion Regulation	21
1.2.3. Executive Functions, Cognitive Control, and Cognitive Inhibitory Control....	24
1.3. Empirical Findings on Self-Regulation in Aggressive and/or Antisocial Individuals.	25
1.3.1. Emotion Regulation in Aggressive and/or Antisocial Individuals	26
1.3.2. Cognitive Control in Aggressive and/or Antisocial Individuals	29
1.4. Open Research Questions	31
1.5. Overall Goals of the Present Work	32
1.6. Structure of the Thesis	33
2. Preliminary Studies	35
2.1. A New Measure – the Cyberball Aggression Task.....	36
2.2. Preliminary Study I: Online Survey.....	38
2.2.1. Methods	38
2.2.2. Results	39
2.2.3. Implications	40
2.3. Preliminary Study II: Testing the Paradigm.....	41
2.3.1. Methods	41
2.3.2. Results	43
2.3.3. Implications	45
3. Main Study, Part I: Similar, Yet Different – Disturbed Emotion Regulation as a Distinctive Feature Among Antisocial as opposed to Non-Antisocial Offenders and Healthy Controls.....	47
3.1. Background	48
3.1.1. Anger Experience, Anger Regulation and (Reactive) Aggressive Behavior....	48
3.1.2. Emotion Regulation – General Abilities and Strategy Use	50

3.1.3. Goals of the Present Work.....	51
3.2. Methods.....	52
3.2.1. Participants	52
3.2.2. Measures	57
3.2.3. Procedure	60
3.2.4. Data Analysis.....	63
3.3. Results.....	64
3.3.1. Anger Experience and Regulation	64
3.3.2. Habitual Emotion Regulation	65
3.3.3. Spontaneous Emotion Regulation: Cyberball Aggression Task.....	68
3.3.4. Predicting Symptom Severity of Antisocial Personality Disorder	74
3.4. Discussion	77
4. Main Study, Part II: Yes, I Can! Antisocial and Non-Antisocial Offenders Show No General Deficit in Cognitive (Inhibitory) Control.....	82
4.1. Background	83
4.2. Methods.....	86
4.2.1. Participants	86
4.2.2. Measures	87
4.2.3. Procedure	89
4.2.4. Data Analysis.....	89
4.3. Results.....	90
4.3.1. Cognitive Inhibitory Control	90
4.3.2. Working Memory and Set-Shifting	92
4.3.3. Trait Impulsivity	92
4.3.4. Anger Experience and Physical Aggression.....	93
4.3.5. Associations between Cognitive Inhibitory Control and Antisocial Symptoms	94
4.4. Discussion	95
5. General Discussion.....	100
5.1. Integration of Results	100
5.1.1. Antisocial Personality Disorder – a Disorder of Habitual Emotion Regulation.....	100
5.1.2. Intact Spontaneous Anger Regulation or Overlooked Deficits due to Methodological Shortcomings?	104

5.1.3. Too Much and Too Little – Miscellaneous Abnormalities in Aggressive Behavior among Offenders with and without Antisocial Personality Disorder	106
5.1.4. Have We Been Overestimating the Importance of Cognitive Inhibitory Control?	110
5.2. Limitations	113
5.2.1. Beyond the Scope of the Current Work.....	114
5.2.2. Sampling Issues	115
5.2.3. Further Methodological Criticism	116
5.3. Implications and Future Perspectives.....	119
5.3.1. Future Research	119
5.3.2. Preliminary Treatment Recommendations	121
5.4. Conclusion.....	123
6. References	126
7. Appendix	143
7.1. Appendix A	143
7.2. Appendix B	145
7.3. Appendix C	146
7.4. Appendix D	148
7.5. Appendix E.....	150
7.6. Appendix F.....	151

List of Abbreviations

ADHD	Attention deficit and hyperactivity disorder
ADHD-SR	ADHD self-rating
AI	Anger induction
ANOVA	Analysis of variance
APD	Antisocial personality disorder
APDs	Individuals with antisocial personality disorder
ASBs	Individuals with antisocial behavior
AQ	Aggression Questionnaire
AR	Anger regulation
AUD	Alcohol use disorder
BIS-15	Short form of the Barratt Impulsiveness Scale
CAT	Cyberball Aggression Task
CERQ	Cognitive Emotion Regulation Questionnaire
CI	Confidence Interval
DERS	Difficulties in Emotion Regulation Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5 th edition
ER	Emotion regulation
HCs	Healthy controls
IC	Cognitive inhibitory control
INCs	Inmate control participants without antisocial personality disorder
MANOVA	Multivariate analysis of variance
M.I.N.I.	Mini International Neuropsychiatric Interview 7.0.2
MWT-B	Multiple Choice Word Fluency Test
PANAS	Positive and Negative Affect Schedule
RT	Reaction time
SAM	Self-Assessment Manikin
SCID-II	Structured Clinical Interview II for DSM-IV
SDS-17	Social Desirability Scale-17
STAXI-2	State-Trait Anger Expression Inventory-2
SUD	Substance use disorder
TMT	Trail Making Test

Table Directory

<i>Table 1.</i> Some of the influential aggression theories in approximate chronological order.....	19
<i>Table 2.</i> Declaration according to § 5 Abs. 2 No. 8 of the PhD regulations of the Faculty of Science	34
<i>Table 3.</i> Demographic information for participants.....	42
<i>Table 4.</i> Wilcoxon signed-rank tests for pairwise comparisons of consequence within anger rounds	45
<i>Table 5.</i> Participants' demographic characteristics and symptom severities	54
<i>Table 6.</i> Detention information for inmates by group.....	55
<i>Table 7.</i> Diagnostic information for inmates by group	55
<i>Table 8.</i> Groups' anger regulation as evident by the State-Trait Anger Expression Inventory-2	64
<i>Table 9.</i> Habitual difficulties in emotion regulation by group	66
<i>Table 10.</i> Habitual emotion regulation strategy use by group	67
<i>Table 11.</i> Mean punishment by credibility and group, depending on condition and consequence	71
<i>Table 12.</i> Hierarchical multiple regression analysis predicting antisocial symptom severity within inmates	76
<i>Table 13.</i> Stroop variables by group	91
<i>Table 14.</i> Reaction times for combinations of current and preceding trial type	91
<i>Table 15.</i> Trail Making Test times by group.....	92
<i>Table 16.</i> Groups' impulsivity as measured by the short form of the Barratt Impulsiveness Scale	93
<i>Table 17.</i> Groups' anger and aggression as measured by the Aggression Questionnaire.....	94
<i>Table 18.</i> Spearman's rank correlation coefficients between cognitive inhibitory control and antisocial symptoms in inmates.....	95

Table of Figures

Figure 1. Sources of emotion dysregulation by time.....	24
Figure 2. Scope of the present work.....	33
Figure 3. Development process of the Cyberball Aggression Task from the preliminary studies to the main study.....	37
Figure 4. Mean PANAS score depending on measuring time.....	39
Figure 5. Mean angry emotions depending on baseline and anger sections.....	40
Figure 6. Mean angry emotions before (pre) and after (post) the Cyberball Aggression Task.....	43
Figure 7. Frequency of punishing behavior among participants in baseline and anger rounds and depending on consequence.....	44
Figure 8. Self-reported lifetime offences by inmate group.....	56
Figure 9. Screenshots of Cyberball.....	59
Figure 10. Sequence of the Cyberball Aggression Task, including the assessment of angry emotions, arousal and emotion regulation strategies.....	62
Figure 11. Self-reported angry emotions and arousal among groups depending on time (pre, post) and credibility (deceived, not deceived).....	69
Figure 12. Participants' punishing behavior.....	72

Abstract

In view of their high propensity for crime and their considerable recidivism rates, understanding the self-regulation of individuals with antisocial personality disorder (APDs) is of great social relevance. Particularly, deficits in emotion regulation (ER) and cognitive inhibitory control (IC) are assumed to contribute to aggressive behavior. However, to date, these aspects of self-regulation are still underexplored in APDs. Therefore, the current thesis aims to identify abnormalities in self-regulation that may underlie the behavioral phenotype of antisocial personality disorder (APD) and that distinguish between inmates with and without APD.

(1) First, two preliminary analyses were conducted to examine the suitability of a newly developed anger induction (AI) and aggression paradigm to be used in the subsequent main study. This instrument assesses a mild form of resource aggression/theft (punishing behavior towards alleged other participants) prior to and during an AI (provocations by alleged other participants). The paradigm was tested among two different male community samples ($N = 324$ in an online survey, $N = 35$ in an experimental study). These studies yielded initial support for the effectiveness of the AI and the sensitivity of the aggression measure.

(2) Part I of the main study comprehensively compared habitual as well as spontaneous ER and, for the first time, aggressive behavior prior to and during an experimental AI between APDs ($n = 31$), inmate control participants without APD (INCs; $n = 33$) and never-incarcerated, healthy controls (HCs; $n = 39$). APDs indicated severe deficits in habitual anger regulation, compared to both, HCs and INCs. However, during the actual regulation attempt in the lab (during the AI), no evidence for a reduced self-reported regulation success or a deviating strategy use was found. Yet, when considering the behavioral measure, resource aggression, abnormalities compared to HCs were revealed in both APDs and INCs – which were however different in nature: APDs showed an increased aggression proneness without the presence of instigating triggers (i.e. prior to the AI), while INCs showed reduced reactive aggression (i.e. during the AI). Regarding overall emotion dysregulation, APDs, but not INCs, reported deficits in comparison to HCs. Particularly APDs' habitual ER strategy use was characterized by an increased use of (generally) maladaptive strategies compared to HCs. Within inmates, deficient ER predicted antisocial symptom severity above and beyond the effects of other variables. Overall, these findings highlight impairments in ER as a distinguishing feature between offenders with and without APD.

(3) Part II of the main study aimed to examine whether APDs suffer from deficient IC performance. Second, potential associations between poor IC and antisocial symptoms were explored. No evidence was found for deficient IC efficiency, disturbed post-conflict

adjustments or impairments in more broad cognitive control abilities – neither for APDs, nor INCs. Within inmates, poor IC was not associated with antisocial symptoms or overall symptom severity. These results challenge the assumption that particularly a poor IC might underlie APDs' symptom domain.

In sum, the current results indicate that impaired ER and elevated aggression proneness are more decisive for APDs' behavioral phenotype than poor IC. The present findings clearly suggest that APD should be recognized as a disorder of ER. Furthermore, divergent mechanisms may underlie APDs' as opposed to INCs' increased aggression. Hence, different treatment options might be suitable for inmates with and without APD. Further implications as well as limitations of the present work – particularly with respect to the measures applied – are discussed.

Zusammenfassung

Da Personen mit antisozialer Persönlichkeitsstörung (APDs) für eine Vielzahl begangener Straftaten verantwortlich sind, ist das Verständnis von für die Störung relevanten Prozessen von außerordentlicher gesellschaftlicher Relevanz. Defizite in der Selbstregulation, speziell der Emotionsregulation (ER) und der kognitiven inhibitorischen Kontrolle (IC), erscheinen aufgrund ihrer Verbindung zu aggressivem Verhalten bei dieser Personengruppe zwar naheliegend, sind aber empirisch keinesfalls gesichert. Die vorliegende Arbeit zielt darauf ab, beeinträchtigte Teilbereiche der Selbstregulation zu identifizieren, die dem Verhaltensphänotyp der antisozialen Persönlichkeitsstörung (APD) zugrunde liegen könnten und die inhaftierte Straftäter mit und ohne APD unterscheiden.

(1) Um die Eignung eines neu entwickelten Ärgerinduktions- (AI) und Aggressions-Paradigmas für die Anwendung in der Hauptstudie zu überprüfen, wurden zunächst zwei Vorstudien durchgeführt. Das untersuchte Instrument erfasst eine schwache Form der Ressourcenaggressivität (monetäres Bestrafungsverhalten gegenüber angeblichen Mitspielern) vor und nach einer AI (Provokationen durch angebliche Mitspieler). Es wurde an zwei unterschiedlichen männlichen Stichproben getestet ($N = 324$ in einer Online-Befragung, $N = 35$ in einer experimentellen Studie). Die Vorstudien ergaben Hinweise auf die Wirksamkeit der AI und des Aggressionsmaßes.

(2) Teil I der Hauptstudie verglich umfassend sowohl die habituelle als auch die spontane ER und erstmalig auch das aggressive Verhalten vor und während einer experimentellen AI zwischen APDs ($n = 31$), inhaftierten Kontrollprobanden ohne APD (INCs; $n = 33$) und niemals inhaftierten gesunden Kontrollen (HCs; $n = 39$). APDs gaben im Vergleich zu HCs und INCs bedeutsame Defizite in der habituellen Ärgerregulation an. Im Gegensatz dazu wurde beim tatsächlichen Regulationsversuch während der AI weder ein verminderter Regulationserfolg noch ein abweichender Strategieeinsatz berichtet. Betrachtet man jedoch das Verhaltensmaß, die Ressourcenaggression, so wurden im Vergleich zu HCs in der Tat Auffälligkeiten sowohl bei APDs als auch bei INCs offenbar, allerdings unterschiedlicher Art: Vor der AI, also ohne klare situative Auslöser, bestrafte APDs am meisten. INCs hingegen zeigten während der AI eine *reduzierte* reaktive Aggressivität. Bezüglich übergeordneter ER-Fertigkeiten berichteten APDs, aber nicht INCs, über Schwierigkeiten im Vergleich zu HCs. Zudem beschrieben insbesondere jene Inhaftierte mit APD gegenüber HCs einen erhöhten habituellen Einsatz (überwiegend) maladaptiver ER-Strategien. Innerhalb der Inhaftierten konnte eine defizitäre ER den Schweregrad der antisozialen Symptomatik voraussagen, selbst wenn für andere Variablen kontrolliert wurde. Insgesamt betrachtet betonen diese Ergebnisse,

dass Beeinträchtigungen in der ER ein Unterscheidungsmerkmal zwischen Inhaftierten mit und ohne APD darstellen.

(3) Teil II der Hauptstudie überprüfte, ob APDs unter einer defizitären IC leiden und ob eine verminderte IC mit antisozialen Symptomen assoziiert ist. Weder bei APDs noch INCs fanden sich im Vergleich zu HCs Belege für eine beeinträchtigte IC-Effizienz, veränderte Konflikthanpassungen oder Defizite in allgemeineren kognitiven Kontrollprozessen. Innerhalb der Inhaftierten war eine schlechtere IC nicht mit spezifischen antisozialen Symptomen oder der allgemeinen Symptomschwere verbunden. Diese Resultate stellen die Annahme infrage, dass speziell Defizite in der IC der Symptomedäne der APD zugrunde liegen.

Die vorliegenden Ergebnisse deuten darauf hin, dass eine gestörte ER und eine verstärkte Aggressionsneigung für den Verhaltensphänotyp der APD entscheidender sind als eine schlechte IC. Die aktuellen Befunde unterstützen die Betrachtung der APD als eine Störung der ER. Darüber hinaus könnten unterschiedliche Mechanismen APDs' und INCs' Aggression unterliegen. Ausgehend von ihren Auffälligkeiten scheinen für Inhaftierte mit und ohne APD unterschiedliche Behandlungsmaßnahmen indiziert. Weitere Implikationen sowie Grenzen der vorliegenden Arbeit – insbesondere hinsichtlich der angewandten Methoden – werden diskutiert.

1. General Introduction

While at the individual experience level crime is a rare event, it is a mass phenomenon at the societal level. In 2018, more than 5 million offences were registered in Germany (Bundeskriminalamt, 2019). Quantifying the damage caused by crime is difficult, as the victims' and other (in)directly affected persons' individual impairments and suffering can hardly be expressed by a mere number. However, if one tries, it can be noted that the annual net expenditure for German prisons alone amounts to approximately € 2.5 billion (Statistisches Bundesamt, 2014). The British government estimated the annual English and Welsh economic costs of crime to be £ 36.2 billion (Home Office, 2005). Undoubtedly, high intangible but also tangible costs are associated with crime.

The most serious and/or repeat offences are usually sentenced with imprisonment. Currently¹, in Germany there are about 51.000 persons incarcerated in pre-trial detention or criminal custody (closed prisons), the vast majority of them men (95%; Statistisches Bundesamt, 2019a). It is assumed that about every second male inmate fulfills the psychiatric diagnosis of antisocial personality disorder (APD; Fazel & Danesh, 2002) – although admittedly there is a wide variation in prevalence estimates between studies (see Moran, 1999; Rotter, Way, Steinbacher, Sawyer, & Smith, 2002). Not only do a large proportion of inmates exhibit APD, but APD is also a predictor of recidivism (Katsiyannis, Whitford, Zhang, & Gage, 2018) and is associated with increased reconvictions and reincarcerations (Shepherd, Campbell, & Ogloff, 2016). Even if one does not look exclusively at individuals with APD (APDs), but at the entire group of male adult offenders², the recidivism rates are already striking: about half of the released male prisoners are reconvicted within three years, almost a quarter even return to prison within this time period (Jehle, Albrecht, Hohmann-Fricke, & Tetel, 2013). Therefore, understanding processes relevant to antisocial behavior, and especially APD, is of extraordinary social relevance.

To gain such a deeper understanding of APD, it is essential to compare APDs with two different groups of people: First, it is important to learn about abnormalities in APDs as compared to never-incarcerated healthy controls (HCs) in order to obtain information about APDs' impairments. Second, it is important to assess whether these deficits are indeed specific to APDs. Hence, to determine the contribution of the psychiatric diagnosis to APDs' antisocial

¹ The cutoff date was September 30, 2019 (Statistisches Bundesamt, 2019a).

² Since only here, reliable figures are available.

behavior as opposed to a mere “criminal lifestyle”, the question arises, as to what differences, but also similarities, exist between inmates with and without APD.

In view of the goal of legal enforcement (§ 2 StVollzG) and the above-mentioned recidivism rates, it seems clear that there is a need for constant revision and improvement of the existing treatment offered in prisons. However, in order to be able to successfully intervene not only therapeutically, but also preventively, contributory factors of offending must be identified. Hence, investigating mechanisms that distinguish incarcerated offenders with (and without) APD from non-offenders might be an important first step to form such hypotheses regarding underlying causes of criminal behavior. The current thesis addresses these issues by exemplarily considering two aspects of self-regulation: emotion regulation (ER) and cognitive control.

1.1. Aggression and Aggression Theories

The diagnosis of APD relies predominantly on behavioral constructs, such as unlawful behavior, aggression and impulsivity (see Diagnostic and Statistical Manual of Mental Disorders, 5th edition; DSM-5, American Psychiatric Association, 2013). Unlike other personality disorders, affect and inner experiences are almost neglected (Ogloff, 2006). Therefore, the diagnosis of APD is often criticized (Baliouis, Duggan, McCarthy, Huband, & Völlm, 2019). There is a discussion whether or not there are other features of APD that are overlooked by diagnostic criteria, which however contribute to the disorders’ behavioral phenotype (Sedgwick et al., 2017). The perhaps most significant behavioral phenotype is aggression³.

Aggression is a “behavior directed toward the goal of harming or injuring another living being, who is motivated to avoid such treatment” (Baron & Richardson, 1994; p. 7). While all forms of aggression intend to harm, it is important to note that physical injury or even violence is not necessarily involved. For example, Parrott and Giancola (2007) distinguish different subtypes besides physical aggression, among them verbal forms of aggression and resource aggression. Furthermore, the ultimate goals of aggressive behavior vary: a common classification distinguishes reactive from proactive aggression (e.g. Berkowitz, 1989). Reactive aggression, which is also called impulsive, hostile, or affective aggression, is a reaction of

³ While aggression is among the main symptoms of APD within the “classic” APD criteria in Section II of the DSM-5, this, however, does not apply to the Alternative DSM-5 Model for Personality Disorders in Section III (American Psychiatric Association, 2013). Here, the relevance of aggression as a diagnostic criterion has been somewhat devalued by solely assigning it to the criterion “callousness”. Nonetheless, diagnoses are still assigned based on Section II.

perceived frustration, typically driven by anger, with harming as the ultimate goal (Anderson & Bushman, 2002; Berkowitz, 1989; Ritter & Eslea, 2005). By contrast, proactive aggression is less emotional. It includes an (additional) goal other than harming. Usually, another person is harmed in order to reach some other goal (e.g. Anderson & Bushman, 2002). Thus, proactive aggression is also called instrumental aggression. However, there are also mixed forms (Anderson & Bushman, 2002).

The DSM-5 does not clearly specify which forms of aggressive behavior APDs are prone to (reactive vs. proactive vs. hybrid). However, other diagnostic criteria such as impulsivity and hostility indirectly suggest that APDs show mainly reactive aggression (American Psychiatric Association, 2013). Given the conceptual link between APD and aggression, as well as the lack of research specifically addressing the underlying processes of APD, it seems useful to take a look at theories of aggression, particularly those that claim to explain reactive forms. These theories may point to abnormalities in APD that are not nosologically specified in the DSM-5. Since there is a vast number of theories which approach the topic from (partly) very different perspectives, a selection of some of the more popular theories is summarized in Table 1.

Table 1. Some of the influential aggression theories in approximate chronological order

Theory	Basic Assumption
Frustration-Aggression Theory (Dollard, Miller, Doob, Mowrer, & Sears, 1939)	Every aggression is preceded by frustration, with frustration being a denial of goals (as opposed to a mere non-achievement).
Learning Theories (e.g. Burgess & Akers, 1966)	These theories basically state that aggression is learned by reinforcement (operant conditioning) and – to a lesser extent – by classic conditioning. Later on, these theories were further developed by emphasizing observational learning processes (social learning).
Excitation-Transfer Theory (e.g. Zillmann & Bryant, 1974)	This is basically a drive theory. Assumes that arousal from a preceding event will amplify the excitatory response to a subsequent stimulus. When the residual excitation is misattributed to anger, the likelihood of aggressive behavior is increased.
The Information-Processing Theories (e.g. hostile attribution bias; Dodge, 1980)	Abnormalities in information processing increase the likelihood for aggressive behavior. For example, the hostile attribution bias describes the phenomenon of inferring hostile intent in ambiguous situations.
Cognitive Neoassociation Theory (Berkowitz, 1989)	Adaptation of the original Frustration-Aggression Theory. Postulates emotional networks, which are interconnected. When a concept is activated in a specific situation, associated concepts are also activated – an inner chain reaction unfolds. An aversive event leads to negative affect, which stimulates various thoughts/memories and behavioral response tendencies. Both, aggressive and fearful concepts are activated. The strength of these associations depend on previous experiences, and genetic factors, among others. Hence, an aversive event might or might not lead to aggression (fight or flight).
A General Theory of Crime (Gottfredson & Hirschi, 1990)	Is a self-control theory. Individuals who lack self-control in the face of temptation (i.e. top-down aspects of self-regulation) are more likely to commit deviant and/or criminal behavior.
Developmental Taxonomy of Antisocial Behavior (Moffitt, 1993; Moffitt & Caspi, 2001)	Distinguishes adolescence-limited from early-onset life-course-persistent antisocial behavior, while only the latter is assumed to be pathological. Life-course-persistent antisocial behavior originates “when the difficult behavior of a high-risk young child is exacerbated by a high-risk social environment” (Moffitt & Caspi, 2001, p. 355). High-risk children are characterized by inherited or acquired neuropsychological deficits (i.e. executive functions), which are relatively stable across the life span.

(continued)

Table 1. Some of the influential aggression theories in approximate chronological order (continued)

Theory	Basic Assumption
General Aggression Model (Anderson & Bushman, 2002)	A prominent metatheory, emphasizing mediating processes between inputs of the person (e.g. trait anger, hostile attribution bias, biological factors) and situation (e.g. provocations, incentives) and aggressive or non-aggressive outcomes. Inputs create a present internal state, comprised of cognition, emotion and arousal. This internal state influences the output of a (non-) aggressive episode. The individual's action can either be automatic or more controlled, depending on available resources and the importance as well as the satisfaction with the (immediate appraisal) outcome. Explains reactive, proactive and hybrid forms of aggression.
I³ Model/Perfect Storm Theory (Finkel, 2014; Finkel & Hall, 2018)	Another integrative framework categorizing risk factors. Simplified, the I ³ Model ("I-cubed Model") focuses on three processes: <u>i</u> nstigating triggers (<i>immediate</i> environmental stimuli that normatively increase aggressive urges; e.g. provocation), <u>i</u> mpelling factors (factors that increase the likelihood to react to instigating triggers; e.g. trait anger, retaliation tendencies, presence of a weapon), and <u>i</u> nhibiting factors (factors that reduce the probability of an aggressive urge being translated into aggressive behavior; e.g. high self-regulatory resources, relationship commitment, no alcohol intoxication). The Perfect Storm Theory emphasizes interactive effects and assumes that the likelihood of aggressive behavior is highest when instigation and impellance are strong, while inhibition is low.

Note. This is a selective choice of existing aggression theories. The selection is subjective, but was mainly based on DeWall, Anderson, and Bushman (2012). Theories on psychopathy (e.g. low-fear hypotheses in primary psychopaths) have been intentionally left out, since the current work primarily deals with the psychiatric disorder antisocial personality disorder.

As depicted in Table 1, more recent (meta-)theories regarding aggression and antisocial behavior particularly stress the importance of self-regulation, especially for reactive aggression. While for example the General Aggression Model (Anderson & Bushman, 2002) highlights ER, among other things, Moffitt's (1993) developmental taxonomy of antisocial behavior particularly underlines deficits in executive functions. The General Theory of Crime (Gottfredson & Hirschi, 1990) as well as the I³ Model (Finkel, 2014) especially emphasize deficits in inhibition. Due to the association between APD and aggression, the question arises as to whether and, if so, how APDs are empirically affected by such impairments in self-regulation.

1.2. Self-Regulation and its Many Facets

Integration of empirical findings regarding self-regulation is often hampered by the fact that no consistent terminology is used (c.f. Nigg, 2017). Sometimes, the underlying constructs of an investigation are not further specified at all. Using the same terminology between studies suggests a comparability of findings that does not maintain a closer scrutiny, which, among others, limits the validity of later meta-analyses (cf. Vazsonyi, Mikuška, & Kelley, 2017 for a meta-analysis on self-control, that does not define self-control). Therefore, in this thesis, key constructs of self-regulation shall be first defined before briefly discussing them in relation to APD.

1.2.1. The Umbrella Term Self-Regulation

According to Nigg's (2017) thorough conceptual framework, self-regulation refers to the *adaptive* and/or goal-directed regulation of one's own actions (i.e. behavior) and internal states (i.e. primarily cognition and emotion) by oneself (i.e. intrinsic). It is a domain-general, relatively broad construct, also including physiological systems (e.g. allostatic mechanisms), which are, however, no subject to the present work. Self-regulation can be managed by both, bottom-up and top-down mechanisms. Reflexively turning the gaze away from a glaring stimulus is equally self-regulation (bottom-up) as is the conscious shift of attention towards a goal-relevant stimulus (top-down).

It seems intuitively plausible that a well-functioning self-regulation (i.e. adaptive, goal-consistent actions) protects against (reactive) aggressive behavior or, in other words, that (reactive) aggression reflects deficits in self-regulation. However, since self-regulation represents a broad construct and aggression theories emphasize various aspects of it, it seems helpful to take a closer look at single components. One important aspect of self-regulation is ER.

1.2.2. Emotion and Emotion Regulation

Despite the increasing body of research in the field of emotion (regulation), there is no precise scientific definition on what an emotion actually is (Izard, 2010). There is, however, a basic agreement that emotions comprise multiple components (e.g. Koole, 2009). According to this multi-aspect approach, an emotional reaction consists of neurobiological processes (central and vegetative nervous system), perceptual-cognitive processes, behavioral response tendencies and a subjective experience – the “feeling” (Izard, 2010). Emotions are associated

with various functions, among them informational content (emotions tell us about the significance of events), communicative value (emotions tell others something about us and have a social and relational function) and motivational components (emotions drive us to act) (e.g. Izard, 2010). Due to emotional experiences in the past an individual learns to choose his or her own behavior according to the anticipated emotional outcomes (Baumeister, Vohs, DeWall, & Liqing, 2007). Hence, current and anticipated emotions are the basis for goal-oriented behavior and determine our entire life (McMurrin, 2011).

The “processes, by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions” is called ER (Gross, 1998, p. 275). ER is goal-directed and targets the emotion-generative process (Gross, Sheppes, & Urry, 2011). Goals are either hedonic (i.e. intensifying or prolonging pleasant emotions, while decreasing unpleasant emotions) or instrumental (i.e. achieving one’s long-term goals; Gross, 2013, 2015). Like self-regulation, ER can be either explicit (i.e. effortful and conscious) or implicit (i.e. effortless and unconscious) – or something in between (see Gross, 2013). However, implicit ER is beyond the scope of the current work. When subsequently referring to ER, this terminology primarily corresponds to intrinsic, explicit applications.

Various accounts have been made trying to categorize the ways emotions can be regulated. Gross’ (1998) process model of ER distinguishes five families of ER strategies, which influence the emotion-generative process at different points in its temporal sequence: Given that emotion generation begins with a personally relevant internal or external situation, the first opportunity for ER is *situation selection*. Accordingly, it is also possible to modify the present situation (*situation modification*). However, it has to be noted that it is the external, physical environment that is changed by these two strategies (Gross, 2015). Only if the situation is attended to, an appraisal of the situation occurs. Hence, *attentional deployment* and *cognitive change* are the next potential points of action, with cognitive change addressing the “internal” environments, i.e. thoughts and appraisals (Gross, 2015). The appraisal in turn determines the emotional response pattern (physiological, behavioral and experiential). Therefore, *response modulation* forms the last ER possibility in the emotion-generative process before the next iteration of the sequence begins (Gross, 1998). Despite its popularity, the process model is not an empirically validated consensual classification (Koole, 2009). One of the main points of criticism refers to the assumption of the necessity of cognitive appraisal for emotion generation (Koole, 2009). Nevertheless, the process model of ER is still the most prominent framework classifying ER strategies (Webb, Miles, & Sheeran, 2012).

An adaptive ER enables the individual to behave in accordance with his or her long-term goals and typically comprises the maintenance of his or her social functioning and well-being. Characteristics of a successful ER are emotional awareness and understanding, but also emotional acceptance (Gratz & Roemer, 2004). These basic abilities serve, so to speak, the analysis of the actual state which is the prerequisite for working towards the target state, the ER goal. ER strategies, in turn, are the tools, by which the ER goal may be achieved. Here, access to a variety of strategies, flexibility in the use and adequate context-sensitive application of ER strategies are important skills for a successful ER (Gratz & Roemer, 2004). It is important to note that the adaptivity/maladaptivity of a particular ER strategy is non-deterministic. Although, for example, problem solving (or reappraisal) are often referred to as “the” adaptive ER strategies, they are less helpful in situations which cannot (or can) be controlled (Barnow, Reinelt, & Sauer, 2016). Hence, context and ER goals must be taken into account when evaluating the (mal)adaptivity of an ER strategy in a given situation (Gross, 2013; McRae & Gross, 2020).

As indirectly outlined above, emotions are generally adaptive (e.g. McMurrin, 2011). However, emotions can be harmful, too (Gross, 2015). Typically, emotions are problematic if they are of inappropriate intensity, duration, frequency, and/or type (for the particular situation; Gross & Jazaieri, 2014). Different kinds of emotion dysregulation can lead to such problematic emotional states. Underregulation, for example, refers to the failure to control the emotional response and may result in a lack of goal-oriented behavior and/or impulsive action (Robertson, Daffern, & Bucks, 2012). Overregulation, by contrast, means the (excessive) use of ER strategies like emotional avoidance and/or expressive suppression in order to prevent the emotional experience as much as possible (Robertson et al., 2012). Paradoxically, this can lead to the opposite, i.e. increased unpleasant emotions (Robertson et al., 2012). It is important to note that not only ER strategy choice and implementation are crucial for emotion dysregulation, but also deficits in the basic abilities mentioned above (awareness, understanding acceptance). Figure 1 shows frequent sources of emotion dysregulation in chronological order (cf. Gross & Jazaieri, 2014).

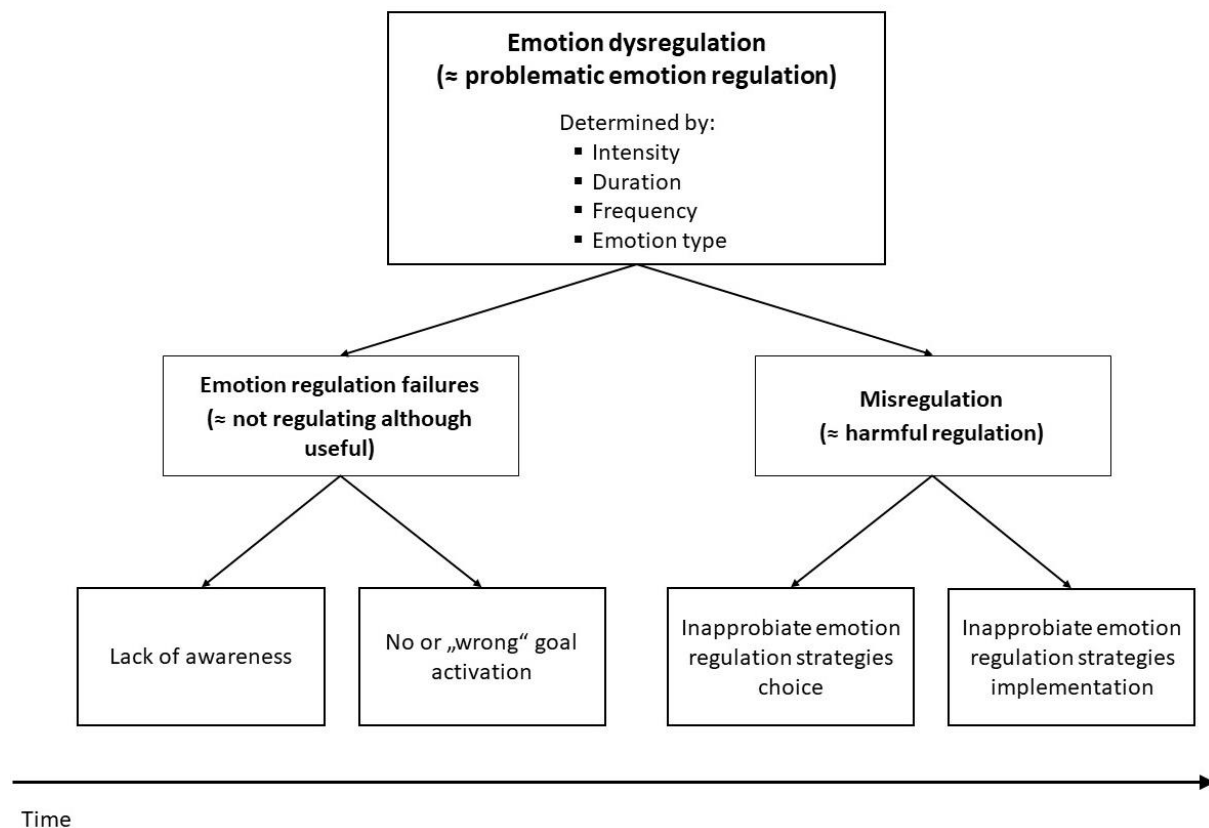


Figure 1. Sources of emotion dysregulation by time.
Illustration created on the basis of Gross and Jazaieri (2014).

Given that almost everything we do or do not do is because of (anticipated) emotional outcomes (see Baumeister et al., 2007), it may, on first glance, seem difficult to separate self-regulation from ER. Indeed, intrinsic ER that enables an adaptive change of state is an important aspect of self-regulation. However, being comforted by a friend is extrinsic ER and is therefore no component of self-regulation (while seeking out a friend indeed is – at least up to this point). An attempt to separate intrinsic ER and self-regulation is based on their primary intention: If self-regulation is *primary* regulation of emotion, then both concepts are equivalent. However, if the regulation of emotion is only secondary, then it is self-regulation of behavior or cognition (Nigg, 2017). One commonality between top-down self-regulation of behavior and cognition, but also explicit intrinsic ER, is their reliance on cognitive processes (cf. McRae, Jacobs, Ray, John, & Gross, 2012; Miyake & Friedman, 2012). An umbrella term for such skills is executive functions.

1.2.3. Executive Functions, Cognitive Control, and Cognitive Inhibitory Control

Executive functions consist of effortful top-down cognitive processes enabling the individual in goal-directed behavior. While executive functions are indispensable for the

adaptive regulation of oneself, they can be used for other purposes as well (e.g. using executive functions to solve a math problem, if ER is not the primary goal here). Executive functions include higher-order skills like reasoning and planning (Diamond, 2013), but also lower-level functions (Nigg, 2017), which are typically divided into three core abilities: (1) response inhibition/interference control, (2) (updating) working memory, and (3) set shifting (Friedman & Miyake, 2017). In the current thesis, these low-level executive functions are understood as cognitive control abilities (see Nigg, 2017). Cognitive control is needed to maintain goal-oriented behavior in situations with directly competing cognitive and behavioral demands, i.e. in situations with *immediate or short-term* conflict (Zeier, Baskin-Sommers, Hiatt Racer, & Newman, 2012). Cognitive control embraces the basic top-down operations that are necessary for more complex cognition (Nigg, 2017). These higher-level executive functions like planning, not only require but go beyond cognitive control, typically cover a larger period of time and manage to resolve *future* conflicts or goals. Therefore, in this thesis, the terms executive functions and cognitive control are not used interchangeably (as, e.g., in Diamond, 2013). The focus of the current work lies on the more basic mechanisms, i.e. cognitive control. Here, particularly response inhibition/interference control is of interest, as it is hypothesized to be the core ability among executive functions (Miyake & Friedman, 2012). Since the term “response inhibition” may be a little misleading, as it might suggest the interruption of an already initiated motor response, it shall be noted, that this work focuses on *cognitive* inhibition. This ability allows the individual to attend to and select a goal-relevant stimulus (component) despite the interference due to a more salient stimulus (component) for which there is a tendency to respond to (Krakowski et al., 2015). Therefore, the construct under investigation shall be labeled cognitive inhibitory control (IC).

Taken together, intrinsic ER is an aspect of self-regulation, while cognitive control mechanisms like IC are typically necessary for both, ER (Ochsner & Gross, 2005) and top-down self-regulation (Nigg, 2017).

1.3. Empirical Findings on Self-Regulation in Aggressive and/or Antisocial Individuals

As depicted above (see Table 1), prominent theories assume impairments in self-regulation such as emotion dysregulation and poor IC to underlie (reactive) aggression and/or antisocial behavior. However, empirical findings on self-regulation within thoroughly diagnosed adult offenders with APD are scarce. This is surprising, considering that this group is responsible for a large number of offences (including severe aggressive acts) and that they have an increased risk of reoffending (Shepherd et al., 2016). Due to this unfortunate lack of

research, findings on the relationship between aspects of self-regulation on the one side and aggression and antisocial behavior on the other, shall also be presented below.

1.3.1. Emotion Regulation in Aggressive and/or Antisocial Individuals

Anger. When looking at specific emotions, the above mentioned aggression theories (see Table 1) particularly highlight the importance of anger for the occurrence of aggressive behavior. And indeed, there is compelling empirical evidence linking anger with physical (reactive) aggression in different samples, including community boys (Sullivan, Helms, Kliewer, & Goodman, 2010), psychiatric patients (Skeem et al., 2006), forensic patients (Doyle & Dolan, 2006) as well as adolescent (Miller, Vachon, & Aalsma, 2012) and adult offenders (e.g. Graña, Redondo, Muñoz-Rivas, & Cantos, 2014). There are also indications for increased trait anger among offender populations (Tonnaer, Siep, van Zutphen, Arntz, & Cima, 2017). Accordingly, anger is referred to as a “driver of violent offending” (Novaco, 2011, p. 72). However, surprisingly few studies explicitly included inmates diagnosed with APD, and to my knowledge, none included both a healthy and an inmate control group. As a consequence, it is unclear, whether increased anger is a general phenomenon among offenders or whether it affects exclusively (or especially) those with APD (more information is given in chapter 3.1.1). Moreover, it has to be noted that it is probably not anger per se that is important for the exertion of aggression but the regulation of angry emotions (Robertson et al., 2012). The dysregulation of anger seems to increase the willingness to engage in aggressive behavior by focusing attention on annoyance-related information, activating aggressive scripts, and biasing interpretation of current events (Robertson et al., 2012).

Anger regulation. With respect to anger regulation (AR), Robertson et al. (2012) suggest that aggressive behavior occurs not only due to underregulation but may also be a result of overregulation. Overregulation may deplete cognitive resources and thus reduce decision making processes, it may prevent the solution of problems and thus extend existing stressors and, as a consequence, paradoxically increase negative affect and arousal, increasing the risk for aggression (Robertson et al., 2012). Indeed, in a study among violent offenders, not only an unregulated AR subtype was identified, but also an overregulated and a regulated one (Low & Day, 2015). Although not entirely consistent with Robertson et al.’s (2012) presumptions, the study by Low and Day (2015) indicates a heterogeneity in AR among inmates and different AR mechanisms that contribute to their aggressive behavior. This is where the question emerges whether inmates with and without APD can be assigned to different regulation types. While it seems reasonable to assume that APDs belong to the unregulated offender type (i.e. increased

trait anger and anger expression, but reduced anger control), there is a lack of empirical evidence – which is outlined in more detail in chapter 3.1.1. To the best of my knowledge, there are again no studies thus far that compare inmates with APD with both, a never-incarcerated HC group (to determine normal or abnormal regulation patterns) and an offender control group (to get an understanding on the influence of the psychiatric disorder). However, in order to be able to tailor interventions, such results would be of great relevance.

Emotion regulation beyond anger regulation. It is assumed that not only anger but also other unpleasant feelings that the individual associates with personal danger and vulnerability can contribute to aggressive behavior (Donahue, Goranson, McClure, & Van Male, 2014; Robertson et al., 2012). Still, there are even fewer studies investigating APDs' ER beyond AR. As outlined above (chapter 1.2.2), basic skills such as emotional awareness and acceptance, but also the (context-sensitive) use of ER strategies, contribute to a functional ER. However, as will be outlined in chapter 3.1, there are hardly any studies examining these areas in offenders, and, to my knowledge, none in inmates with APD.

Furthermore, almost all ER research in offender populations focuses on *habitual* ER in participants' usual environment. It seems questionable whether these results are transferable to an actual regulation attempt in the lab, i.e. to *spontaneous* ER. Besides, emotion dysregulation as assessed with self-reports on basic skills and frequency of strategy use are not necessarily transferable to ER success (Gruber, Harvey, & Gross, 2012; McRae, 2013). Hence, it is of interest, how APDs' regulation pattern and their emotional reactivity (i.e. their ER success) turn out when spontaneously regulating affect during a standardized experimental anger induction (AI). Similar considerations apply to research on aggression: Many of the studies assessing aggression rely on self-reports and/or information from the Federal Central Register (e.g. Kolla, Meyer, Bagby, & Brijmohan, 2017). While the former is prone to social desirability and memory bias, the latter displays only the “bright” field of crime, which probably differs greatly from reality. Therefore, in order to identify abnormalities in aggressive behavior in offenders with and without APD, it would be reasonable to also *observe* actual aggressive behavior in the lab.

Spontaneous anger reactivity and aggression. There is, to my knowledge, only one study that conducted an AI among APDs, but none that conducted both, an AI and an aggression paradigm. Lobbestael, Arntz, Cima, and Chakhssi (2009) used a stress-interview method to elicit angry emotions among APDs. During this method, participants are asked to recall and describe a biographical, anger-evoking situation from their past. The interviewer takes an active role and asks non-standardized questions with the goal of reactivating the participants'

emotional experience (Lobbestael, Arntz, & Wiers, 2008). Indeed, this interview method proved effective, as evident by a significant increase in angry emotions across participants (Lobbestael et al., 2009). However, and contrary to expectations, APDs did not report increased anger reactivity as compared to the overall group, consisting of APDs, patients with other personality disorders and non-patient controls (Lobbestael et al., 2009). Though when looking at the descriptive data of the study (Lobbestael et al., 2009), it seems likely that APDs indeed reported increased reactivity as compared to non-patient controls. This effect was probably overlooked by merging the non-patient controls with the personality disordered patients into a very heterogeneous overall group. But even if these differences were significant, there would still be another difficulty in interpreting the results: Due to its unstandardized character, the stress-interview method lacks internal validity. So, if observing group differences in emotional change scores it cannot be ruled out that one group simply experienced less severe conflicts in the past instead of being “better” in AR. Furthermore, in light of the specific instructions and method used, it seems questionable whether an increase in anger was actually a "poor performance" (e.g. not sufficiently controlling the emotional experience) or rather a "good performance" (e.g. access to own emotions, awareness, emotional acceptance).

In view of the lack of studies on behavioral aggression in APDs it is obvious to extend the field of research to studies focusing on AIs and aggression paradigms among individuals with antisocial behavior (ASBs). But even then, barely any findings are found. The few existing AI studies show a great heterogeneity in sample characteristics and often lack an adequate psychiatric assessment, which makes it difficult to interpret findings at all. Samples of these studies range from mixed gender students and community participants with and without high psychopathic tendencies (Yoon & Knight, 2015), male undergraduate students without a criminal record but high psychopathic tendencies (Osumi et al., 2012), domestically violent men from the community (Babcock, Green, Webb, & Yerington, 2005; Barbour, Eckhardt, Davison, & Kassinove, 1998), male violent offenders from a forensic psychiatric institution (Tonnaer et al., 2017) to samples of male forensic and penitentiary offenders (Tonnaer, Cima, & Arntz, 2019). To induce angry emotions, frustrating tasks (ultimatum and dictator game; Osumi et al., 2012), film clips (Yoon & Knight, 2015), conflict discussions (Babcock et al., 2005), audiotaped vignettes⁴ (the Articulated Thoughts in Simulated Situations paradigm; Babcock et al., 2005; Barbour et al., 1998; Tonnaer et al., 2019; Tonnaer et al., 2017) and harassing feedback by a mannequin (the harassing body opponent bag; Tonnaer et al., 2019)

⁴ Participants were instructed to listen to audiotaped (anger) situations and to imagine they were in the situation themselves (see Tonnaer et al., 2019).

were used. However, some methodological issues need to be considered: While film clips do not seem to be a suitable method for inducing angry feelings (Rottenberg, Ray, & Gross, 2007), the conflict discussion paradigm lacks internal validity and did not prove to work sufficiently (Babcock et al., 2005). Although the Articulated Thoughts in Simulated Situation task has been associated with increases in angry emotions in Barbour et al. (1998) and Babcock et al. (2005), it has to be considered that this task necessarily requires participants' imagination ability. Yet, it cannot be assured that all participants are equally capable of imagination. Furthermore, the articulation of thoughts and feelings in a face-to-face context with the investigator can be biased by various factors. Besides, what people report they *would* do if they were in a situation like that, is not necessarily what they actually do (for criticism on mood inductions that rely on autobiographical memory see also Tang & Schmeichel, 2014). Furthermore, only Osumi et al. (2012) and Tonnaer et al. (2019) not only conducted an "AI", but additionally assessed aggressive behavior. While Osumi et al.'s (2012) sample (Japanese students) is only of minor interest for the present work, the suitability of the body opponent bag (Tonnaer et al., 2019) for measuring aggression has to be put into question: in view of the definition of aggression (i.e. the intent to harm another person who is believed to be motivated to avoid that behavior), the dependent variable of the body opponent bag task (i.e. the force of each punch when instructed to punch) does not seem to reflect aggressive behavior, since it is about punching a mannequin.

Taken together, there is a lack of studies investigating APDs' habitual AR and their ER beyond anger. Furthermore, there are, to my knowledge, no studies to date that examine both, APDs' spontaneous AR and their (reactive) aggressive behavior. Measures used within different samples of ASBs show significant weaknesses. Hence, to address this research question, we are clearly in need of a more internally valid AI method that is able to induce a significant amount of anger, which is however simultaneously feasible in the specific setting of a prison.

1.3.2. Cognitive Control in Aggressive and/or Antisocial Individuals

As outlined above, cognitive control performance, and particularly IC, contrasts with impulsiveness and is a prerequisite for tolerating frustration, resisting aggressive urges and thus preventing punishment (Zeier et al., 2012) – symptom areas by which APD is defined (American Psychiatric Association, 2013). Hence, a deficit in IC seems plausible in offenders with APD. In line with this, it has recently been shown that IC protects adolescents with deviant peers against delinquency (Hinnant & Forman-Alberti, 2019). However, research that specifically investigated IC in adult APDs yielded mixed results and either contained no HC

group or no inmate control group (Roszyk, Izdebska, & Peichert, 2013; Schiffer et al., 2014; Zeier et al., 2012; these studies are discussed in more detail in chapter 4.1).

When broadening from IC to superordinate executive functioning, there are two meta-analyses that are often considered as evidence for deficient executive functions in ASBs and APDs (Morgan & Lilienfeld, 2000; Ogilvie, Stewart, Chan, & Shum, 2011). Ogilvie et al. (2011), who extended the earlier meta-analysis of Morgan and Lilienfeld (2000), found intermediate deficits in executive functioning among ASBs compared to control subjects (weighted effect size $d = .44$). However, it has to be considered, that the operationalization of antisocial behavior included the diagnosis of APD, but also psychopathic personality traits, conduct disorder, as well as crime and delinquency. Hence, the sample assignments were partly based on psychiatric diagnoses, partly based on more general legal or even social norms. Moreover, samples consisting of male and female adults, adolescents and children were merged. Accordingly, it was revealed that ASBs did not originate from a single underlying population but represented a heterogeneous group of people with high variability (Morgan & Lilienfeld, 2000; Ogilvie et al., 2011). By merging such heterogeneous groups (into ASBs), there is a risk that large differences between more severely impaired individuals and unimpaired individuals (possibly APDs vs. HCs) may appear small or may be overlooked because less severely impaired individuals (possibly offenders without APD) are included into the antisocial behavior group. However, when looking at specific subgroups of ASBs, results were the opposite, as one might expect: Criminals (i.e. individuals with a presumably increased level of functioning compared to APDs) were quite severely impaired in executive functioning ($d = .61$), while this was not the case for the more homogenous group of APDs, for whom the deficits were not even clinically relevant ($d = .19$) (Ogilvie et al., 2011). Though these differences in effect sizes may have reflected mere control group effects: That is, APDs' performances were more frequently compared to those of other inmates' and patients', while ASBs' performances were usually compared to HCs' (Ogilvie et al., 2011). Hence, APDs' impairments in executive functions might have been underestimated. Moreover, it should be considered that there was not only heterogeneity in samples and control groups, but also between the different measures of executive functioning (Morgan & Lilienfeld, 2000; Ogilvie et al., 2011). For ASBs, results varied from $d = -.13$ (for $n = 1$ study conducting a two-back test assessing working memory) to $d = .38$ (for $n = 1$ study using the Eriksen Flanker Task, and thus measuring IC; Eriksen & Eriksen, 1974) (Ogilvie et al., 2011). This is not surprising, given that each task requires (slightly) different underlying skills. Merging results across tasks is therefore associated with a huge loss of information, as the origin of the deficits remains unclear.

Although both meta-analyses indeed suggest deficits in executive functioning among APDs (Morgan & Lilienfeld, 2000; Ogilvie et al., 2011), it cannot be inferred which specific abilities are impaired (e.g. IC). Furthermore, no conclusion can be drawn as to whether APDs are more severely impaired than inmate control participants without APD (INCs). However, since the results varied considerably between the specific antisocial behavior groups sampled, it is once again emphasized that an appropriate subsampling within prison populations is indispensable (as is the adequate choice of a control group). Thus, in the overall consideration of these meta-analyses and the contradictory recent findings on IC (see above), it can be stated that the assumption of a deficient cognitive (inhibitory) control among APDs lacks empirical evidence and is by no means certain.

1.4. Open Research Questions

Even though the field of ER is receiving more and more attention (Gross, 2015), research focusing on ER among offender populations and specifically among APDs is scarce. As Gross and Jazaieri (2014) already outlined, there is a “gap between clinical intuition and empirical findings” (Gross & Jazaieri, 2014, p. 396). Although a deficient ER seems plausible in APDs, both the exact impairments (~ abnormalities compared to HCs) and the specificity of these potential deficits (~ differences compared to INCs) are still unclear. With respect to actual regulation patterns in the lab, AR success as well as behavioral aggression, I am not aware of any research including inmates with and without APD.

The situation is similarly inconclusive regarding cognitive (inhibitory) control: Not only is the clinical relevance of APDs’ impairments in executive functions questionable, but it is also unclear, whether or not such potential deficits are already apparent in more basic cognitive control processes, such as IC. No studies thus far examined IC in APDs and included both, an offender control group and a never-incarcerated HC group. Examining APDs’ IC abilities would be an important first step in delineating possible deficits in executive functioning. If only the higher-level executive functions (such as planning) are considered, it remains unclear which underlying skills are responsible for potential deficits, since the higher-order executive functions also require lower-level cognitive control abilities.

In sum, as intuitively plausible or even obvious deficits in ER and cognitive (inhibitory) control may seem in offenders with APD, the former still needs to be specified and the latter must be proven first. Should impairments in self-regulation distinguish inmates with and without APD, this would speak for the validity and relevance of the psychiatric diagnosis of APD, which is often criticized because of its behavioral phenotype. Furthermore, and although

the present study is to be classified as basic research, impairments in inmates compared to HCs may indicate potential starting points for interventions programs, since the abilities under investigation are a possible explanatory mechanism for delinquent behavior within a multi-causal construct.

1.5. Overall Goals of the Present Work

The main goal of the present work was to find out whether inmates with APD differ from INCs and HCs in aspects of self-regulation, more precisely, with respect to their ER, including AR and aggressive behavior, and their IC performance.

In order to be able to assess APDs' spontaneous AR, first, a new AI and aggression paradigm had to be developed. Therefore, preliminary studies aimed to evaluate this new paradigm and, based on the results, to improve it where necessary. It was intended to provide evidence for the instruments' effectiveness in inducing angry feelings and its general suitability for measuring (reactive) aggressive behavior (more details with respect to these preliminary studies are given in chapter 2). Regarding ER, it was aimed to specify APDs' (potential) deficits by investigating habitual ER, including, but not limited to AR, as well as spontaneous AR and (reactive) aggressive behavior – the latter by using the aforementioned AI and aggression paradigm (precise hypotheses concerning ER are specified in chapter 3.1.3). With respect to IC, the study's purpose was to examine whether APDs indeed exhibit deficits as compared to HCs. Furthermore, it was aimed to determine whether these (potential) impairments are specific for offenders with APD and the domain of IC or also apply to offenders without APD (i.e. INCs) and/or other aspects of cognitive control (more information regarding these objectives is provided in chapter 4.1). The scope of the present work is sketched out schematically in Figure 2.

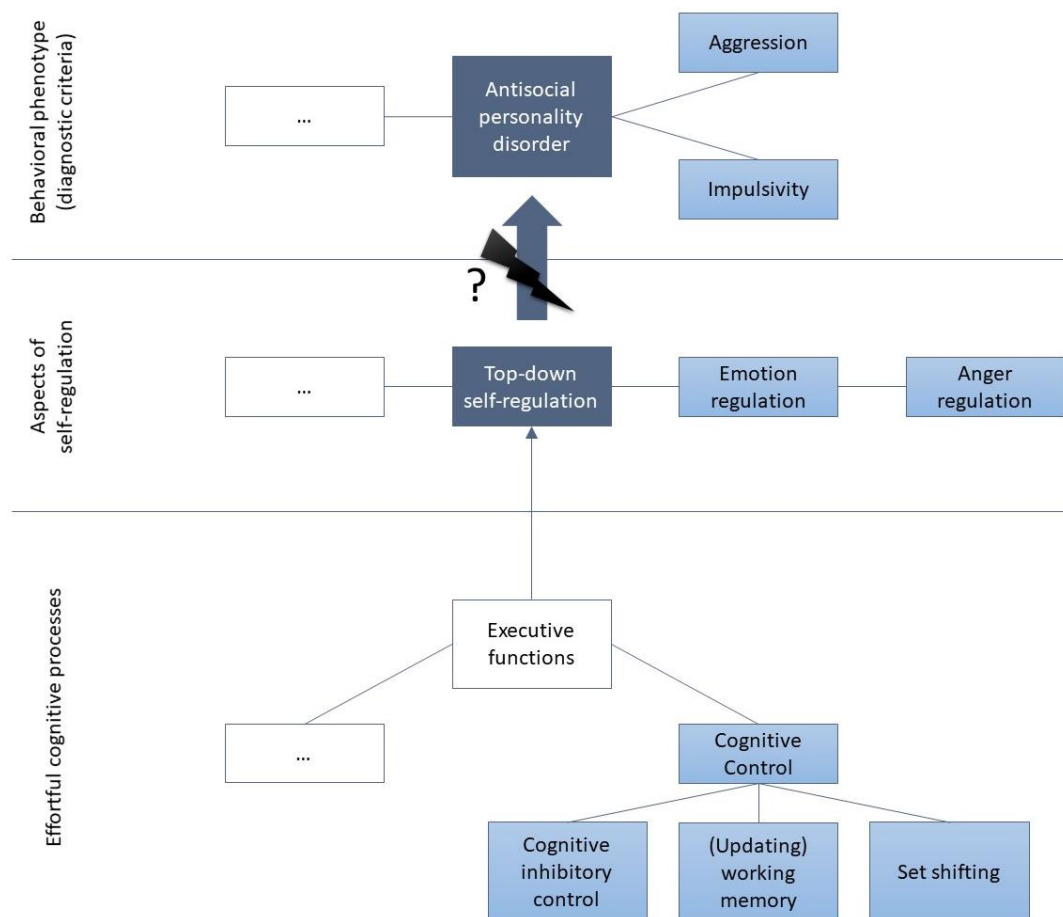


Figure 2. Scope of the present work.

Symptoms of antisocial personality disorder and aspects possibly contributing to its behavioral phenotype, which are investigated in the present study, are marked in light blue.

1.6. Structure of the Thesis

The present work is divided into three manuscript-like chapters. In chapter 2 results of two preliminary studies are briefly outlined for reasons of transparency. As stated above, these were conducted to evaluate a newly developed instrument that combines an AI method with an aggression paradigm that was later used in the main study. The findings of the main study are then individually presented in chapter 3 and chapter 4. While chapter 3 focuses on ER and particularly AR, including aggression, chapter 4 mainly examines cognitive control and IC. These manuscripts are to be published in a peer reviewed journal but have not yet been submitted. A declaration on the share of collaborative work is depicted in Table 2. In chapter 5 the thesis concludes with an overall discussion of the results of the main study.

Table 2. Declaration according to § 5 Abs. 2 No. 8 of the PhD regulations of the Faculty of Science

Author	Scientific ideas	Research design	Data generation	Analysis & interpretation	First draft manuscript writing	Manuscript revision
Elena Schreiner	50%	90%	100%	90%	100%	75%
Larissa Wolkenstein	50%	10%	0%	10%	0%	25%

Note. This declaration applies to both manuscripts intended for publication (chapter 3 and chapter 4).

2. Preliminary Studies

As outlined in chapter 1.3.1, AIs previously conducted with ASBs show considerable shortcomings. Looking at existing aggression paradigms, similar concerns emerge. The classic “Big Four” (Tedeschi & Quigley, 1996, p. 165) – the aggression machine (also known as teacher/learner paradigm; Buss, 1961), essay evaluation paradigms (Berkowitz, Corwin, & Heironimus, 1963), competitive reaction time games (Taylor, 1967) and the Bobo modeling paradigm (Bandura, 1973) – do not seem suitable for application in a prison context: With respect to teacher/learner and essay evaluation paradigms (and in parts the competitive reaction time game), the availability of severe response options such as the (apparent) delivering of electric shocks (or, in newer versions: apparent delivering and receiving of loud noise blasts) in retaliation for a previous insult raises considerable ethical concerns. Second, cover stories intend to lower the inhibition threshold for aggressive behavior and describe the delivering of shocks or noxious sounds as an aid for the (confederate) participant to improve his or her learning (teacher/learner and essay evaluation paradigms). Accordingly, the “harm doing behavior” is more or less defined as a prosocial tool. Hence, should participants believe in the cover story, construct validity must be questioned, because what is measured is most likely not aggression (Tedeschi & Quigley, 1996). Serious doubts about construct validity also apply to the Bobo modeling paradigm, where a mannequin is punched. Obviously no harm can be delivered to a mannequin – therefore no aggressive behavior is assessed here either. Furthermore, it seems questionable whether a rather academic ego-threat (insults in the essay evaluation paradigms) or a competitive character (competitive reaction time games) is just as successful in inducing anger in a sample of incarcerated offenders as it is in students, with whom the paradigms were mostly applied. Slightly newer aggression paradigms also seem debatable, either due to ethical reasons (e.g. bungled procedure; Russell, Arms, Loof, & Dwyer, 1996), practical reasons (e.g. hot sauce paradigm; Lieberman, Solomon, Greenberg, & McGregor, 1999) credibility reasons (e.g. negative evaluation tasks, e.g. see DeWall, Twenge, Gitter, & Baumeister, 2009) or serious objections to construct validity (e.g. uncomfortable pose task; Finkel, DeWall, Slotter, Oaten, & Foshee, 2009). More detailed information and criticism on recent aggression paradigms is given in Ritter and Eslea (2005) and McCarthy and Elson (2018).

As a result of the aforementioned shortcomings, it was intended to create a new AI/aggression paradigm. Former research comparing the effectiveness of AI instruments revealed that methods with personal contact are superior to those without personal contact, while the utilization of insults proved to be particularly effective (Lobbestael et al., 2008).

Novaco (2011) outlined that perceived provocations typically comprise not only insults, but also unfair treatment or intended thwarting. Therefore, the AI paradigm to be used should contain both, insults and unfair treatment. However, it was intended to limit the investigator's involvement in order to decrease experimenter effects (c.f. Tedeschi & Quigley, 1996) and enhance internal validity. Furthermore, it was aimed to elicit a significant amount of angry emotions, while taking into account the special circumstances of a prison environment, i.e. legal, ethical and practical considerations. It was therefore decided to use a standardized (fake) chat conversation as a (supposedly) social interactive context. Through various provocations a personal involvement should be established and anger should be elicited (Harmon-Jones, Amodio, & Zinner, 2007). Given that the assessment of a behavioral measure of physical aggression is hardly feasible, let alone in a prison context, it was decided to study an indirect and active form of monetary harm, i.e. resource aggression (classification according to Parrott & Giancola, 2007). While the AI was aimed to be embedded in a cover story, the aggression task should not be explicitly justified in order to be able to interpret participants' responses more plausibly as intended to cause harm (see criticism by Tedeschi & Quigley, 1996). Furthermore, and in contrast to some previously used aggression paradigms, it was intended to include a non-aggressive response option. Thus, a first version of the Cyberball Aggression Task (CAT) was developed.

2.1. A New Measure – the Cyberball Aggression Task

In this chapter, only rudimentary information about (first versions of) the CAT is given. Further details on the final version of the CAT are described in chapter 3.2.2.

The first component of the CAT comprises a modified version of Cyberball 4.0 (Williams, Yeager, Cheung, & Choi, 2012), used for the AI. Cyberball is an online-ball tossing game, originally used to study ostracism (Williams & Jarvis, 2006). It fakes a social interaction by making participants believe that they are playing the game with two other (alleged) participants. In fact, these other players do not exist, the whole game is operated by the computer (Williams et al., 2012). It was aimed to present a bogus chat conversation between the two (alleged) other players, which is capable of inducing angry emotions. Each time the participants tries to send a chat comment on his own, an error message occurs and hinders the message from being sent. Therefore, he is not able to participate in the conversation. This “bug” was needed to establish the credibility of the chat conversation between the two (faked) players and at the same time to maintain internal validity. During baseline rounds of the game, the two (alleged) other players involve the participant in their communication by asking him questions.

However, they are apparently getting annoyed due to the fact that the participant does not answer (by reasons of the faked system error “hindering” messages from being sent). This is when the (alleged) other players start to insult the participant and the AI begins (so-called anger rounds).

The second component of the CAT is a forced-choice punishment decision, aimed to assess aggressive behavior against the (alleged) other players. After each round of the Cyberball game, the participant has to decide whether he takes away parts of one of the other players’ reward (active, indirect resource aggression, i.e. theft). Thereby, spontaneous aggression (during baseline rounds, without any apparent reason to aggress) and reactive aggression (during anger rounds, when there is an incentive to punish) is assessed. To strengthen the need to inhibit the aggressive behavior, punishment was associated with different probabilities of an own negative consequence (i.e. own money loss). The development process of the CAT, from creation of the chat comments until its final application in the main study, is depicted in Figure 3.

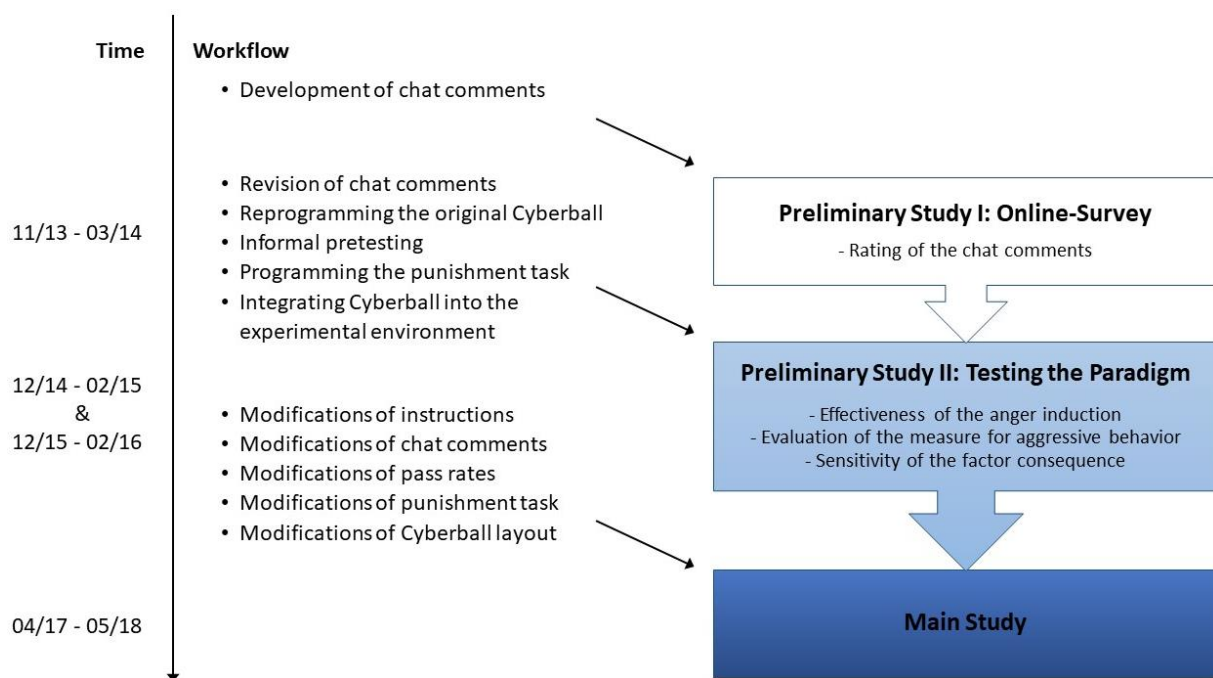


Figure 3. Development process of the Cyberball Aggression Task from the preliminary studies to the main study.

The following is a brief outline of the most important results of two larger preliminary studies, which were conducted to test and adapt the new AI paradigm, the CAT. First, results of an online survey are outlined (see chapter 2.2). Hereafter, a second preliminary study (see chapter 2.3) is briefly summarized.

2.2. Preliminary Study I: Online Survey

A first step in the development of the CAT was to create insulting chat comments that might work as AI. To receive a first evaluation of the chat comments' suitability, an online survey was conducted. Participants were expected to report increased levels of angry emotions after insulting chat comments compared to comments assigned to the baseline.

2.2.1. *Methods*

Participants. Only men were included in the online survey, as the later main study was also restricted to men. Of the $N = 324$ male adults who completed the entire survey, the majority was highly educated (96.6% had a university entrance diploma). The mean age was $M = 27.20$ years ($SD = 9.10$). A minority indicated one or more lifetime offences (17.6%), while 44.4% reported prior involvement in a brawl. Only $n = 1$ participant (0.3%) indicated a previous conviction.

Measures and procedure. All participants who completed the survey had the opportunity to win one of four vouchers. At the very beginning of the survey, informed consent was obtained, followed by a short assessment of demographic information to assess eligibility for this study (inclusion criteria: men, age ≥ 18 years). At the beginning of the rating task, participants were asked to imagine that they were playing an online ball-tossing game with two other unknown players. They were told that these players could communicate with each other via chat. However, the chat function does not work properly, so the participant can read other players' comments but cannot write anything himself. After this instruction, a chat conversation between the two players "Player 1" and "Player 3", divided into 18 short sections, was sequentially presented. Each sequence contained two to four comments. Before the first section (pre) and after each following section, participants were asked to indicate their current emotional state if they imagined the chat conversation would actually happen. The adjectives irritable, upset and hostile (from the Positive and Negative Affect Schedule, PANAS; Watson, Clark, & Tellegen, 1988) and the filler items nervous, scared, helpless, sad and happy were presented. Items were rated on a 5-point Likert scale ranging from 1 ("not at all") to 5 ("extraordinary"). Section 1-5 represented the baseline (no insults), while from section 6 on, chat comments became increasingly insulting (beginning of the AI). After this rating task additional measures, which are beyond the scope of the present work, were conducted. Then, participants had the opportunity to give qualitative feedback on how to improve the chat

conversations' credibility. At the end of the survey, further demographic information was assessed.

2.2.2. Results

Participants' mean ratings of the items upset, hostile and irritable prior to the rating task and after each section of the chat are depicted in Figure 4.

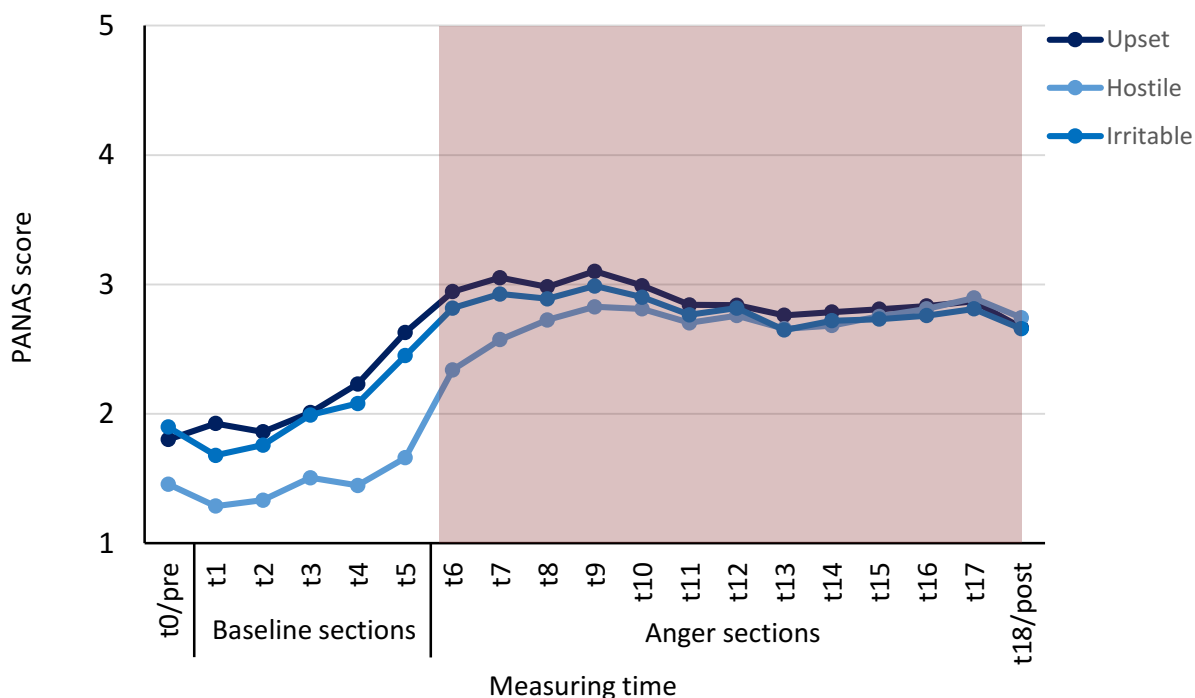


Figure 4. Mean PANAS score depending on measuring time.

PANAS = Positive and Negative Affect Schedule. t = measuring time (i.e. chat section number). Red background color indicates sections containing the anger induction.

Repeated measures multivariate analysis of variance (MANOVA) with the factor measuring time on the dependent variables upset, hostile and irritable indicated a significant difference in the combined dependent variables between measuring times, $V = 0.40$, $F(54, 17442) = 50.15$, $p < .001$. Follow-up univariate analyses of variance (ANOVAs⁵) revealed significant alterations over time for all dependent variables: upset, $F(5.88, 1900.33) = 73.12$, $p < .001$, hostile, $F(4.39, 1417.08) = 160.02$, $p < .001$, and irritable, $F(5.45, 1759.00) = 74.73$, $p < .001$. To further investigate whether baseline sections differed from anger sections with respect to each angry emotion, mean values of upset, hostile and irritable were calculated separately for baseline and anger sections (see Figure 5).

⁵ In case of violation of sphericity, degrees of freedom were adjusted by either using the Greenhouse-Geisser (for $\epsilon < 0.75$) or the Huynh-Feldt (for $\epsilon > 0.75$) correction.

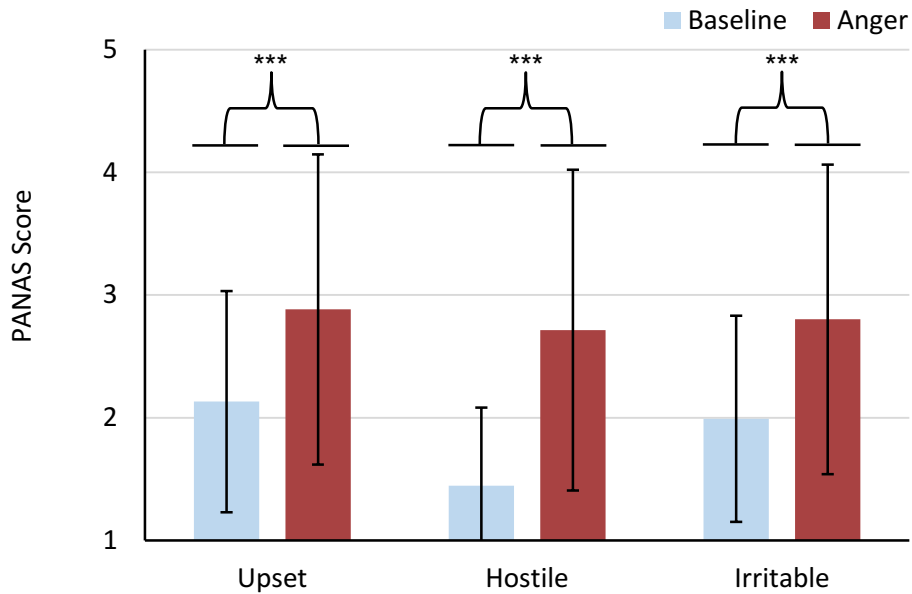


Figure 5. Mean angry emotions depending on baseline and anger sections.

Error bars represent standard errors of the means. Asterisk indicate significance level of effects.

*** $p < .001$

Dependent t -tests revealed significant differences between baseline and anger sections for each angry emotion: upset, $t(323) = 12.41$, $p < .001$, $d = 0.865$, 95% confidence interval (95% CI)⁶ [0.704, 1.026], hostile $t(323) = 19.79$, $p < .001$, $d = 1.942$, 95% CI [1.755, 2.129], and irritable, $t(323) = 13.73$, $p < .001$, $d = 1.017$, 95% CI [0.854, 1.181].

2.2.3. Implications

In view of the large effect sizes, there was preliminary evidence that the AI actually works. However, and as a result of the quantitative data depicted above, but also based on qualitative feedback and own reflections, the chat comments were subsequently modified in order to be able to evoke actual “real-life” changes in angry emotions (as opposed to the mere imagination of affect reactivity). For example, weak insults were excluded, spelling mistakes were included, and the chat conversation was abbreviated from 18 to 12 sections. Hereafter, the original Javascript code of Cyberball was reprogrammed in order to implement the AI (i.e. the chat conversation, the bogus error message, etc.). After informal pretesting of this adapted version of Cyberball, it was put together with a task intended to measure aggression and embedded in the E-Prime 2.0 experimental environment. Thus, a first version of the CAT was created.

⁶ 95% CIs for Cohen’s d were calculated by using the freeware Psychometrica (Lenhard & Lenhard, 2016).

2.3. Preliminary Study II: Testing the Paradigm

Preliminary Study II was conducted to assess the suitability of the computerized CAT (for a description of the task see chapter 2.1). More specifically, it was aimed to test the effectiveness of the AI concerning real-life emotions. The second main goal was to evaluate the measurement of aggressive behavior.

A significant increase in angry emotions due to the CAT was expected (i.e. from pre to post). While no punishing behavior (i.e. money deduction for one of the alleged other players) was hypothesized during baseline rounds of Cyberball (no provocation and thus no incentive to reduce frustration), punishing behavior was indeed hypothesized during anger rounds. Accordingly, analyses should reveal a significant increase in punishing behavior from baseline to anger rounds. In addition to that, visual data inspection should yield no floor effects of punishing behavior during the AI. Further, a significant variation of punishment due to the factor consequence (0%, 15%, 50% or 100% probability of own money loss) was expected during anger rounds. As briefly outlined above, this factor was included to strengthen the participants' need to inhibit aggressive behavior in order not to jeopardize his own goal of profit maximization (i.e. cognitive control task). As a general pattern, the punishing behavior was expected to decrease with increasing probability of own negative consequence. There were no hypotheses regarding consequence in baseline rounds (due to assumed floor effects of punishing behavior in this condition).

2.3.1. *Methods*

Participants. Thirty-five participants were included in the final sample of this preliminary study. They were recruited in the community through direct approach (e.g. in homeless centers, job centers, discount stores, fast food restaurants, at the train station) and through advertisements (e.g. online platforms, flyer). Inclusion criteria were: male gender, $18 \leq \text{age} \leq 70$, sufficient knowledge of the German language and no university entrance diploma (German: "Abitur"). Before study enrollment participants were screened for past physical aggression (desired criteria). Demographics of the final sample are depicted in Table 3.

Table 3. Demographic information for participants

Characteristic	%	
Highest level of school education		
Without graduation	5.8	
Certificate of Secondary Education (German: Hauptschulabschluss)	42.9	
General Certificate of Secondary Education (German: Mittlere Reife)	51.4	
Lifetime illegal drug consumption (yes)	51.4	
Lifetime brawl		
never	27.3	
1	15.2	
2 - 3	33.3	
≥ 4	24.2	
Lifetime offence committed (yes)	58.8	
Criminal record (yes)	26.5	
	<i>M</i>	<i>SD</i>
Age	28.80	11.39
MWT-B IQ	92.79	11.87

Note. MWT-B IQ = Intelligence quotient assessed with the Multiple Choice Word Fluency Test. $N = 35$ participants.

Measures and procedure. All participants received financial compensation for study participation. As in the later main study, participants were deceived about the true purpose of the study. They were told that the study is about cognitive skills in different groups of people. First, demographic information was assessed. Then, (an earlier version of) the CAT was presented: A total of 12 rounds of Cyberball were played. Chat comments and pass rates were slightly different from those shown in Appendix A, which depicts the final version. The first four rounds served as baseline, while the AI was conducted in rounds 5 to 12 (anger rounds). After each round of Cyberball, the so-called punishment task (i.e. the aggression measure) was conducted: participants had to decide whether they wanted to deduct money from Player 1 or not. The decision to punish the other player was interpreted as aggressive behavior (monetary harm). Punishment was coupled to a specific probability of own money loss in order to enhance the need to control one's aggressive behavior (factor consequence). There were four factor levels of consequence (0%, 15%, 50% and 100% probability of own loss). Before (pre) and after the CAT (post) the complete PANAS, consisting of ten negative and ten positive affect terms, was conducted. For the current thesis only the items upset, hostile and irritable are of

interest. Each item was rated on a 5-point Likert scale. Subsequently, participants' beliefs about the task were assessed to receive a qualitative feedback on credibility and improvement opportunities. Then, the Multiple Choice Word Fluency Test (MWT-B; Lehl, 2005), for which IQ estimates are reported, and additional measures, which are beyond the scope of this thesis, were conducted.

2.3.2. Results

Angry emotions. A repeated measures MANOVA with the factor measuring time (pre, post) on the dependent variables upset, hostile and irritable was carried out. Means are depicted in Figure 6.

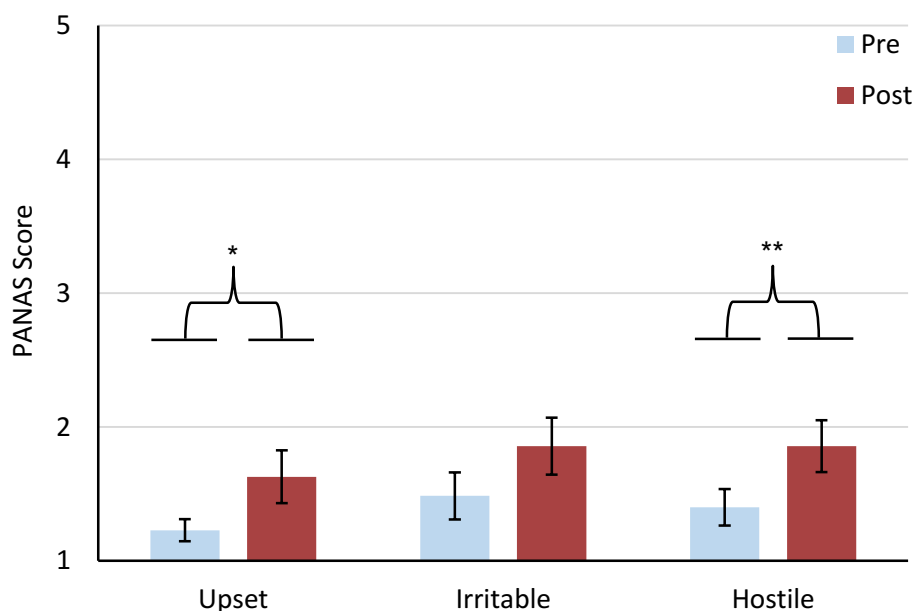


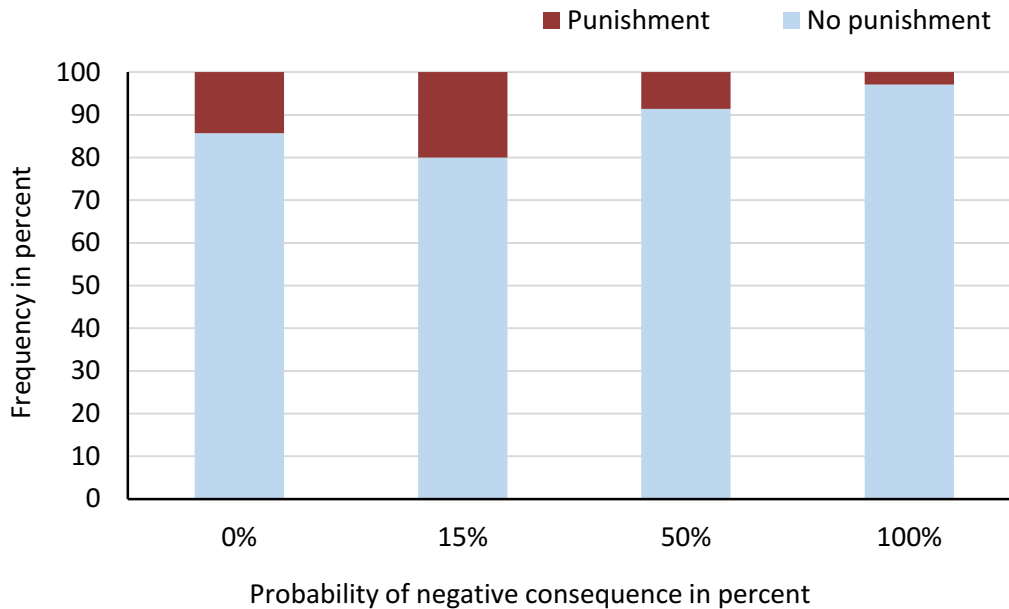
Figure 6. Mean angry emotions before (pre) and after (post) the Cyberball Aggression Task. Error bars represent standard errors of the means. Asterisks indicate significance level of effects. * $p < .05$. ** $p < .01$.

A significant effect of measuring time occurred, $V = 0.22$, $F(3, 32) = 3.06$, $p = .043$. Follow-up dependent t -tests showed significant increases of emotional experiences between pre and post regarding upset, $t(34) = 2.07$, $p = .046$, $d = 0.669$, 95% CI⁶ [0.187, 1.150] and hostile $t(34) = 2.94$, $p = .006$, $d = 0.632$, 95% CI [0.152, 1.112]. However, with respect to irritable, participants' ratings did not significantly differ between pre and post, $t(34) = 1.68$, $p = .102$.

Taken together, and although results indicated that angry emotions were successfully induced, the AI was clearly improvable, as evident by low means of angry emotions after the CAT (post).

Punishing behavior. Figure 7 depicts frequencies of punishing behavior for baseline (Figure 7a) and anger rounds (Figure 7b).

a) Baseline rounds



b) Anger rounds

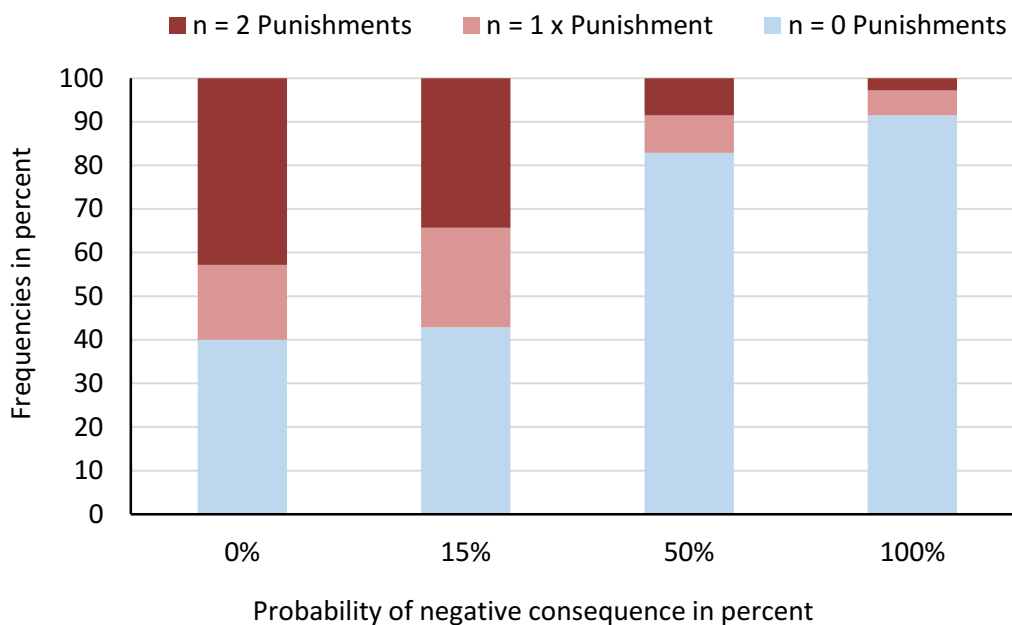


Figure 7. Frequency of punishing behavior among participants in baseline and anger rounds and depending on consequence.

Each consequence (0%, 15%, 50%, 100%) was presented once in baseline rounds ($n = 4$ rounds) and twice in anger rounds ($n = 8$ rounds), resulting in a different number of maximum punishments depending on condition.

Punishing behavior was separately analyzed for condition (baseline rounds, anger rounds). For baseline rounds, Cochran's Q test was conducted due to dichotomous data. Analysis revealed no significant effect of consequence, $\chi^2(3) = 6.67, p = .083$. Hence, during baseline rounds, participants showed similar punishing behavior across consequence, while hardly deducting any money (see Figure 7a).

As punishing behavior was only expected during the AI, analyses within anger rounds were of greater importance for the evaluation of the paradigm. Here, Friedman's ANOVA indeed showed a significant effect of consequence, $\chi^2_F(3) = 35.41, p < .001$. Follow-up Wilcoxon tests are depicted in Table 4. With the exceptions of 0% and 15% and 50% and 100%, respectively, all pairwise comparisons reached significance. Generally speaking, and as expected, punishing behavior decreased with increasing probability of negative consequence (see Figure 7b).

Table 4. Wilcoxon signed-rank tests for pairwise comparisons of consequence within anger rounds

	15%	50%	100%
0%	$z = 0.89, p = .372$	$z = 3.63, p < .001$	$z = 3.96, p < .001$
15%		$z = 3.35, p = .001$	$z = 3.76, p < .001$
50%			$z = 1.89, p = .059$

Note. Percentages (0%, 15%, 50%, 100%) indicate the probability of negative consequence and represent factor levels of consequence.

Taken together, the paradigm proved to be successful in inducing anger, though to a relatively small extent. Effectiveness of the AI was suggested rather by looking at behavioral data: While floor effects occurred regarding punishment in baseline rounds, participants indeed showed aggressive behavior when there was an incentive to punish, i.e. when being provoked during anger rounds. Moreover, consequence had an effect on participants' decisions to punish during anger rounds and thus proved its suitability.

2.3.3. Implications

Based on the findings mentioned above, the task was revised again. For better analysis, and due to the fact that the consequences 0% and 15% or rather 50% and 100% achieved similar results, the factor consequence was dichotomized by using two end points (no consequence vs. potential negative consequence). By contrast, dichotomous punishment behavior (yes vs. no) seemed disadvantageous due to its low sensitivity when detecting effects. As continuous data is usually preferred as a dependent variable, the response mode was changed from forced-choice

to a visual analogue scale, ranging from 0 to 100. Furthermore, the extent of punishing behavior was linked to the probability of the negative consequence, i.e. more money deduction was linked to increased risk of own reduction in remuneration in the potential negative consequence condition. In order to slightly increase the subjective experience of angry emotions, the chat conversation was modified again: Presumably weak comments were removed and a new provocation was designed by maximizing social exclusion (i.e. no ball-tosses to the participant from round 9 on, for information on pass rate see Appendix A). To further improve the credibility of the paradigm, a wireless USB modem was allegedly used in prisons to get internet access. Moreover, instructions were modified, a long waiting time before the first round of Cyberball (“waiting for other players”) was implemented to reflect the difficulty in temporal coordination with the alleged other participants, typing errors were increased and the temporal fit between ball tosses and chat comments was improved. The final chat conversation is depicted in Appendix A. The resulting modified version of the CAT was conducted in the main study of this thesis and is described in more detail in chapter 3.2.2.

3. Main Study, Part I: Similar, Yet Different – Disturbed Emotion Regulation as a Distinctive Feature Among Antisocial as opposed to Non-Antisocial Offenders and Healthy Controls⁷

Abstract:

Despite the growing body of emotion regulation (ER) research within offender populations, the exact pattern of potential ER deficits among inmates with antisocial personality disorder (APDs) remains unclear. Therefore, the current study comprehensively assessed $n = 31$ APDs' self-reported habitual and spontaneous ER, including, but not limited to, anger regulation (AR). In addition, we investigated abnormalities in (reactive) aggressive behavior using a newly developed anger induction (AI) paradigm. $N = 33$ inmates without APD (inmate controls; INCs), and $n = 39$ never-incarcerated healthy controls (HCs) served as control groups. With respect to habitual AR, APDs reported chronic anger experience, accompanied by increased anger suppression and expression compared to both, INCs and HCs. By contrast, all groups reported similar anger reactivity and strategy use in response to the AI. Whereas APDs did not show increased reactive aggression, they behaved more aggressively than INCs and HCs *without* prior provocation, thus suggesting an elevated spontaneous aggression proneness. INCs, on the contrary, showed less reactive aggression than APDs and HCs. Regarding ER beyond AR, APDs, but not INCs, indicated overall emotion dysregulation and impulse control difficulties. Further, APDs reported increased habitual use of the strategy of blaming others compared to INCs. Within inmates, maladaptive ER predicted antisocial symptom severity – even when controlling for other variables. Overall, the current study provides clear evidence for increased (habitual) ER deficits in APDs as opposed to INCs. Different intervention programs might be suitable for these offender subgroups.

General scientific summary: The current study emphasizes emotion regulation deficits as a distinctive feature among antisocial compared to non-antisocial offenders. Antisocial personality disorder should be considered as a disorder of emotion regulation.

Keywords: Antisocial Personality Disorder, Offender, Emotion Regulation, Emotion Regulation Strategies, Anger Induction, Aggression

⁷ An abridged version of this manuscript is intended for publication but has not yet been submitted. A declaration on the share of collaborative work is given in Table 2 in chapter 1.6.

3.1. Background

APD is mainly defined by behavioral aspects, whereas affect and inner experiences are less stressed (American Psychiatric Association, 2013). However, irritability and aggressive behavior are among the key symptoms of APD (American Psychiatric Association, 2013). While irritability (cognition) could reflect increased anger reactivity, aggression (behavior) is assumed to be evoked by unpleasant emotions, particularly anger (e.g. Miller et al., 2012; Novaco, 2011; Robertson, Daffern, & Bucks, 2015). Accordingly, one might expect APDs to also exhibit abnormalities in ER. But is this assumption also empirically tenable?

3.1.1. *Anger Experience, Anger Regulation and (Reactive) Aggressive Behavior*

There is indeed evidence suggesting that more severe APD symptomatology is associated with increased anger experience: Within a typology of men convicted of intimate partner violence, those batterers who belonged to the cluster with the most APD symptoms, also reported increased state and trait anger (Graña et al., 2014). Moreover, higher trait anger within male APDs was found to be associated with the number of violent convictions (Kolla et al., 2017). However, among other limitations – number of convictions is not equateable to number of offences committed, self-report assessments of APD have to be questioned – it must be taken into account that both studies refer to a dimensional view of APD and do not relate APDs' reports to a HC group. Interestingly, two recent studies that both assessed APD dichotomously and included HCs, found evidence for increased trait (Yavuz, Şahin, Ulusoy, İpek, & Kurt, 2016) as well as state anger (Timmermann et al., 2017) among APDs as compared to HCs. However, Yavuz et al. (2016) recruited APDs from an outpatient clinic. Thus, it is unclear whether the abnormalities found are due to the APD diagnosis or are a result of (unreported) comorbid psychiatric disorders, which, after all, had to be severe enough to indicate therapeutic treatment. Timmermann et al.'s (2017) findings seem more meaningful, although generalizability to male APDs of a broader age range seems questionable, since they recruited a young sample of mixed gender, including female participants with comorbid borderline personality disorder.

Considering that about one in two male prisoners exhibits APD (Fazel & Danesh, 2002), it seems meaningful to also clarify whether or not increased anger experience is a distinct phenomenon amongst the – supposedly more severely impaired – subgroup of APDs or whether it is a broader phenomenon to be found in criminals in general. So far, evidence is mixed: Some studies suggest higher trait anger among criminals not limited to APDs (Barbour et al., 1998),

whereas others contradict this assumption (Garofalo, Velotti, & Zavattini, 2018), and still others found different results between offender groups (Gillespie, Garofalo, & Velotti, 2018). However, due to the lack of diagnostic information regarding APD, results are hardly meaningful to interpret. Future research should not only recruit a HC group but also further subdivide offender populations based on a thorough diagnosis of APD.

Given that anger alone does not inevitably lead to aggressive urges, but instead the effective ER is crucial (Hawes et al., 2016; Robertson et al., 2012), one might expect adults with APD to not only exhibit increased trait anger but also deficient AR. Indeed, recent evidence supports this assumption: Compared to HCs, APDs reported a maladaptive AR pattern as evident by increased outward anger expression and inward anger suppression (Timmermann et al., 2017; Yavuz et al., 2016), alongside decreased anger control (Timmermann et al., 2017), even though null findings occurred regarding the latter strategy (Yavuz et al., 2016). Variety of AR patterns within offenders (Gillespie et al., 2018; Low & Day, 2015) again emphasizes the need for adequate subsampling.

Beyond this unclear state of research, it has to be considered that information on APDs' habitual AR strategy use provides only limited information on actual AR and its success. Hence, it would be beneficial to additionally assess APDs' *spontaneous* AR. Despite its high relevance, to our knowledge, there is only one study so far that has experimentally induced anger in APDs: Following an unstandardized stress-induction interview, in which participants had to recall and describe an anger-evoking situation from the past, Lobbestael et al. (2009) did not find differences in self-reported anger reactivity between female and male patients with APD as opposed to participants with and without personality disorders other than APD. However, methodological issues must be borne in mind: It remains unclear to what extent APDs have been mentally ill offenders and non-incarcerated psychiatric patients, thus challenging generalizability of results, for example to a prison sample. Given that APDs' anger reactivity was not compared to HCs' anger reactivity but to the reactivity of the overall sample, potential deficits in APDs could have been masked. Beyond that, it cannot be ruled out that APDs have been less (more) capable of empathizing with past situations than the other groups and/or reported less (more) severe anger-evoking situations, which could have distorted the results. Consequently, there is a need for a standardized AI paradigm to investigate AR in APD.

Unfortunately, anger is difficult to induce by using conventional methods such as film clips, considering that this emotion requires a high degree of personal involvement and temporal immediacy (Rottenberg et al., 2007). Therefore, social psychological methods with cover stories and personal contact seem to be necessary for a successful AI (Harmon-Jones et al.,

2007; Lobbestael et al., 2008). Previous research particularly emphasized the effectiveness of harassment (Lobbestael et al., 2008). At best, the nature of the manipulation is concealed, while internal validity is still ensured (Harmon-Jones et al., 2007). However, formerly used AI methods either have high demand characteristics due to their face validity (e.g. variations of the Articulated Thoughts in Simulated Situations task; Davison, Robins, & Johnson, 1983), have a possibly reduced efficacy within offender populations due to their competitive and/or performance based character (e.g. essay-evaluation paradigms; Berkowitz et al., 1963) or are difficult to implement in a prison setting due to questionable ethical and legal aspects (e.g. Bungled Procedere paradigm; Russell et al., 1996). Therefore, there is a need for a standardized AI paradigm that creates a supposedly interactive context sufficient to achieve a significant increase in angry emotions, while accounting for the specific restrictions of law enforcement (i.e. organizational and safety-related aspects).

Given that anger, as every other emotion, initiates action tendencies, it can, but does not need to, result in aggressive behavior (Berkowitz, 1989). Assuming AR difficulties within APDs and considering their diagnostic criteria, it seems obvious to also expect an increased readiness for hostile, affective aggression in response to provocations (i.e. increased reactive aggression as a result of an AI in the lab). To our knowledge, however, no laboratory controlled measurement of aggression in APDs has yet been conducted. Correspondingly, empirical evidence lacks.

3.1.2. Emotion Regulation – General Abilities and Strategy Use

Research suggests that not only anger, but also other unpleasant emotions are antecedents of aggressive behavior – though assumably to a lesser extent (Robertson et al., 2012). Gratz and Roemer (2004) proposed several abilities needed for an adaptive ER: the capacity to experience, understand and differentiate emotions, accepting and valuing emotional responses, inhibiting impulsive behaviors, behaving analogue to own goals, and flexibly applying ER strategies depending on the context. According to the authors, deficits in one of these abilities indicate emotion dysregulation. Previous work has found a link between such emotion dysregulation and violence *within* different offender populations (Robertson, Daffern, & Bucks, 2014; Robertson et al., 2015; Tager, Good, & Brammer, 2010). However, when looking at studies that relate offenders' ER to comparison groups, offenders do not show consistent deficits, instead, there is only evidence for reduced emotional acceptance (Garofalo et al., 2018; Gillespie et al., 2018) of negligible size (Garofalo et al., 2018). Yet again it is quite possible that existing deficits in APDs (e.g. compared to HCs) were masked by assigning

heterogeneous groups of people (offenders with and without APD) into a single group and not recruiting an adequate HC group. Further research is needed to clarify whether APDs exhibit problematic ER beyond anger.

Besides the aforementioned more general ER abilities, transdiagnostic ER research also emphasizes the relevance of ER strategy choice. Robertson et al. (2012) stress that rather dysfunctional strategies like suppression, avoidance and rumination contribute to aggressive behavior. This raises the question whether and to what extent APD is (also) affected by an altered ER strategy use as opposed to INCs and HCs. To our knowledge there is, surprisingly, only one study that assessed ER strategy use within offenders: Gillespie et al. (2018) found no differences in reappraisal and suppression use among male violent, sexual and homicide offenders compared to non-incarcerated controls. Again, no adequate psychiatric assessment was carried out and only a limited number of ER strategies were examined.

3.1.3. Goals of the Present Work

Taken together, the exact pattern of potential ER deficits among APDs as opposed to INCs and HCs is still unclear. Thus, the present work aimed to broadly assess APDs' *habitual* ER by specifically examining their (1) AR, but also their (2) more general emotion dysregulation and their (3) ER strategy use in the context of unpleasant emotions that are not limited to anger and to compare their reports to HCs and INCs. Furthermore, we addressed APDs' *actual* ER pattern by using a newly developed experimental AI. In this regard we investigated APDs' (4) spontaneous ER strategy use, their (5) self-reported experience of angry emotions and arousal, as well as the behavioral correlate, their (6) (reactive) aggression before and during the AI.

Based on previous research, we expected APDs to report (1) increased state and trait anger as well as increased outside anger expression and inside anger suppression compared to HCs and INCs, while exploring differences in anger control. We further expected APDs to report (2) overall emotion dysregulation compared to HCs and INCs. Since previous research lacks, we neither had specific hypothesis regarding their exact pattern of ER difficulties (i.e. the different facets proposed by Gratz, Moore, & Tull, 2016) nor (3) their habitual and (4) spontaneous ER strategy use. However, given their diagnostic criteria, we expected APDs to report (5) heightened anger reactivity following the AI and to show (6) increased aggressive behavior, especially when provoked – both compared to INCs and HCs. Due to the lack of relevant findings, no further predictions were made regarding differences between APDs and INCs, while potential differences between INCs and HCs should be explored. Last, we aimed

to explore (7) whether APD symptom severity in offenders is associated with ER impairments when viewing APD dimensionally.

3.2. Methods

3.2.1. Participants

Since epidemiological research has shown that APD is particularly prevalent in male inmates (Moran, 1999), APDs as well as INCs were recruited in prisons, while the entire sample was limited to men. Of originally $N = 137$ individuals, $N = 103$ participants were included in the final sample. Data was collected between April 2017 and July 2018. Inmates were recruited from three German prisons. In two prisons located in Baden-Wuerttemberg ($n = 11$ and $n = 8$) inmates were pre-selected by prison staff, while inmates of a Bavarian prison ($n = 45$) were briefly screened in advance to study participation by the investigator. Here, participants either applied in response to advertisements at bulletin boards or were specifically addressed by the investigator during prison routine. HCs were recruited in the community through advertisements (e.g. online platforms, flyer) and were thoroughly screened with regard to inclusion criteria prior to study participation.

Inclusion criteria for all participants were: male gender, $18 \leq \text{age} \leq 69$, no psychotropic medication, unless stable dosage for at least 4 weeks, sufficient knowledge of the German language, and verbal IQ > 80 . APDs ($n = 31$) met the diagnostic criteria for current APD (i.e. significant symptomatology within the last years following the Structured Clinical Interview II for DSM-IV, SCID-II; First, Spitzer, Gibbon, & Williams, 1996), whereas INCs ($n = 33$) did not. Both were either in pre-trial detention or in criminal custody (closed prison) and self-reported at least one criminal offence beyond traffic offences and offences against foreigners' law to ensure criminality. In contrast, HCs ($n = 39$) reported no lifetime imprisonment and did not meet any current or past DSM-5 disorder, according to the Mini International Neuropsychiatric Interview – The M.I.N.I. 7.0.2 (M.I.N.I.; Sheehan et al., 1998). Exclusion criteria for APDs and INCs were: current episode of major depression, current bipolar disorder, current social anxiety disorder, current posttraumatic stress disorder, current substance use disorder (SUD) or alcohol use disorder (AUD) if moderate or severe and no abstinence in the past 6 months⁸, current or lifetime psychotic disorder and current anorexia nervosa.

⁸ Given the increased psychopathology within offender populations and APDs' association with comorbid disorders, particularly SUD and AUD (e.g. Black, Gunter, Loveless, Allen, & Sieleni, 2010; Gottfried & Christopher, 2017), excluding mild forms of SUD and AUD would have reduced external validity of the sample.

Before the application of final inclusion criteria, HCs, but not INCs, were matched to APDs by age (\pm 5 years) and education (university entrance diploma: yes vs. no). Final sample characteristics are depicted in Table 5. Table 6 shows demographic information and Table 7 diagnostic information for the inmate groups. Self-reported committed offences among inmates are depicted in Figure 8.

Table 5. Participants' demographic characteristics and symptom severities

Characteristic	APDs (<i>n</i> = 31)		INCs (<i>n</i> = 33)		HCs (<i>n</i> = 39)		Group comparisons
	%	95% CI	%	95% CI	%	95% CI	
University entrance diploma	6.45		24.24		15.38		Fisher's Exact Test = 3.77, <i>p</i> = .133
Psychotropic medication	16.13^a		15.15^a		0.00 ^b		Fisher's Exact Test = 7.98, <i>p</i> = .014
German citizenship	64.52^a		69.70^a		97.44 ^b		$\chi^2(2) = 13.46$, <i>p</i> < .001
	<i>Mdn (IQR)</i>	95% CI	<i>Mdn (IQR)</i>	95% CI	<i>Mdn (IQR)</i>	95% CI	
Age	28.00^a (15.00)	[26.00, 34.00]	37.00^b (18.00)	[34.00, 44.00]	29.00^a (18.00)	[27.00, 36.00]	H(2) = 6.25, <i>p</i> = .044
ADHD-SR	14.00^a (7.00)	[12.00, 14.00]	13.00 ^{ab} (8.00)	[9.00, 14.00]	10.00^b (9.00)	[9.00, 11.00]	H(2) = 6.17, <i>p</i> = .046
	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	<i>M (SD)</i>	95% CI	
MWT-B IQ	93.84^a (6.95)	[91.67, 96.41]	100.64^b (10.06)	[97.53, 103.62]	104.21^b (14.13)	[100.36, 108.43]	<i>F</i> (2, 64.89) = 10.19, <i>p</i> < .001
SDS-17	8.87^a (3.01)	[7.70, 9.94]	10.12 ^{ab} (3.68)	[8.88, 11.23]	10.64^b (3.12)	[9.66, 11.63]	<i>F</i> (2, 100) = 2.58, <i>p</i> = .080

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. ADHD-SR = Total score of the ADHD self-rating. MWT-B IQ = Intelligence Quotient assessed with the Multiple Choice Word Fluency Test. SDS-17 = Social Desirability Scale-17. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval.

Different superscripts indicate significant differences (reported in bold face) at *p* < .05 in pairwise comparisons, whereas the letters a denote smaller sum of ranks/means.

Table 6. Detention information for inmates by group

Characteristic	APDs (<i>n</i> = 31)		INCs (<i>n</i> = 33)		Group comparisons
	<i>Mdn (IQR)</i>	95% CI	<i>Mdn (IQR)</i>	95% CI	
Duration of current detention (in months)	11.03 (10.07)	[7.17, 14.60]	9.27 (32.60)	[6.71, 23.67]	$U = 488.00, z = 0.32, p = .752$
Duration of lifetime detentions (in months)	22.00 (52.00)	[17.00, 42.91]	32.00 (66.00)	[17.00, 46.00]	$U = 503.50, z = 0.11, p = .914$
Number of lifetime detentions	2.00 (4.00)	[2.00, 2.00]	1.00 (2.00)	[1.00, 2.00]	$U = 388.50, z = 1.75, p = .081, d = 0.422$
	%		%		
Convict (vs. pretrial detainee)	45.16		54.55		$\chi^2(1) = 0.56, p = .453$

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. 95% CI = 95% bias corrected and accelerated confidence interval.

Table 7. Diagnostic information for inmates by group

Diagnosis	APDs	INCs	Group comparisons
	(<i>n</i> = 31)	(<i>n</i> = 33)	
	%	%	
Major depression (lifetime)	32.26	27.27	$\chi^2(1) = 0.19, p = .663$
Bipolar disorder (lifetime)	0.00	3.23	Fisher's Exact Test: $p = .484$
Alcohol use disorder			
Lifetime	67.74	33.33	$\chi^2(1) = 7.57, p = .006, OR = 4.20$
Current	16.13	0.00	Fisher's Exact Test: $p = .022, OR^* = 13.91$
Substance use disorder			
Lifetime	90.32	57.58	$\chi^2(1) = 8.79, p = .003, OR = 6.88$
Current	29.00	9.09	$\chi^2(1) = 4.17, p = .041, OR = 4.09$

Note. Frequencies of psychiatric disorders as assessed by the Mini International Neuropsychiatric Interview 7.0.2. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. Significant differences at $p < .05$ are reported in boldface.

* The Haldane-Anscombe correction was used for calculation of Odds Ratio due to division by zero error.

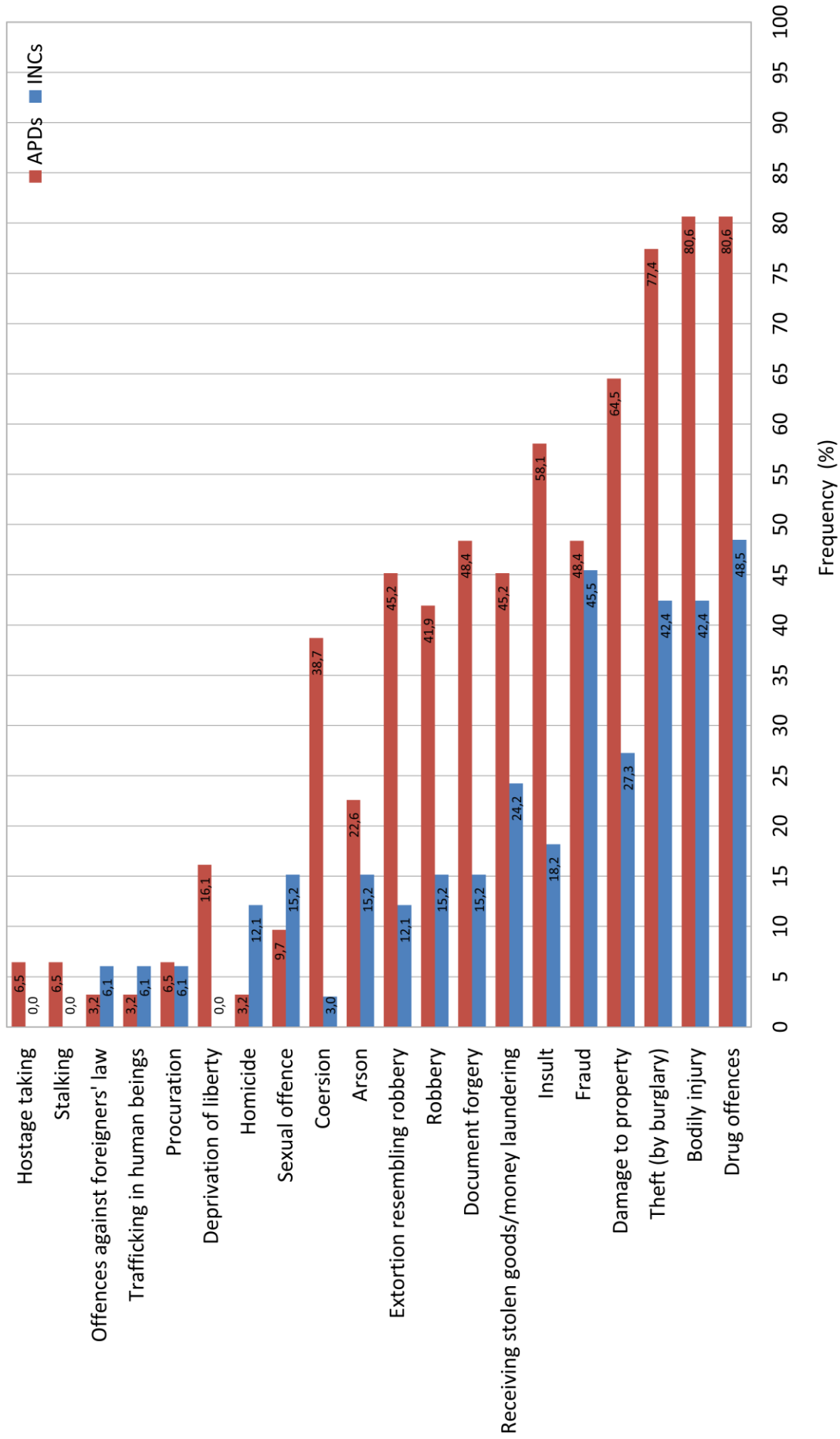


Figure 8. Self-reported lifetime offences by inmate group. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. Multiple answers possible.

3.2.2. Measures

Diagnostic assessment. To assess psychiatric disorders, the M.I.N.I. was carried out. We adapted this semi-structured interview by asking additional questions in order to assess *lifetime* SUD and AUD. Furthermore, sections B (suicidal tendencies), P (APD) and Q (borderline personality disorder) were skipped. Since diagnostic criteria for APD have not changed from DSM-IV to DSM-5 (see American Psychiatric Association, 2013), and the SCID-II offers more detailed acquisition of both, conduct disorder and adult APD symptoms, we assessed APD by the corresponding section of the SCID-II. To obtain a dimensional measure of APD (i.e. symptom severity) we did not follow skip rules but completed the whole interview with all participants. Recoded scores for each item (1 = absent was recoded to 0, 2 = subthreshold was recoded to 1, and 3 = true was recoded to 2) were added, resulting in an APD symptom severity index ranging from 0 to 14. In order to assess symptom severity of attention deficit and hyperactivity disorder (ADHD), all participants completed the ADHD self-rating (ADHD-SR; Rösler, Retz-Junginger, Retz, & Stieglitz, 2008). This questionnaire instructs participants to estimate the severity of 18 ADHD symptoms on the scales inattention and hyperactivity/impulsivity according to DSM-5. In the current study the total score was used. When applied as a screening instrument for ADHD, the authors suggest a cut-off score of 15 (sensitivity = .77, specificity = .75; Rösler et al., 2008). To rule out cognitive and linguistic impairment, all participants completed the MWT-B, for which IQ estimates are reported. As a measure of response bias the Social Desirability Scale-17 (SDS-17; Stöber, 2001) was conducted. Higher values indicate a higher degree of socially desirable response style.

Anger experience and regulation. The State-Trait Anger Expression Inventory-2 (STAXI-2; Spielberger, 1999) is a self-report questionnaire that measures intensity of state anger, frequency of anger experience (trait anger), outward physical or verbal expression of anger (anger expression-out), inward suppression of angry feelings (anger expression-in) and frequency of attempts to control existing angry feelings (anger control). Items are rated on 4-point Likert scales, with higher scores indicating more anger/AR. The STAXI-2 has proven its psychometric property within prison inmates (Etzler, Rohrman, & Brandt, 2014) and is a widely used instrument.

Habitual emotion regulation. To assess *habitual* ER, two self-report measures were used. The Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004) was completed to measure more general emotion dysregulation. It consists of the scales awareness (i.e. not attending to/acknowledging one's emotions), clarity, non-acceptance, impulse, goals (i.e. difficulties engaging in goal-directed behavior), and limited access (i.e. poor confidence in

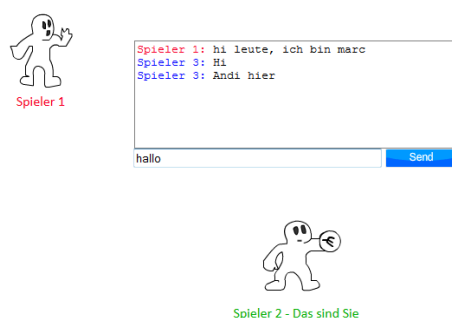
the effectiveness of one's ER). Items are rated on a 5-point Likert scale, ranging from "almost never" to "almost always". Higher scores indicate more severe difficulties. A composite score is available to reflect overall emotion dysregulation. In addition, the Cognitive Emotion Regulation Questionnaire (CERQ; Garnefski, Kraaij, & Spinhoven, 2001) was conducted to assess the frequency of cognitive ER strategy use (i.e. self-blame, blaming others, rumination, catastrophizing, acceptance of the situation, positive refocusing, putting into perspective, refocus on planning and positive reappraisal). Again, items are scored on a 5-point Likert scale.

Cyberball Aggression Task. The CAT is composed of a modified version of Cyberball 4.0 (Williams et al., 2012) and a forced-choice punishment task. The paradigm subsequently described was carefully tested in advance. Information on these prior studies is given in chapters 2.2 and 2.3.

Inducing anger. Cyberball 4.0 is a computerized virtual ball-tossing game in which the participant supposedly plays with other participants to train his mental visualization ability. In fact, the other players are faked. The original version was adapted in order to (a) simulate a chat conversation between the two other players, meant to induce anger, (b) display an error message each time the participant tries to send a chat comment on his own and (c) hinder the message from being sent. This alleged bug is embedded in a cover story to strengthen deception (see chapter 3.2.3). The game was programmed in such a way that the two other players first involve the participant by tossing him the ball and asking him questions (round 1-4). Due to the fact that the participant does not answer (alleged system error when trying to send messages), the other players are apparently becoming annoyed. As an (ostensible) result, they begin to insult the participant via (pre-programmed) chat and exclude him by not tossing him balls anymore (round 5-12). Hence, there are two conditions: baseline (inclusion via chat and ball throws, round 1-4) and anger (insults and ostracism, round 5-12). Screenshots of the conditions and the error message can be seen in Figure 9. A translation of the chat conversation is depicted in Appendix A.

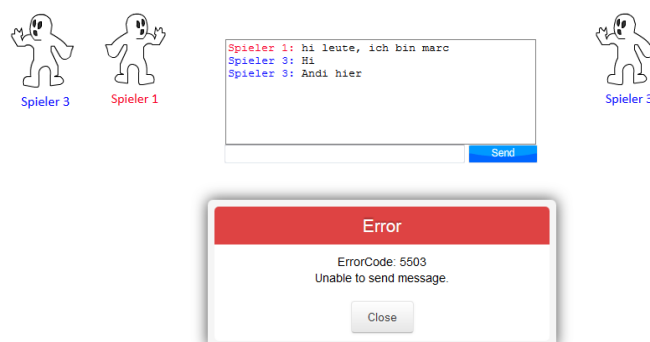
a) Baseline round 1: participant typing

Werfen Sie den Ball, indem Sie mithilfe der Maus auf den Namen oder das Bild eines Mitspielers klicken.



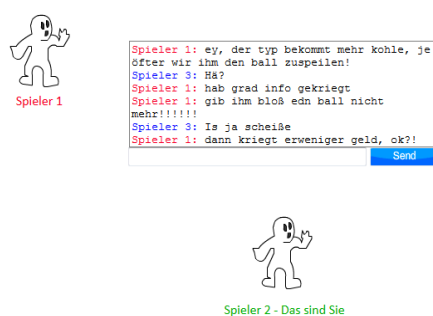
b) Baseline round 1: error message when trying to send a comment

Werfen Sie den Ball, indem Sie mithilfe der Maus auf den Namen oder das Bild eines Mitspielers klicken.



c) Anger, round 8

Werfen Sie den Ball, indem Sie mithilfe der Maus auf den Namen oder das Bild eines Mitspielers klicken.



d) Anger, round 12

Werfen Sie den Ball, indem Sie mithilfe der Maus auf den Namen oder das Bild eines Mitspielers klicken.



Figure 9. Screenshots of Cyberball.

A chat box is displayed in the middle of the screen. Each round 3-7 comments from the alleged other players are shown. The English translation of the chat conversation can be found in Appendix A. Figure a) displays the participant possessing the ball and typing “hello” in the input field. Figure b) shows the error message, which appears each time the participant tries to send a chat comment by himself. Note that the message (“hello”) is not shown in the chat box. Figure c) and d) display anger rounds, when the participant is provoked and does no longer receive the ball.

Assessing aggressive behavior. After each round of Cyberball (1 round = 15 ball tosses), the participant is forced to choose whether and to what extent he wants to punish “Player 1” by deducting money from him (i.e. behaving aggressively). The answer is given on a visual analogue scale with the end points “no money deduction for Player 1” (0) and “maximum possible money deduction for Player 1” (100). In order to enhance the need to control aggressive behavior, participants’ decision to deduct money was allegedly coupled with the threat of own money deduction in half of the trials (in fact, the decision did not affect their remuneration at

any time, all participants received full payment). The instruction was as follows: “If you deduct money from Player 1, there is the risk of losing a proportion of your own reward. The more money you deduct, the more likely it is that you will suffer a reduction in financial compensation yourself”. In the other half of the trials, the participants’ decision had no personal consequence. Slides of the forced-choice task are depicted in Appendix B.

Anger reactivity. Before the first Cyberball round of the CAT (pre) and immediately after the last punishment slide (post), participants were presented a short version of the PANAS. The adjectives irritable, hostile and upset as well as the filler items determined, enthusiastic and proud were displayed. Participants had to indicate to what extent they felt the particular emotion at the present moment. Items were scored on a 9-point Likert scale ranging from “not at all” to “extremely”. The aggregated mean of irritable, hostile and upset was used as a score for angry emotions. Cronbach’s α was $\alpha = .858$ for pre-scores of angry emotions and $\alpha = .920$ for post-scores. Furthermore, a graphic scale consisting of pictograms expressing emotional arousal, the Self-Assessment Manikin (SAM; Bradley & Lang, 1994), was used. Participants had to indicate to what extent they felt aroused in the precise moment. Scores ranged from “not at all” (1) to “very much” (9). The SAM was presented two times, each after the PANAS.

Spontaneous emotion regulation strategy use. To assess participants’ use of ER strategies during an actual regulation situation (i.e. during and after the CAT), data for the strategies positive refocusing, putting into perspective, positive reappraisal, acceptance (of the situation), experience suppression, and the rumination strategies understanding of causes and angry afterthoughts was collected. The items consisted of slightly rephrased questions from the CERQ (positive refocusing, putting into perspective, positive reappraisal, acceptance of the situation), the Heidelberg Form for Emotion Regulation Strategies (Izadpanah, Barnow, Neubauer, & Holl, 2019; experience suppression) and the Anger Rumination Scale⁹ (Sukhodolsky, Golub, & Cromwell, 2001; understanding of causes, angry afterthoughts). Responses were scored on a 5-point scale, ranging from “almost never” to “almost always”. The scales had moderate to good internal consistencies, reflected by Cronbach’s α ranging from $\alpha = .602$ to $\alpha = .844$.

3.2.3. Procedure

The current study was approved by the local ethical committee, and the Criminological Services of the corresponding Departments of Justice. During recruitment and at the beginning

⁹ The original items were translated into German by ES and modified after a discussion with LW. A bilingual researcher whose native language is English verified the accuracy of the translated text.

of the session participants were deceived as to the true purpose of the study. After the debriefing written informed consent was obtained. All participants received financial compensation for study participation.

Initially, diagnostic interviews (M.I.N.I. and SKID-II) were conducted. Then, the MWT-B and two other instruments¹⁰ were assessed. Subsequently, the CAT was introduced. Within the computerized CAT environment, the participant first completed the PANAS and the SAM. Then, further instructions and parts of the cover story were presented¹¹, both visually and orally to ensure deception. The participant was led to believe that Cyberball was conducted to train mental visualization. Therefore, he should try to visualize himself playing the game in real life. When the chat function was introduced, the investigator paused her reading of the instructions and informed the participant, that there have been some technical issues with the chat function today. However, due to the strict timeline and the fact that all participants were summoned, the task has to be carried out anyway. The participant shall try to visualize the scenario and the other players as well as possible, regardless of whether or not the chat works. After ensuring that the participant had no further questions, the first round of Cyberball started with a loading bar (“waiting for players”). The waiting time before the first game round was 7 minutes to reflect the difficulty in temporal coordination with the alleged other participants and to increase credibility. Meanwhile, the participant was given a paper-pencil version of the SDS-17 with the request to start filling in, since it is difficult to estimate how long he has to wait for the other players. The remaining waiting times before the other Cyberball rounds were between 0.1 and 17 seconds. There was a total of 12 rounds of Cyberball. For all participants, the first four rounds served as the baseline, while the last eight rounds contained the AI. After each round of Cyberball the punishment slide appeared as a measurement of aggressive behavior. Within each condition (baseline, anger), half of the decisions were coupled to a potential negative consequence, whereas the other half was not. The order of consequence (no consequence, negative consequence) was randomized. After 12 rounds of Cyberball the PANAS and the SAM were presented a second time. Then, manipulation checks were conducted: Since negative consequence should have no effect on participants’ punishing behavior if payment was not important to them, we assessed subjective importance of financial

¹⁰ The data reported in this manuscript are part of a larger study design. Participants completed additional measures. Due to data quantity and different research questions these results are reported elsewhere (see manuscript in chapter 4).

¹¹ The deceptions and instructions presented below are not complete. Due to the face-to-face interaction, instructions could not be presented fully standardized without jeopardizing credibility. Further oral additions were made depending on participants’ behavior and especially towards inmates in order to meet their doubts (e.g. regarding internet access in prison or precautions we had to make in order to obtain the permission of the Criminological Services). The guidelines for our instructions are available upon request.

compensation in order to estimate validity of the variable consequence. Furthermore, participants were asked if they read the chat comments to ensure potential efficacy of the AI. Finally, participants' use of ER strategies during the AI was assessed. Strategies were presented in randomized order. For each item, participants were asked to indicate how much the statement applied to them during and after the ball tossing game. The sequence of the CAT, including the assessment of anger reactivity and spontaneous ER strategy use, is depicted in Figure 10.

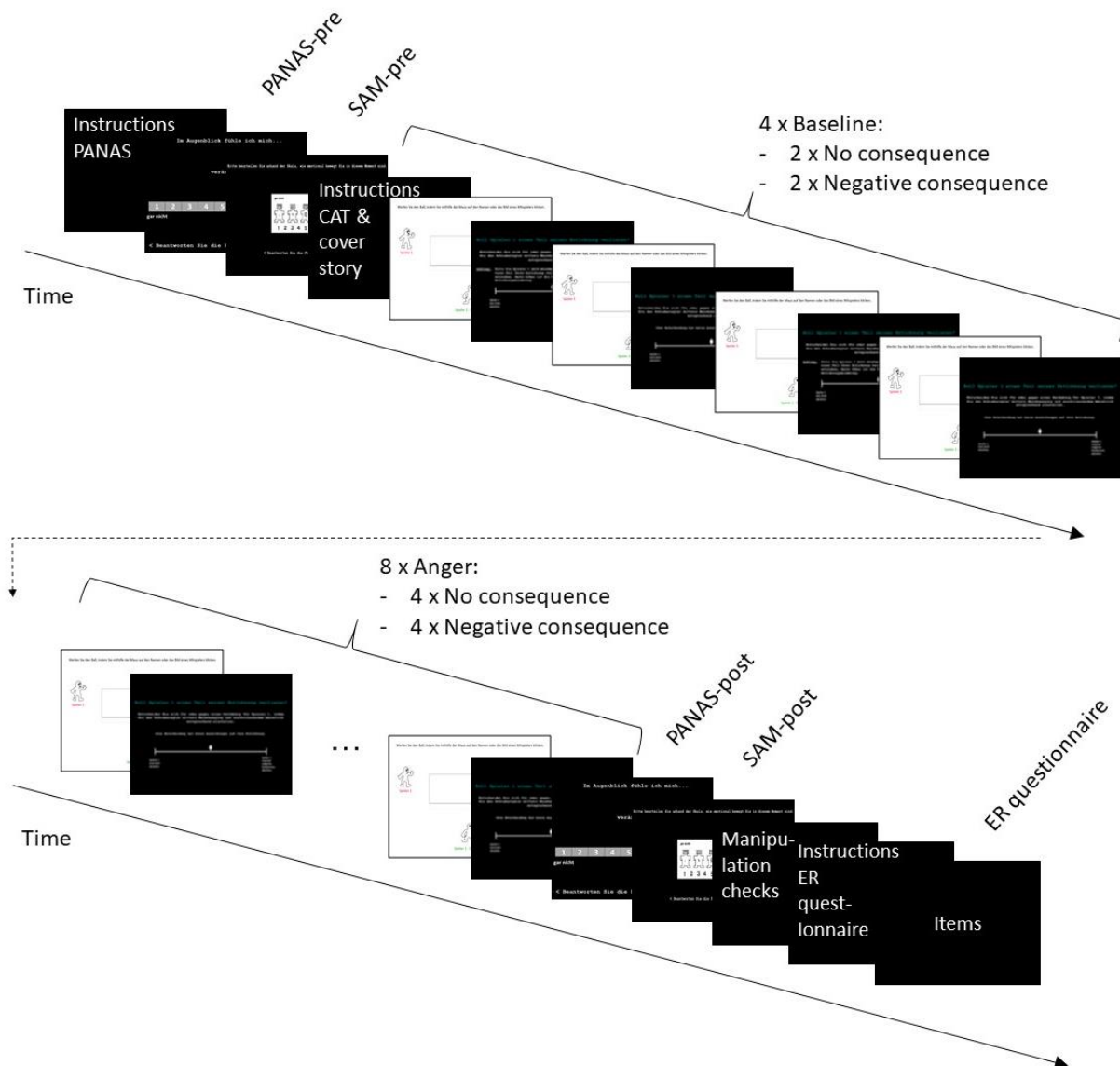


Figure 10. Sequence of the Cyberball Aggression Task, including the assessment of angry emotions, arousal and emotion regulation strategies.

PANAS = Positive and Negative Affect Schedule. SAM = Self-Assessment Manikin. ER = emotion regulation. Cyberball and punishment slides are presented alternately 12 times. The order of condition is fixed (starting with four rounds of baseline, followed by eight rounds of anger), while consequence (no consequence, negative consequence) occurs in randomized order within each condition (but equally often).

After the strategy assessment, participants' beliefs about the existence of the other players, the bogus chat and the purpose of the task were subtly probed during an interview. Credibility (deceived vs. not deceived) was conservatively assessed and included as a factor in subsequent analyses. Further information about credibility assessment and coding is given in Appendix C. Immediately thereafter, participants were fully debriefed in written form (see Appendix D). Next, two additional questionnaires¹⁰ were presented. Afterwards, participants completed the (habitual) ER questionnaires CERQ and DERS. Then, anamnestic information was assessed, whereas HCs received slightly different demographic questionnaires, as inmates were asked additional information about their detention. The ADHD-SR and the STAXI-2 were completed at the end of the session, to ensure a relatively neutral emotional state (brief check of the state anger scale).

3.2.4. Data Analysis

To investigate potential group differences in demographic characteristics, clinical symptoms and neuropsychological functioning, Pearson's chi-squared tests, Fisher's exact tests (if expected cell frequency < 5), analyses of variance (ANOVAs), and the non-parametric alternatives Kruskal-Wallis tests and Mann-Whitney *U* tests were performed. Regarding hypothesis testing, we additionally applied (mixed) ANOVAs as well as (dependent) *t*-tests. A multiple linear hierarchical regression analysis using the forced entry-method was conducted to predict antisocial symptom severity within inmates. A significance level of $\alpha = .05$ was used, with $.05 < \alpha < .10$ denoting marginal significance. In the case of heteroscedasticity and if available in SPSS, robust methods (i.e. Welch's *F* and Welch's *t*-test) were applied. In case of violations of normality, we used nonparametric analyses, wherever possible. Concerning the analysis of the CAT, we tried to approximate normality more closely by transforming the data. For the dependent variable ER strategy use results did not change when using transformed data, so we report raw data to ensure better readability. For the dependent variables punishing behavior, angry emotions and arousal, transforming data was not successful. Despite this and due to the lack of robust methods for 3- or 4-factor mixed-designs in SPSS, parametric analyses were nevertheless conducted. However, we provide 95% bias corrected and accelerated bootstrap confidence intervals (95% BCa CI) using 1000 samples for non-repeated measures means (or rather medians). Odds ratio (OR) or Cohen's *d* are offered as effect sizes following marginally significant omnibus tests (for a classification of effect sizes see Cohen, 1988). While *ds* were calculated using the freeware Psychometrica (Lenhard & Lenhard, 2016), 95% BCa CIs for means and medians and 95% CIs for regression coefficients were taken from SPSS.

Test statistics and effect sizes are reported as absolute values – direction of effects can be taken from descriptives. Missing values in questionnaires ($n = 5$) were replaced by the participants' corresponding scale means. Due to necessary data exclusions, sample sizes vary slightly across analyses.

3.3. Results

3.3.1. Anger Experience and Regulation

Regarding the STAXI-2, groups differed significantly on all (state anger: $H(2) = 12.59$, $p = .002$, trait anger: $H(2) = 11.48$, $p = .003$, anger expression-out: $H(2) = 16.67$, $p < .001$, anger expression-in: $F(2, 100) = 3.35$, $p = .039$) but one scale (anger control, $H(2) = 1.82$, $p = .402$). Descriptive data can be seen in Table 8.

Table 8. Groups' anger regulation as evident by the State-Trait Anger Expression Inventory-2

Scale	APDs ($n = 31$)			INCs ($n = 33$)			HCs ($n = 39$)		
	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI
State anger*		15.00^b (2.00)	[15.00, 15.00]		15.00^b (3.00)	[15.00, 15.00]		15.00^a (0.00)	n.a.
Trait anger		20.00^b (7.00)	[18.00, 24.00]		18.00^a (6.00)	[16.00, 20.00]		16.00^a (4.00)	[16.00, 16.50]
Anger expression-out		14.00^b (6.00)	[13.00, 16.00]		10.00^a (4.00)	[9.00, 10.00]		11.00^a (5.00)	[10.00, 11.00]
Anger expression-in	20.39^b (5.35)		[18.49, 22.32]	17.39^a (5.40)		[15.60, 19.30]	17.15^a (6.03)		[15.27, 19.08]
Anger control		28.00 (9.00)	[26.00, 33.00]		32.00 (8.00)	[31.00, 32.00]		31.00 (8.00)	[30.00, 32.00]

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval. Different superscripts indicate significant differences (reported in bold face) at $p < .05$ in pairwise comparisons. The letters a denote smaller sum of ranks/means.

* Results for this scale are to be interpreted with caution (floor effects).

In line with our hypothesis, APDs reported higher state anger than HCs, $U = 425.00$, $z = 3.35$, $p = .001$, $d = 0.524$. Yet, INCs also reported increased state anger compared to HCs, $U = 449.00$, $z = 3.39$, $p = .001$, $d = 0.536$, while inmates did not differ from each other, $U = 505.50$, $z = 0.10$, $p = .923$. However, results for state anger should be interpreted carefully due to floor effects and different shapes of distribution between groups. As expected, APDs reported higher

trait anger than both, HCs, $U = 324.50$, $z = 3.33$, $p = .001$, $d = 0.862$, and INCs, $U = 353.50$, $z = 2.13$, $p = .033$, $d = 0.550$. No differences between INCs and HCs were found concerning trait anger, $U = 533.50$, $z = 1.26$, $p = .209$. Looking at AR, the same pattern of results was found for anger expression-out and anger expression-in: As expected, APDs reported increased anger expression as compared to HCs (anger expression-out: $U = 329.00$, $z = 3.28$, $p = .001$, $d = 0.845$, anger expression-in: $t(68) = 2.34$, $p = .022$, $d = 0.564$), but also as compared to INCs (anger expression-out: $U = 232.50$, $z = 3.77$, $p < .001$, $d = 1.061$, anger expression-in: $t(62) = 2.23$, $p = .030$, $d = 0.558$). INCs and HCs did not differ with respect to anger expression-out, $U = 576.00$, $z = 0.77$, $p = .440$, or anger expression-in, $t(70) = 0.18$, $p = .860$.

Taken together, with the exception of state anger INCs indicated no abnormalities in anger experience and regulation compared to HCs. In contrast, APDs reported increased (trait) anger, as well as increased anger expression- in and expression-out than both, HCs and INCs.

3.3.2. *Habitual Emotion Regulation*

Regarding the DERS, groups differed with marginal significance on the total score, $H(2) = 4.96$, $p = .084$. Follow-up tests indicated that this was due to APDs reporting increased emotion dysregulation compared to HCs, $U = 418.50$, $z = 2.20$, $p = .028$, $d = 0.545$, but not INCs, $U = 440.50$, $z = 0.95$, $p = .340$, while INCs and HCs did not differ from each other, $U = 533.50$, $z = 1.24$, $p = .214$. The only significant group difference on scale level was found for impulse, $H(2) = 7.71$, $p = .021$. APDs reported increased difficulties to refrain from impulsive behavior when distressed as compared to both, INCs, $U = 336.50$, $z = 2.37$, $p = .018$, $d = 0.615$, and HCs, $U = 398.50$, $z = 2.45$, $p = .014$, $d = 0.609$. Again, no group differences between INCs and HCs occurred, $U = 600.00$, $z = 0.50$, $p = .620$. As to the other scales, groups reported comparable difficulties (awareness: $F(2, 100) = 1.17$, $p = .314$, clarity: $H(2) = 1.45$, $p = .484$, non-acceptance: $H(2) = 0.52$, $p = .773$, goals: $H(2) = 2.56$, $p = .278$, limited access: $H(2) = 4.55$, $p = .103$). Descriptives for the DERS are depicted in Table 9.

Table 9. Habitual difficulties in emotion regulation by group

DERS scale	APDs (<i>n</i> = 31)			INCs (<i>n</i> = 33)			HCs (<i>n</i> = 39)		
	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI
Awareness	17.97 (5.49)		[16.11, 20.11]	17.64 (5.44)		[15.79, 19.69]	16.28 (3.40)		[15.10, 17.57]
Clarity		10.00 (4.00)	[10.00, 10.00]		10.00 (4.00)	[9.00, 10.00]		9.00 (4.00)	[9.00, 9.00]
Non- acceptance		11.00 (6.00)	[10.00, 14.00]		12.00 (5.50)	[11.00, 12.00]		11.00 (8.00)	[10.50, 11.00]
Impulse		10.00^b (6.00)	[9.00, 14.00]		9.00^a (4.00)	[9.00, 9.00]		9.00^a (4.00)	[8.00, 10.00]
Goals		12.00 (5.00)	[11.50, 12.00]		11.00 (6.00)	[10.00, 13.00]		10.00 (6.00)	[9.00, 10.00]
Limited access		15.00 (7.00)	[14.50, 15.00]		13.00 (6.00)	[13.00, 13.00]		12.00 (6.00)	[10.00, 14.00]
Total score		80.00^b (19.00)	[77.50, 83.00]		75.00 ^{a,b} (23.00)	[71.50, 78.00]		68.00^a (22.00)	[64.00, 70.00]

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. DERS = Difficulties in Emotion Regulation Scale. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval. Different superscripts indicate significant differences (reported in bold face) at $p < .05$ in pairwise comparisons, whereas the letters a denote smaller sum of ranks/means.

Regarding the CERQ, analyses yielded group differences for self-blame, $H(2) = 13.13$, $p = .001$, catastrophizing, $H(2) = 16.84$, $p < .001$, and acceptance, $F(2, 62.85) = 6.23$, $p = .003$. Descriptives for the CERQ can be seen in Table 10.

Table 10. Habitual emotion regulation strategy use by group

	APDs (<i>n</i> = 31)			INCs (<i>n</i> = 33)			HCs (<i>n</i> = 39)		
	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI
Self-blame		8.00^b (4.00)	[7.00, 9.00]		9.00^b (3.00)	[9.00, 9.00]		6.00^a (5.00)	[5.50, 6.00]
Blaming others		6.00^b (3.00)	[5.00, 6.50]		4.00^a (3.00)	[3.00, 5.00]		5.00 (2.00)	[4.00, 5.50]
Rumination	7.94 (2.42)		[7.13, 8.80]	8.30 (2.31)		[7.51, 9.11]	7.46 (2.76)		[6.65, 8.33]
Catastro- phizing		7.00^b (2.00)	[6.00, 7.00]		6.00^b (3.50)	[5.50, 7.00]		5.00^a (2.00)	[4.00, 6.00]
Acceptance (of the situation)	11.45^b (3.14)		[10.39, 12.46]	11.24^b (1.94)		[10.59, 11.91]	9.51^a (2.58)		[8.61, 10.38]
Positive Refocusing	8.23 (3.36)		[7.05, 9.55]	9.21 (2.77)		[8.20, 10.20]	8.10 (2.99)		[7.15, 8.98]
Putting into perspective	10.03 (3.77)		[8.57, 11.40]	10.45 (2.11)		[9.72, 11.23]	9.54 (2.85)		[8.64, 10.48]
Refocus on planning		12.00 (4.00)	[9.00, 13.00]		12.00 (3.00)	[10.00, 13.00]		11.00 (3.00)	[10.00, 12.00]
Positive reappraisal	9.77 ^{a,b} (3.53)		[8.49, 11.11]	11.18^b (2.93)		[10.17, 12.28]	9.85^a (2.51)		[9.03, 10.66]

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. CERQ = Cognitive Emotion Regulation Questionnaire. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval. Different superscripts indicate significant differences (reported in bold face) at $p < .05$ in pairwise comparisons, whereas the letters a denote smaller sum of ranks/means.

Pairwise comparisons revealed the same pattern for all three strategies: APDs and INCs reported increased use of self-blame (APDs: $U = 344.00$, $z = 3.41$, $p = .001$, $d = 0.792$, INCs: $U = 394.00$, $z = 2.51$, $p = .012$, $d = 0.705$), catastrophizing (APDs: $U = 287.50$, $z = 3.79$, $p < .001$, $d = 1.002$, INCs: $U = 380.00$, $z = 3.02$, $p = .002$, $d = 0.750$) and acceptance (APDs: $t(68) = 2.84$, $p = .006$, $d = 0.682$, INCs: $t(70) = 3.16$, $p = .002$, $d = 0.749$) compared to HCs, while not differing from each other (self-blame: $U = 431.55$, $z = 1.08$, $p = .278$, catastrophizing: $U = 433.50$, $z = 1.06$, $p = .290$, acceptance: $t(49.37) = 0.32$, $p = .752$). Moreover, marginally significant group differences were found for blaming others, $H(2) = 5.20$, $p = .074$, and reappraisal, $F(2, 61.81) = 13.13$, $p = .097$. Interestingly, APDs reported significantly increased use of blaming others compared to INCs, $U = 361.50$, $z = 2.05$, $p = .040$, $d = 0.521$, and marginally significant increased use compared to HCs, $U = 455.00$, $z = 1.80$, $p = .073$, $d = 0.432$. INCs and HCs did

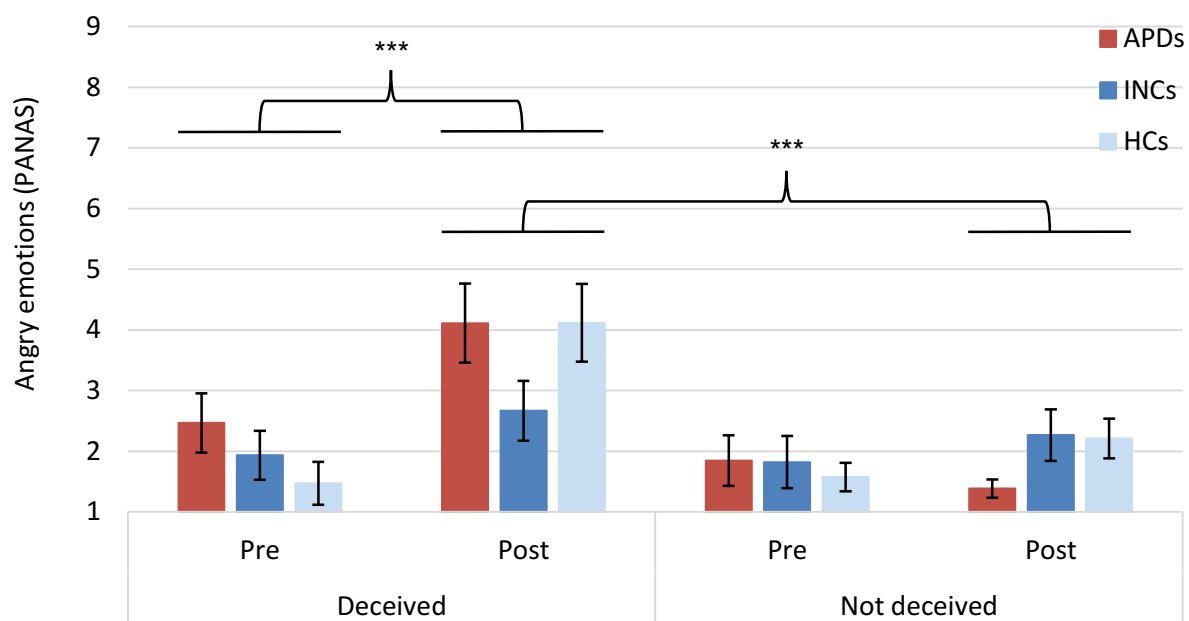
not differ regarding blaming others, $U = 578.00$, $z = 0.76$, $p = .449$. Concerning positive reappraisal, INCs surprisingly reported increased use compared to HCs, $t(70) = 2.08$, $p = .041$, $d = 0.491$ and APDs, $t(62) = 1.74$, $p = .087$, $d = 0.436$, though the latter difference was only marginally significant. HCs and APDs did not differ from each other regarding positive reappraisal, $t(52.29) = 0.96$, $p = .924$. With respect to rumination, $F(2, 100) = 1.01$, $p = .369$, and the rather adaptive strategies positive refocusing, $F(2, 100) = 1.36$, $p = .260$, putting into perspective, $F(2, 61.95) = 1.21$, $p = .305$, and refocus on planning, $H(2) = 0.79$, $p = .674$, all groups reported comparable strategy use.

3.3.3. Spontaneous Emotion Regulation: Cyberball Aggression Task

Manipulation checks indicated that all participants read the chat comments (92.8% completely and 7.2% in part). Of note, only a minority reported that receiving full financial compensation was a worthy aspiration (48.5% vs. 51.5%, groups did not differ significantly, $\chi^2(2) = 4.56$, $p = .102$), thus questioning the validity of our variable consequence. Credibility assessment indicated that 48.5% of participants had been deceived about the true nature of the task, whereas 51.5% had (partly) seen through the cover story sometime during the CAT. No group differences occurred regarding credibility, $\chi^2(2) = 0.69$, $p = .708$.

Angry emotions and arousal. We conducted separate mixed design ANOVAs with the within-subjects factor time (pre, post) and the between-subjects factors group (APDs, INCs, HCs) and credibility (deceived, not deceived) on angry emotions and arousal. Means are seen in Figure 11.

a) Angry emotions



b) Arousal

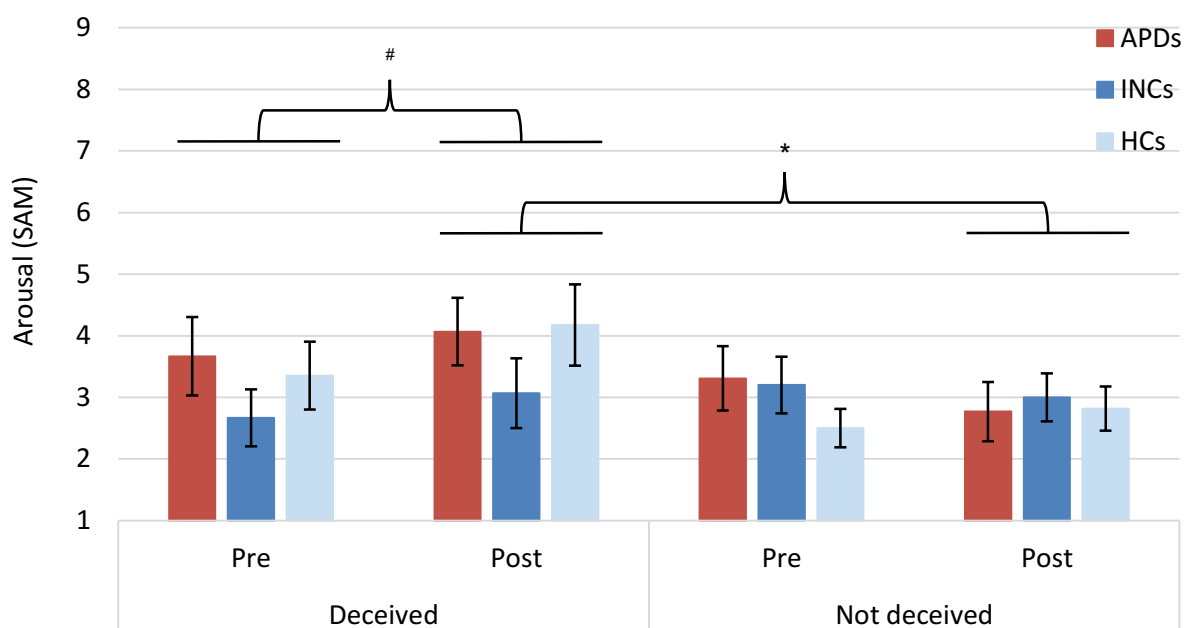


Figure 11. Self-reported angry emotions and arousal among groups depending on time (pre, post) and credibility (deceived, not deceived).

APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. PANAS = Positive and Negative Affect Schedule. SAM = Self-Assessment Manikin. Error bars represent standard errors of the means. Asterisks indicate significance level of effects. Sample sizes were $n = 28$ for APDs, $n = 30$ for INCs, and $n = 39$ for HCs.

$p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$

Concerning angry emotions, ANOVA showed a significant main effect of time, $F(1, 91) = 17.44, p < .001$, as well as a significant main effect of credibility, $F(1, 91) = 11.46, p = .001$, qualified by a significant interaction of Time \times Credibility, $F(1, 91) = 10.63, p = .002$. Furthermore, a marginally significant effect of Time \times Group occurred, $F(2, 91) = 2.66, p = .075$. No other effects were found (group: $F(2, 91) = 0.32, p = .729$, Credibility \times Group: $F(2, 91) = 1.98, p = .144$, Time \times Credibility \times Group: $F(2, 91) = 1.68, p = .193$). Regarding the Time \times Credibility interaction, pairwise comparisons revealed that angry emotions prior to the CAT did not differ between deceived and not deceived participants, $t(95) = 0.70, p = .484$. By contrast, and as could be expected, participants who believed in the cover story reported higher angry emotions after the CAT than participants who saw through the deception, $t(72.81) = 4.02, p < .001, d = 0.829$. In line with that, only deceived participants reported a significant increase of angry emotions due to the CAT, $t(46) = 4.21, p < .001, d = 0.776$, while not deceived participants did not: $t(49) = 1.35, p = .183$. Still, when examining the marginally significant Time \times Group interaction, it becomes evident that the effect of time (i.e. the AI) was strong enough within INCs and HCs to even persist when merging deceived and not deceived participants into one group, INCs: $t(29) = 2.07, p = .047, d = 0.402$, HCs: $t(38) = 4.15, p < .001, d = 1.003$. However, within APDs the increase of angry emotions was no longer significant, $t(27) = 1.18, p = .247$. Again, no differences between groups were found, neither before, $F(2, 55.51) = 1.55, p = .220$, nor after the CAT, $F(2, 94) = 0.62, p = .540$.

Regarding arousal, analyses yielded a marginally significant interaction of Time \times Credibility, $F(1, 91) = 2.99, p = .087$. No other effects emerged (time: $F(1, 91) = 1.04, p = .312$, credibility: $F(1, 91) = 2.49, p = .118$, group: $F(2, 91) = 0.52, p = .595$, Time \times Group: $F(2, 91) = 1.02, p = .367$, Credibility \times Group: $F(2, 91) = 1.32, p = .273$, Time \times Credibility \times Group: $F(2, 91) = 0.11, p = .897$). Follow-up tests again indicated that prior to the CAT all participants reported similar arousal, regardless of later deception, $t(95) = 0.79, p = .429$, whereas credibility indeed influenced arousal-ratings after the CAT: Participants who had been deceived, reported higher arousal ratings after the CAT than participants who had not been deceived, $t(80.42) = 2.23, p = .028, d = 0.460$. Accordingly, increase of arousal from pre to post was marginally significant only for participants who believed in the cover story, $t(46) = 1.71, p = .095, d = 0.263$, whereas arousal did not change for participants who saw through the deception, $t(49) = 0.28, p = .783$.

Taken together, results indicate that our AI was successful – but only for those participants who believed in the cover story. Contrary to expectations, groups reported comparable arousal and anger reactivity.

Aggressive behavior. To examine group differences in aggressive behavior, a Group (APDs, INCs, HCs) \times Credibility (deceived, not deceived) \times Condition (baseline, anger) \times Consequence (no consequence, negative consequence) mixed-design ANOVA on punishing behavior (i.e. money deduction) was conducted. Means and standard deviations are shown in Table 11. The analysis yielded significant main effects of condition, $F(1, 91) = 32.69, p < .001$, and credibility, $F(1, 91) = 12.51, p = .001$, as well as a marginally significant main effect of group, $F(2, 91) = 2.85, p = .063$. However, these effects were qualified by a significant interaction of Condition \times Credibility, $F(1, 91) = 16.65, p < .001$, and Condition \times Group, respectively, $F(1, 91) = 4.38, p = .015$ (see Figure 12). There were no other significant effects, all $F_s < 2.25$, all $p_s > .113$ (details are shown in Appendix E). Hence, and in line with participants' low valuation of the study reward (see above), consequence did not influence their decision to deduct money at any time.

Table 11. Mean punishment by credibility and group, depending on condition and consequence

Credibility and group	<i>n</i>	Baseline				Anger			
		No consequence		Negative consequence		No consequence		Negative consequence	
		<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>
Deceived									
APDs	15	23.90	(32.87)	24.77	(32.85)	37.70	(34.84)	48.27	(42.54)
INCs	15	9.43	(18.29)	8.83	(17.08)	23.28	(30.61)	19.57	(22.39)
HCs	17	13.50	(21.05)	8.53	(15.40)	44.76	(39.88)	44.96	(36.61)
Not deceived									
APDs	15	17.88	(33.55)	9.31	(14.17)	13.81	(27.83)	12.69	(29.11)
INCs	15	4.27	(7.30)	3.13	(8.23)	3.82	(10.18)	8.27	(21.20)
HCs	22	6.32	(16.24)	5.68	(16.32)	15.74	(29.61)	13.88	(28.79)

Note. Mean Punishment (Range = 100) depending on the within-subjects factors condition (baseline, anger) and consequence (no consequence, negative consequence) and the between-subjects factors credibility (deceived, not deceived) and group (APDs, INCs, HCs). APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls.

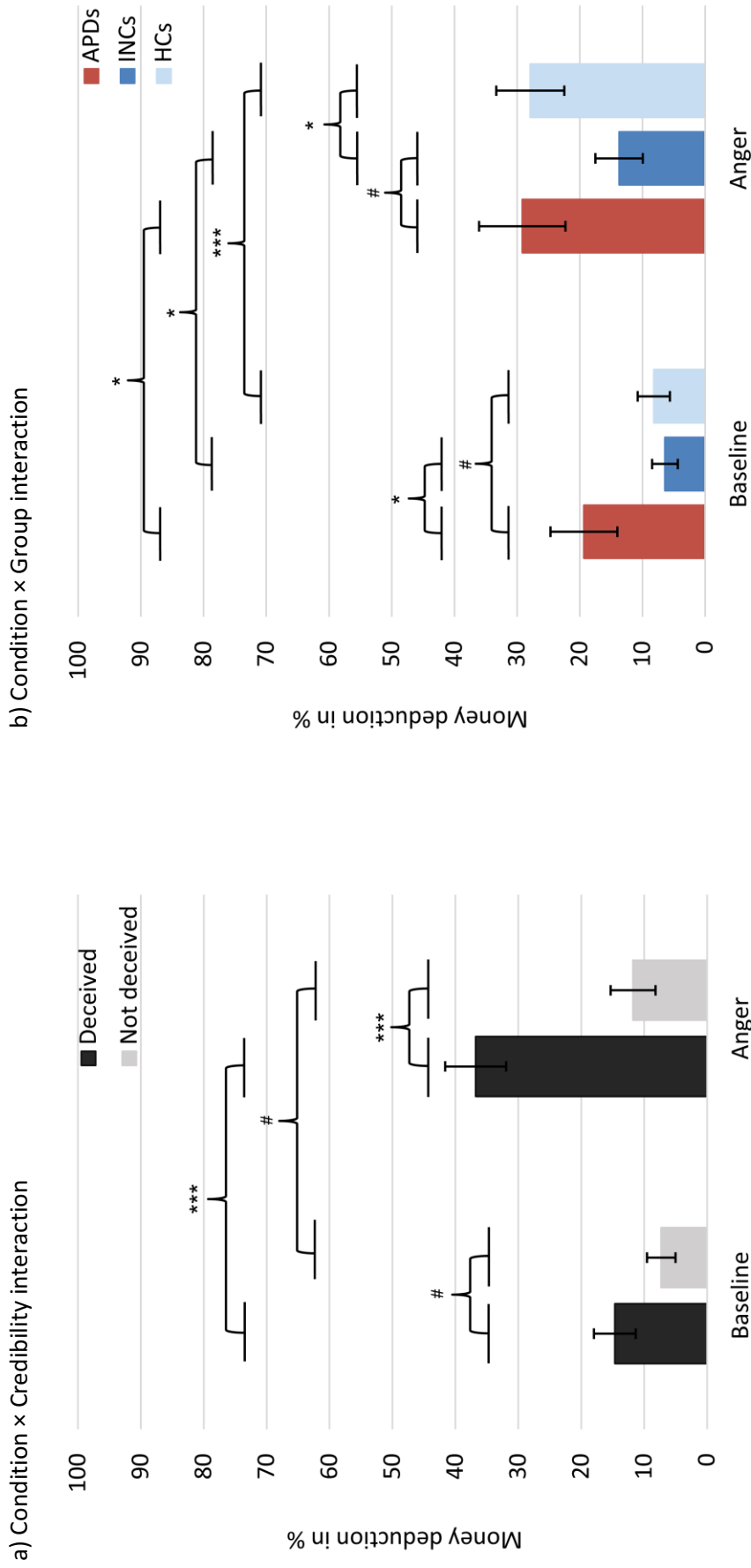


Figure 12. Participants' mean punishing behavior (i.e. money deduction).

Figure 12a depicts the Condition (baseline, anger) x Credibility (deceived, not deceived) interaction, with $n = 47$ deceived and $n = 50$ not deceived participants. Figure 12b shows the Condition x Group (APDs, INCs, HCs) interaction with $n = 28$ APDs, $n = 30$ INCs, and $n = 39$ HCs. APDs = Inmates with antisocial personality disorder. INCs = Inmate controls. HCs = Healthy controls. Error bars represent standard errors of the means. Asterisk indicate significance level of effects. # $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Regarding the Condition \times Credibility interaction, follow-up tests revealed that deceived participants punished significantly more in response to the anger condition than in response to the baseline condition, $t(46) = 5.65, p < .001, d = 0.741$. Within not deceived participants this effect of the AI was only marginally significant, $t(49) = 1.92, p = .061, d = 0.259$. Correspondingly, credibility had only a marginally significant effect in baseline, $t(82.20) = 1.83, p = .071, d = 0.376$, but a significant effect in anger, $t(85.62) = 4.17, p < .001, d = 0.854$. The interaction is depicted in Figure 12a.

With respect to the Condition \times Group interaction, follow-up tests indicated that the effect of condition was present across groups (APDs: $t(27) = 2.06, p = .049, d = 0.362$, INCs: $t(29) = 2.15, p = .040, d = 0.336$, HCs: $t(38) = 4.93, p < .001, d = 0.777$; see Figure 12b). However, the magnitude of the effect varied: When comparing groups with respect to change scores (i.e. subtracting baseline punishment from anger punishment), HCs surprisingly showed the largest increase in punishing behavior due to the AI, as evident by a significantly higher difference score than INCs, $t(66.95) = 2.36, p = .021, d = 0.552$. However, the difference to APDs did not reach significance, $t(65) = 1.59, p = .117$. Furthermore, there was no difference between INCs and APDs, $t(56) = 0.43, p = .666$. To further elucidate potential group differences within conditions, we conducted separate ANOVAs. While the effect of group was marginally significant in baseline, $F(2, 54.78) = 2.51, p = .090$, it was significant within anger, $F(2, 57.72) = 3.26, p = .045$. Interestingly, different patterns of results occurred: in baseline, APDs punished the most, reflected by a significant difference compared to INCs, $t(34.96) = 2.26, p = .031, d = 0.608$, and a marginally significant difference compared to HCs, $t(39.37) = 1.88, p = .067, d = 0.508$. INCs and HCs did not differ in their baseline punishing behavior, $t(67) = 0.51, p = .593$. However, and due to HCs' aforementioned comparatively strong reaction to the AI, APDs' and HCs' punishing behavior did no longer differ in the anger rounds, $t(65) = 0.15, p = .884$. Here, surprisingly, INCs punished the least, as evident by a significant difference to HCs, $t(64.18) = 2.14, p = .036, d = 0.489$, and a marginally significant difference to APDs, $t(42.30) = 1.96, p = .056, d = 0.525$.

Taken together, the AI was also successful on a behavioral level: Participants of all groups primarily punished when they had a reason to do so, i.e. when they were provoked (condition = anger), and especially when they did not see through the deception (credibility = deceived). Moreover, and independent of credibility, abnormalities in inmates' punishing behavior were revealed, though different than expected: APDs showed the most punishing behavior without (obvious) reason, i.e. during baseline. By contrast, they showed comparable reactive aggression as HCs. Alterations within INCs were revealed during the AI: INCs showed

lower overall punishment in anger rounds and reduced increase in aggressive behavior due to the provocation as compared to HCs.

Emotion regulation strategy use during and after the anger induction. After visual inspection of reaction time (RT) distributions, all items with RT < 1.5 seconds were excluded from further analysis. A 2 × 3 (Credibility × Group) ANOVA was conducted for each ER strategy. Analysis yielded a main effect of credibility for afterthoughts, $F(1, 91) = 6.00, p = .016, d = 0.464$, and understanding causes, $F(1, 91) = 4.41, p = .039, d = 0.442$, while no other effect of credibility was found (all F s < 1.74 all p s > .191, see Appendix F). Participants who believed in the cover story reported to have ruminated significantly more (afterthoughts: $M = 2.34, 95\% \text{ BCa CI } [2.11, 2.57], SD = 0.83$, understanding causes: $M = 2.44, 95\% \text{ BCa CI } [2.15, 2.71], SD = 0.92$) than participants who saw through the deception (afterthoughts: $M = 1.96, 95\% \text{ BCa CI } [1.74, 2.22], SD = 0.81$, understanding causes: $M = 2.07, 95\% \text{ BCa CI } [1.88, 2.27], SD = 0.75$). Of note, no group effect occurred for any ER strategy (all F s < 2.10, all p s > .129, see Appendix F). Hence, APDs, INCs and HCs reported similar use of ER strategies during an actual regulation situation.

3.3.4. *Predicting Symptom Severity of Antisocial Personality Disorder*

Beyond group differences in ER it was of interest whether ER is also capable of predicting APD symptom severity within the often as homogenous considered group of (incarcerated) offenders. Such a result would provide further evidence for APD as a disorder of ER. Variables for which we found (marginal) significant group differences between APDs and INCs were entered as predictors in a hierarchical multiple linear regression analysis. Data from the CAT was not included in the regression analysis due to diminished sample size when excluding not deceived participants. As it is known that SUD/AUD is a predictor of APD (Fossati et al., 2007), that APD declines in age (e.g. Baliouisis et al., 2019; Black et al., 2010) and that it is associated with low verbal IQ (Sedgwick et al., 2017), these variables were entered in the first step of the regression (model 1: basic model). The second step consisted of ER variables (model 2: ER model). All analyses are based on data from $n = 64$ inmates ($n = 31$ APDs, and $n = 33$ INCs).

Five standardized residuals exceeded +/- 2 SD in the first step, only two in the second step, which is statistically expectable (95% of the sample should lie within +/- 2 SD). No outliers above +/- 3 SD were identified. Combined information on Cook Distance (all cases < 1), Mahalanobis Distance (model 1: two cases greater than the critical value of 7.81, model 2: three cases greater than the critical value of 18.31), centered leverage (all leverage values within

twice the average leverage; see Hoaglin & Welsch, 1978), DFBeta (all absolute values < 1) and the covariance ratio (all cases within the critical interval [0.48, 1.52]) indicated no unduly influential cases but suggested a quite reliable model (see Field, 2013). Linear relationship of the variables, linearity and homoscedasticity of the residuals and normality of the residuals were confirmed by visual inspection (i.e. partial regression plots, standardized predicted values vs. standardized or rather studentized residuals, histograms and p-p-plots). The Durbin-Watson test statistic for the final model was $d = 1.90$ and lay within the proposed boundary [1.27, 1.96] (see Savin & White, 1977), thus indicating independent residuals. No correlation coefficient was above the $r = .80$ mark (only two correlations greater than $r = .60$), no variance inflation factor exceeded a value of 10 (max = 3.46) and no tolerance level was below 0.2 (min = 0.29), suggesting that multicollinearity was no major threat to the current model (Field, 2013).

Regression coefficients and model fit are depicted in Table 12. Both models were significant, model 1: $F(3, 60) = 6.47, p = .001$, model 2: $F(10, 53) = 3.86, p = .001$, so was the change score, $\Delta F(7, 53) = 2.31, p = .039$. Hence, ER accounted for additional variance in APD symptomatology beyond basic variables. The final model explained 42% of the total variance, indicating a high goodness of fit. However, of basic variables only SUD/AUD significantly contributed to predicting APD symptom severity, while only reappraisal and anger expression-in turned out as (marginally) significant predictors among ER variables.

Table 12. Hierarchical multiple regression analysis predicting antisocial symptom severity within inmates

Predictor	APD symptom severity									
	Model 1: Basic model					Model 2: Emotion regulation				
	<i>B</i>	95% CI	SE <i>B</i>	β	<i>p</i>	<i>B</i>	95% CI	SE <i>B</i>	β	<i>p</i>
Constant	11.41	[2.39, 20.50]	4.53		.014	7.17	[-3.60, 17.94]	5.37		.188
Lifetime SUD/AUD ^a	3.32	[1.34, 5.30]	0.99	.39	.001	3.00	[1.04, 4.96]	0.98	.35	.003
Age	-0.01	[-0.09, 0.07]	0.04	-.04	.762	-0.03	[-0.11, 0.05]	0.04	-.09	.462
MWT-B	-0.07	[-0.17, 0.02]	0.05	-.19	.138	-0.02	[-0.12, 0.07]	0.05	-.06	.655
Trait anger						-0.04	[-0.26, 0.18]	0.11	-.06	.738
Anger expression-out						0.01	[-0.28, 0.29]	0.14	.01	.965
Anger expression-in						<i>0.13</i>	<i>[-0.02, 0.29]</i>	<i>0.08</i>	<i>.21</i>	<i>.095</i>
Emotion dysregulation						-0.02	[-0.09, 0.06]	0.04	-.07	.686
Impulse Control						0.24	[-0.10, 0.59]	0.17	.27	.165
Blaming others						0.13	[-0.23, 0.49]	0.18	.09	.476
Positive reappraisal						-0.36	[-0.64, -0.09]	0.14	-.35	.012
Modell summary										
<i>R</i> ²	.24					.42				
ΔR^2						.18				

Note. APD = Antisocial Personality Disorder. 95% CI = 95% confidence interval. *B* = unstandardized regression coefficient. β = standardized regression coefficient. *N* = 64. Boldface indicates significant variables at $p < .05$, italics indicate marginal significant variables at $p < .10$. Multiple regressions were conducted using forced entry with listwise deletion.

^a Dummy code for no lifetime substance or alcohol use disorder vs. lifetime substance or alcohol use disorder.

3.4. Discussion

The current study comprehensively investigated APDs' AR and ER, both habitually as well as spontaneously and compared to two different control groups – INCs and age and education-matched HCs. Besides self-reports we also used a more objective approach and assessed aggressive behavior with a newly developed AI paradigm.

In line with hypotheses, APDs reported chronic anger experience and impairments in habitual AR. Of note, these deficits were found both in comparison to HCs *and* INCs. Disturbed AR was evident by increased use of maladaptive strategies (anger suppression and expression) rather than decreased use of adaptive strategies (anger control). These results mainly replicate previous findings (Timmermann et al., 2017; Yavuz et al., 2016) and expand them by emphasizing the significance of the psychiatric diagnosis of APD, since INCs, in contrast to APDs, reported a normal AR pattern. Regarding the state anger differences between inmates and non-incarcerated HCs, findings should be interpreted cautiously, due to floor effects. Nevertheless, this result would not be surprising, considering the special life circumstances (i.e. incarceration) of both inmate groups (see also Velotti et al., 2017).

Of note, the pattern of impairments among APDs continued when broadening from AR to more general ER: APDs, as opposed to INCs, reported overall emotion dysregulation compared to HCs. Thus, prior research's failure to detect ER impairments in offenders could be due to the fact that psychiatric diagnoses were not considered and offenders with and without APD were merged (Garofalo et al., 2018; Gillespie et al., 2018). With respect to the specific skills underlying an adaptive ER, APDs reported impulse control difficulties compared to HCs and INCs. Given that groups differed with respect to prevalence of SUD/AUD, one might object that impulse control difficulties are due to addiction, not APD. Since more than half of all INCs also exhibited lifetime SUD/AUD, whereas none of HCs did, but no differences between these two groups were found, this objection seems unlikely. Concerning ADHD symptomatology, inmates did not differ from each other, while symptom severity was below cutoff. Hence, impulse control difficulties cannot be (fully) explained by ADHD either and represent a distinctive feature within APDs. The present null findings regarding the ER abilities awareness, clarity, non-acceptance, goals and limited access are difficult to interpret. When looking at CIs, it seems possible that existing differences between APDs and HCs were not discovered due to insufficient power.

Regarding ER strategy use, results are not unambiguous. APDs, but also INCs, reported increased *habitual* use of self-blame, catastrophizing and acceptance (of the situation). However, it is possible that these similarities primarily reflect inmates' life situation: Dealing

with one's index offence(s) and detention would almost inevitably lead to admitting one's guilt (i.e. self-blame), becoming aware of the implications of imprisonment (i.e. catastrophizing) and realizing that there is no way to change the situation (i.e. acceptance). Hence, aforementioned results might depict a rather normal reaction to an abnormal situation instead of being dysfunctional ER in the narrower sense. Interestingly, differences between inmates were found for the strategy of blaming others. APDs' increased use of this strategy possibly reflects a rather external attributional style, which might be particularly adverse in the context of their offences. Surprisingly, and once again in contrast to Gillespie et al. (2018), INCs reported increased use of reappraisal compared to HCs. However, when looking at *spontaneous* ER during the AI, there were no group differences for this strategy (or any other strategy). Hence, future research has to clarify, whether or not reappraisal is more frequently used among INCs and if so, whether it is indeed functional (for the disconnect between reappraisal use and success see e.g. McRae & Gross, 2020) or rather reflects a dysfunctional overregulation (Robertson et al., 2012). While discrepancies between results in habitual and spontaneous ER are not per se unusual in ER research (e.g. Schreiner, Joormann, & Wolkenstein, 2020), methodological reasons might be responsible for the absence of any group differences: Original items for each ER strategy had to be slightly rephrased to adapt to the context of our AI. Therefore, items could have lacked validity (see also internal consistencies). Behavioral observations further suggest that satisficing due to symptoms of fatigue and/or low motivation may have been a problem here (see Matjašič, Vehovar, & Manfreda, 2018).

Remarkably, maladaptive ER (and more specifically: decreased use of reappraisal and increased use of (anger) suppression) explained variability in inmates' APD symptom severity even when controlling for age, verbal intelligence and SUD/AUD. In other words, ER impairments within inmates can neither be (fully) explained by offending per se, the situational context due to incarceration or frequent comorbidities. Instead, ER dysregulation seems to be a distinctive feature among offenders with APD, and especially pronounced in those with increasing symptom severity.

During the actual regulation attempt, the AI, APDs surprisingly indicated no impairments in AR success. While this result contrasts our findings on habitual AR, it is somewhat in line with prior research (Lobbestael et al., 2009; however, for a critical account on this study see above). Future research needs to clarify, why APDs were (apparently) able to regulate angry affects in the lab, but indicated suffering from AR difficulties in everyday life. Interestingly, when looking at a behavioral measure of the CAT, a different pattern of results occurred: APDs as opposed to HCs and INCs showed increased aggressive behavior during the

baseline – in a situation where there was no incentive to punish the other players. This increased spontaneous aggression might reflect a lower threshold for aggressive behavior among APDs. However, APDs' reactive aggression was comparable to HCs' (i.e. similar increase in aggressive behavior from baseline to anger and similar aggressive behavior during the provocation). This suggests that APDs' impairment mainly is to react aggressively in situations where there are no justified external cues. Unfortunately, our study design does not enable us to draw conclusions about participants' reasons for the punishing behavior or potential mediators. According to prominent aggression theories it is possible that APDs' increased spontaneous aggression was due to their increased trait anger (e.g. Finkel, 2014), different beliefs and attitudes (e.g. Anderson & Bushman, 2002) and/or a hostile attribution bias (see Wilkowski, Crowe, & Ferguson, 2015). INCs, by contrast, showed a reduced increase in aggressive behavior due to the AI and behaved least aggressively when provoked. On first glance, such a diminished reactive aggression seems desirable. However, HCs' behavior implies that aggression might be functional under certain conditions – at least if no physical harm is involved, but a rather weak form of aggressive behavior is conducted, as was the case in the current study. Hence, INCs' diminished adjustment to provocations could reflect a reduced assertiveness. As with increased and inflexible use of reappraisal (see above), this could lead in the long run to an accumulation of stressors – thereby, at some point, being the straw that breaks the camel's back – and resulting in an aggressive outburst. Future research should examine whether or not this decreased reactive aggression among INCs indeed reflects dysfunctional behavior, i.e. it is inflexible and increases the risk for later aggression, while APDs suffer from a more general tendency toward aggressive behavior in situations without external cues. Such different mechanisms contributing to aggressive behavior (decreased threshold vs. increased threshold) would have important implications for treatment programs.

In addition to the limitations already mentioned, further restrictions have to be kept in mind when interpreting the present results. First, as lying, deception, and manipulation represent symptoms of APD (American Psychiatric Association, 2013), the question arises whether APDs deliberately manipulated the outcome variables. However, APDs' similar (vs. INCs) or rather decreased (vs. HCs) extent of social desirability and their openness regarding potentially detrimental information (e.g. admitting drug use in prison) do not suggest a general dishonest response style – quite the contrary. Second, it is suggested that aggressive behavior in individuals with narcissistic tendencies is triggered by different provocations than in individuals with psychopathic tendencies (ego threat vs. physical threat; Jones & Paulhus, 2010). We neither assessed psychopathic tendencies nor personality disorders other than APD.

Hence, we cannot rule out that these personality features biased the current results. Future work should cover the corresponding symptoms. Third, several limitations apply to the CAT: About half of the participants saw through the cover story. Although it seems advantageous to have assessed credibility at all (which is, unfortunately, not common practice when applying AI methods) and to have included it as a factor into our analyses, this inevitably led to a reduced sample size in the subgroups of interest (i.e. deceived participants) and thus to a reduced power to detect effects. Moreover, it must be criticized that our AI contained rather weak provocations. Nonetheless, the AI was successful as evident by moderate to large increases in angry emotions across groups. This is remarkable, since credibility (we have no information on credibility beyond the last round of the CAT) and masculine norms (e.g. avoiding the display of vulnerable emotions, see Berke, Reidy, & Zeichner, 2018) might have influenced participants' post-ratings, thus resulting in a potential underestimation of the true effects. Future research might consider using more intense AIs (if ethically justifiable) and assessing additional physiological measures during the AI, as they seem more sensitive to detect changes in mood and/or arousal (Lobbestael et al., 2008). Furthermore, our AI might have been less suitable for APDs than HCs (i.e. less threatening APDs' self-concept/not affecting them on a personal level). As a consequence, it is possible that group effects between APDs and HCs were underestimated, not only in terms of anger reactivity but also with respect to reactive aggressive behavior. Clearly, future research needs to address this issue. Another major limitation is that we assessed aggressive behavior only with respect to a mild form of indirect and active resource aggression (theft). Hence, generalizability to other forms of real-life aggression is questionable. We clearly agree with McCarthy and Elson (2018, p. 10) who state that "claims about 'aggression' as a general construct" are only valid "when there is converging and replicable evidence from several different lab-based aggression paradigms". Hence, the current study is a starting point in a hopefully growing field of aggression research in APDs. Further studies that use different, but also ecologically valid, AI and aggression paradigms are necessary to broaden our knowledge of APDs' aggression (proneness). Only this will enable us to infer empirically well-founded treatment decisions.

Despite the aforementioned limitations, the current study is the first to provide preliminary evidence for abnormalities in aggressive behavior among both, APDs and INCs – though different in nature. Moreover, APDs, unlike INCs, seem to suffer from a wide range of ER deficits, including, but not limited to, AR. Different mechanisms may therefore be responsible for APDs' and INCs' aggressive behavior. In this case, different treatment methods would be appropriate depending on the APD diagnosis. Although clearly more research is

needed to provide empirically grounded recommendations for specific intervention programs, the current findings suggest that interventions for APDs should include anger management trainings, but also go beyond, by additionally covering more general ER abilities and/or strategies. Overall, APD should not be viewed as a mere behavioral disorder, but also as a disorder of ER.

4. Main Study, Part II: Yes, I Can! Antisocial and Non-Antisocial Offenders Show No General Deficit in Cognitive (Inhibitory) Control¹²

Abstract

Although it is assumed that individuals with antisocial personality disorder (APDs) suffer from deficits in executive functioning, exactly affected components have yet to be specified. Cognitive control, and particularly cognitive inhibitory control (IC), represent the elementary basis of executive functions and may well contribute to APDs' symptom domain (e.g. impulsivity, anger experience and physical aggression). The current study examined IC in $n = 31$ inmates with APD as compared to $n = 32$ inmates without APD (inmate control participants; INCs) and $n = 39$ education and aged-matched healthy controls (HCs). To determine whether potential impairments in APDs are specific to IC, we conducted a second measure for other components of cognitive control as well. Within inmates, relationships between IC and antisocial symptoms were examined. Contrary to expectations, no evidence was found for a deficient IC among APDs as compared to INCs or HCs – neither with respect to overall performance level nor post-conflict adjustments. Deficits in more broad cognitive control abilities could not be identified either. APDs indeed reported increased impulsivity, anger experience and physical aggression compared to both, INCs and HCs. However, there was no evidence for associations between poor IC performance and antisocial symptoms within inmates. Although further research with increasing task demands and different modes of IC is required, the present results clearly challenge the assumption of a diminished IC underlying APDs' symptom domain. Implications for future research are provided to enhance our understanding of APD.

General scientific summary: Inmates with and without antisocial personality disorder showed no deficits in cognitive inhibitory control and did not differ from each other. The current results challenge the idea that specifically, impairments in cognitive inhibitory control may underlie the antisocial symptom domain.

Keywords: Antisocial Personality Disorder, Offender, Inhibitory Control, Cognitive Control

¹² An abridged version of this manuscript is intended for publication but has not yet been submitted. A declaration on the share of collaborative work is given in Table 2 in chapter 1.6.

4.1. Background

Cognitive control abilities represent low-level executive functions (Nigg, 2017) and are needed to maintain goal-oriented behavior in situations with immediate conflict, i.e. in situations with directly competing cognitive and behavioral demands (Zeier et al., 2012). Often, an automatic but goal-conflicting response has to be suppressed in favor of a more complex and cognitively demanding reaction that matches internalized current goals and intentions (Miller & Cohen, 2001). Hence, besides focusing and switching attention (i.e. shifting), updating and manipulating working-memory, cognitive control also embraces IC (Nigg, 2017). It is assumed that cognitive control processes are essential for an adaptive ER (e.g. Nigg, 2017; Tang & Schmeichel, 2014). Particularly a poor IC might lead to increased processing of mood-congruent and reduced processing of mood-incongruent memory contents (e.g. Joormann & Vanderlind, 2014), thus increasing the risk for maladaptive ER. Beyond that, and almost by definition, IC is needed to overcome aggressive urges (Zeier et al., 2012) and resist impulsive behavior (Nigg, 2017). Given APDs' increased trait anger (Timmermann et al., 2017), their impaired ER (see chapter 3), their aggression proneness and their increased impulsiveness (e.g. diagnostic criteria, see American Psychiatric Association, 2013), one could assume a common underlying deficit in IC. Poor IC might even distinguish offenders with APD from those without APD. Determining whether (a) APDs indeed suffer from a deficient IC and (b) whether this ability is associated with APD symptoms, could therefore improve the understanding of the disorders' underlying processes. Corresponding results might serve as a starting point for tailoring prison interventions for APDs and INCs.

Despite the relevance of this issue, there are, to our knowledge, only three¹³ recent studies that assessed IC in male offenders with APD (symptoms), revealing inconsistent results: when using the Eriksen-Flanker task (Eriksen & Eriksen, 1974), a task which is assumed to measure IC, offenders exhibiting increased APD symptoms did not differ from those with low symptom severity regarding IC efficiency (i.e. regarding RTs), while results were inconclusive with respect to IC effectiveness (accuracy) (Zeier et al., 2012). However, it is not possible to determine whether inmates' similar results reflected an abnormal or a normal IC performance, since there was no second comparison group composed of non-incarcerated HCs (Zeier et al., 2012). Roszyk et al. (2013) at least recruited a non-incarcerated control group. They used a paper-version of the Stroop Colour Word Task (Stroop, 1935) and operationalized IC by

¹³ Although two other recent studies claim to have assessed IC (Baliouis et al., 2019; Chamberlain, Derbyshire, Leppink, & Grant, 2016), closer examination reveals that they rather captured the ability to suppress an already primed motor reaction. Therefore, these studies are not outlined in detail.

slowing from the automated reading condition to the color naming task (instead of the more common operationalization by the interference due to incongruent stimuli). In this study, APDs showed significantly enhanced delays compared to non-incarcerated control participants, which was interpreted as a deficit in IC (Roszyk et al., 2013). Unfortunately, sample characteristics limit the interpretation of results: As the primary interest of Roszyk et al. (2013) was the investigation of sexual crimes, the majority of APDs were sex offenders. Second, no information on (comorbid) disorders was provided, neither for APDs nor for controls, so it cannot be ruled out that comorbidities (and peculiarities of sex offenders) were responsible for APDs' impairments. Another study that excluded participants with psychiatric disorders other than SUD/AUD found contrasting results to Roszyk et al. (2013): Schiffer et al. (2014) compared offenders with APD with a carefully matched (e.g. by IQ and former SUD) non-incarcerated control group: using a computerized Stroop task, no increased interference and thus no deficient IC was found in APDs. Surprisingly, APDs even showed a *better* IC efficiency. However, the pattern of results was different for post-conflict adjustments. The so-called conflict adaptation effect refers to the observation that participants usually show less interference in response to a preceding incongruent trial (i.e. the conflict) as compared to a preceding congruent trial (see Botvinick, Braver, Barch, Carter, & Cohen, 2001). This phenomenon is assumed to mirror a continuous adjustment of cognitive (inhibitory) control, which is more highly engaged after conflict detection (Carter & van Veen, 2007). Although unexpectedly no clear behavioral adjustment effect was found for either group (as evident by a non-significant omnibus test), APDs exhibited similar (non-)adjustments compared to controls (Schiffer et al., 2014). However, some limitations of the study have to be considered: Despite the study's thorough and undoubtedly very elaborate matching it should be questioned, how suitable a SUD-matched control group is. To obtain an understanding of APD in its entirety it could be argued that although the influence of characteristic features such as SUD/AUD (e.g. Black et al., 2010) should be assessed, no attempt should be made to (artificially) control it. Or at the least, a HC group should be included. In addition, it has to be noted that a substantial part of Schiffer et al.'s (2014) sample was recruited from forensic psychiatric services for offenders with SUDs, which makes transferability to a regular prison sample more difficult. In sum, research on IC in APDs is not only scarce, but additionally inconclusive (e.g. Roszyk et al., 2013 vs. Schiffer et al. 2014).

If, in the absence of further studies on APDs' IC performance, the research scope is expanded to more broad samples and superordinate cognitive abilities, studies examining IC in offender populations (Pasion, Cruz, & Barbosa, 2018; Seruca & Silva, 2015, 2016) and (meta)

analyses on executive functions in ASBs can be found (Ogilvie et al., 2011). While more recent studies revealed no IC impairments among different offender populations as assessed with Stroop tasks, these findings may have been biased by using a paper-pencil Stroop (Pasion et al., 2018) and lacking a thorough psychiatric assessment (Seruca & Silva, 2015, 2016). By not assessing APD, impairments in this offender subgroup may have been masked. Thus, on closer inspection, these studies provide only limited informative value regarding the current research question. With respect to superordinate executive functions, Ogilvie et al.'s (2011) meta-analysis is often considered as evidence for executive impairments in APDs. However, the minor effect size of APDs' "deficits" ($d = .19$) questions clinical relevance and thus urges caution in the (over)interpretation of this finding. With regard to the sub-component IC, as measured by Stroop tasks, results were only provided for ASBs, and revealed small deficits in IC ($d = .35$). At first, this result might suggest a poor IC also among APDs. However, due to the heterogeneous samples subsumed under antisocial behavior (e.g. children, adolescents, social drinkers, psychiatric patients, incarcerated psychopaths, APDs etc.), the various control groups used (e.g. clinical vs. inmate vs. HC groups) and the different Stroop operationalizations conducted, it cannot be inferred whether offenders with APD indeed exhibit deficits in IC and if so, whether these impairments are actually clinically relevant, whether they occur in different dimensions (e.g. overall performance, conflict adaptation), whether they are unique to inmates diagnosed with APD or whether they also occur in non-antisocial offender populations instead.

Taken together, although there are numerous studies investigating executive functioning in ASBs, there is a lack of research focusing on thoroughly diagnosed inmates with APD and including both, an incarcerated control group (i.e. INCs) and a HC group. Moreover, looking only at executive functions (in ASBs) carries the risk of masking specific deficits (in specific subgroups; see also Zeier et al., 2012) or, on the contrary, ascribing deficits, which do not exist. Despite the high relevance, there are only few studies on APDs' IC, which unfortunately yielded ambiguous results. Therefore, the main goal of the present work was to clarify whether male offenders with APD differ from those without APD (i.e. INCs) and non-incarcerated HCs regarding their IC and if so, whether this deficit is specific or also occurs in other cognitive control abilities.

While, to our knowledge, no study thus far investigated relationships between a poor IC and specific APD symptoms in offenders with and without APD, some preliminary evidence for this assumption comes from studies in non-clinical samples: In undergraduate students poor IC was associated with higher anger experience (Zajenkowski & Zajenkowska, 2015) and increased aggression (Holley, Ewing, Stiver, & Bloch, 2015). Hence, the second goal of this

study was to examine whether impairments in IC are indeed accompanied by increased self-reported impulsivity, anger experience and aggression within inmates. Additionally, we aimed to assess the relationship between IC and overall antisocial symptom severity when looking on APD dimensionally.

Based on theoretical considerations we predicted a deficient IC in APDs as compared to HCs. However, due to inconclusive previous findings, we had no specific hypothesis regarding differences between INCs and APDs. The exact dimensions (mean performance, control engagement after conflict) in which potential deficits of APDs become apparent should be explored. To evaluate whether deficits are specific to IC we report a short screening measure for cognitive control abilities beyond IC. Based on the considerations depicted above we expected associations between a deficient IC and increased impulsive behavior, anger experience and aggression within inmates, while relations between poor IC and overall antisocial symptom severity should be explored.

4.2. Methods

4.2.1. *Participants*

One hundred and three participants were enrolled in this study¹⁴. Due to data quantity, results were split between two manuscripts. Both inmate groups (APDs, INCs) were recruited in three German prisons, while HCs were recruited in the community. Inmates were either in pre-trial detention or in criminal custody in closed prison. Inclusion criteria for all participants were: male gender, $18 \leq \text{age} \leq 69$, no psychotropic medication, unless stable dosage for at least 4 weeks, sufficient knowledge of the German language, verbal IQ > 80, and no color blindness. APDs ($n = 31$), but not INCs ($n = 33$), had to fulfill diagnostic criteria for current APD (i.e. significant symptomatology within the last years following the SCID-II). Exclusion criteria for inmates were: current episode of major depression, current bipolar disorder, current social anxiety disorder, current posttraumatic stress disorder, current SUD or AUD if moderate or severe and no abstinence in the past 6 months, current or lifetime psychotic disorder, and current anorexia nervosa. Furthermore, APDs and INCs had to report at least one lifetime criminal offence (beyond traffic offences and offences against foreigners' law) to ensure criminality. Further inclusion criteria for HCs ($n = 39$) were no lifetime imprisonment and no current or past psychiatric disorder, as assessed with the M.I.N.I. HCs were matched to APDs by age (+/- 5 years) and education (university entrance diploma: yes vs. no). Sample characteristics are

¹⁴ It was the same sample as in Main Study, Part I (see chapter 3.2.1)

depicted in chapter 3.2.1: Table 5 shows demographic information and symptom severities for all groups, while Table 6 and Table 7 show detention information and diagnostic information for inmates. Figure 8 portrays self-reported criminality among inmates.

4.2.2. Measures

Diagnostic assessment. Diagnoses were given according to DSM-5 using the M.I.N.I. While sections B (suicidal tendencies), P (APD) and Q (borderline personality disorder) were skipped, additional questions were asked to receive information on *lifetime* SUD and AUD. Due to its more detailed assessment, the SCID-II was carried out to assess conduct disorder and APD. Since diagnostic criteria has not changed from DSM-IV to DSM-5 (see American Psychiatric Association, 2013), conducting the SCID-II provided up-to-date diagnostics. Skip rules were not followed in order to gain a dimensional measure of APD for all participants. APD symptom severity was calculated by adding up recoded scores for each of the seven APD criteria (1 = absent was recoded to 0, 2 = subthreshold was recoded to 1, 3 = true was recoded to 2), resulting in a score ranging from 0 to 14. For ADHD symptom severity, the ADHD-SR was conducted. This questionnaire measures the 18 DSM-5 symptoms of ADHD, distributed on the scales of inattention and hyperactivity/impulsivity. For the present purpose, the total score was used. A cutoff-score ≥ 15 is proposed (sensitivity = .77, specificity = .75; Rösler et al., 2008). To assess verbal intelligence and ensure language skills, the MWT-B was conducted. IQ estimates are also reported.

Cognitive control. To check whether potential impairments in IC are specific, we investigated two different aspects of cognitive control.

Cognitive inhibitory control. A computerized version of the Stroop Color Word Task was applied to assess IC. The Stroop task is frequently used and warrants a relatively good comparability to past research. Four German color words were presented in one of four font colors (red, green, blue, yellow). Participants had to indicate the font color of the presented word as fast as possible. We included two conditions: In the congruent condition the font color matches the presented word (red, green, blue, yellow), while in the incongruent condition the font color and the presented word differ (e.g. red, green, blue, yellow). The task consisted of two experimental blocks with 96 trials each. Half of the trials were congruent. All color words or font colors occurred with the same frequency. Stimuli were presented randomized without repeats. Each trial began with a fixation cross appearing against a black screen for 1 second. Thereafter, the Stroop stimulus followed. It was presented until the participant indicated an answer by pressing a corresponding computer key, but for a maximum of 4 seconds. Then, the

next trial began. We assessed RTs for correct responses and accuracy, with non-responses coded as errors. Strong interference (i.e. the deceleration in RT or the deterioration of accuracy from congruent to incongruent trials) is thought to reflect poor IC, since in incongruent trials the dominant and relatively automated reading ability must be inhibited in favor of the less automated color naming.

Working memory and switching ability. The Trail Making Test (TMT; Reitan, 1992) is a neuropsychological test sensitive to detect brain damage (Tombaugh, 2004). Participants have to connect numbers printed on a sheet of paper in ascending order as quickly as possible and without removing the pen (part A). In part B numbers and letters must be connected alternately. The dependent variable is the time in seconds needed to complete. Although TMT-A and TMT-B performances are highly correlated (Tombaugh, 2004), the TMT-A primarily reflects visual search and perceptual speed, while the TMT-B additionally requires cognitive control abilities, particularly working memory (Sánchez-Cubillo et al., 2009). The difference score TMT-B - TMT-A (B-A) hardly reflects visuo-perceptual abilities anymore, but is mainly related to switching ability (Sánchez-Cubillo et al., 2009).

Self-reported trait impulsivity. The short form of the Barratt Impulsiveness Scale (BIS-15; Spinella, 2007) was conducted to assess impulsivity. Besides the total score, the BIS-15 provides measures for lack of future orientation and foresight (scale nonplanning), cognitive instability/lack of persistence (scale attention impulsivity) and acting on the spur of the moment (scale motor impulsivity). Motor impulsivity corresponds most closely to more narrow definitions of impulsivity (Nigg, 2017) and was therefore selected as a measure of trait impulsivity in the current work. However, all scales' results are reported for reasons of transparency. Items are rated on a 4-point Likert scale, with higher values indicating increased impulsivity.

Anger and aggression. The Aggression Questionnaire (AQ; Buss & Perry, 1992) measures different aggression factors and related constructs. It consists of two scales assessing behavior (physical aggression, verbal aggression), and two scales measuring emotion (anger) or cognition (hostility), respectively (Buss & Perry, 1992). For the current study only emotion (anger experience) and behavior (aggression) were of interest. Due to the validation of the German adaption (Herzberg, 2003), an interpretation of the scale verbal aggression was omitted. Hence, only the scales anger and physical aggression were interpreted. Again, for reasons of transparency, the results of the other scales are nevertheless reported. However, no total score is provided as it would be inconsistent with definitions of aggression (Parrott &

Giancola, 2007). Items of the German version are rated on a 5-point Likert scale. For each scale, mean values are reported.

4.2.3. Procedure¹⁵

The study protocol was approved by the local ethical committee and the corresponding Criminological Services. Written informed consent was obtained. All participants received financial compensation. At the beginning of the session, the semi-structured interviews (i.e. M.I.N.I. and SKID-II) were carried out to assess psychiatric disorders and to check for inclusion criteria. Hereafter, the TMT-A and the TMT-B were applied. Then, the MWT-B was completed and subsequently, participants conducted the computerized Stroop task. First, written instructions and examples were presented. Afterwards, participants performed a short exercise, consisting of four trials. They received visual feedback and support from the investigator to ensure task comprehension. Then, a practice block, consisting of eight trials, was conducted. Visual feedback was still provided but there was no assistance by the investigator. Participants had to reach an accuracy level of $\geq 75\%$ to continue with the experimental block. Otherwise, the practice block was repeated (with different trials) until participants' performance met the requirements. After successful completion, two experimental blocks à 96 trials followed. No feedback was provided. Between blocks, a resting period of variable length, regulated by the participant, was conducted. After the Stroop, another experiment¹⁵ was accomplished. Then, participants were asked to complete the BIS-15 and the AQ. Furthermore, some additional questionnaires¹⁵ and demographic information were assessed. At the end of the session, the ADHD-SR was completed.

4.2.4. Data Analysis

Regarding hypotheses testing, we used nonparametric analyses (Kruskal-Wallis tests, Mann-Whitney *U* tests) in case of violations of normality and if offered by SPSS. Otherwise, we initially employed transformations to approximate normality more closely. However, as analyses with transformed data did in no case differ from analyses with original data, we report raw data for better interpretation. Several (mixed) ANOVAs as well as (dependent) *t*-tests were carried out. For correlation coefficients, Spearman's rho is reported. 95% CIs for correlation coefficients were calculated using the freeware Psychometrica (Lenhard & Lenhard, 2014). For

¹⁵ The present study was part of a larger study design (see chapter 3). Additional measures included a newly developed AI paradigm and several ER questionnaires. Due to the large amount of data, the corresponding results are reported in a separate manuscript (see chapter 3.3).

non-repeated measures means and medians 95% BCa bootstrap CIs were taken from SPSS (using 1000 samples). Absolute values of Cohen's d (for a classification of effect sizes see Cohen, 1988) are offered as effect sizes following significant omnibus tests and were again calculated using Psychometrica (Lenhard & Lenhard, 2016). A significance level of $\alpha = .05$ was used, where $.05 < \alpha < .10$ was considered as marginally significant.

4.3. Results

4.3.1. Cognitive Inhibitory Control

Regarding the Stroop task, $n = 1$ participant had to be excluded from analyses due to arthrosis of the left hand. Accuracy was generally very high, $M = .983$, 95% CI [.978, .987], $SD = .020$, indicating a ceiling effect. Therefore, no further analyses were conducted with respect to accuracy. Regarding RTs, only trials with correct responses were analyzed. Visual inspection of RT distribution did not indicate the need for an absolute lower cutoff, no extremely fast RTs below 200 ms were detected. Data was not trimmed¹⁶.

First, to analyze overall performance level, we conducted a mixed design ANOVA with the within-subjects factor condition (congruent, incongruent) and the between-subjects factor group (APDs, INCs, HCs) on RTs. To determine whether groups differed in conflict adaptation effects, we subsequently conducted a 2 (previous trial type: congruent, c; incongruent, i) \times 2 (current trial type: congruent, C; incongruent, I) \times 3 (group: APDs, INCs, HCs) mixed design ANOVA on RTs. Indications for post-conflict adjustments would come from an interaction between previous and current trial time (Botvinick et al., 2001): Stronger interference effects (I - C) should occur after preceding congruent (c) compared to preceding incongruent (i) trials: $(cI - cC) > (iI - iC)$ (Botvinick et al., 2001). Hence, a three-way interaction would reveal different conflict adaptations between groups.

Mean performance level. Descriptives are depicted in Table 13. ANOVA yielded a significant main effect of condition, indicating an interference effect for all groups $F(1, 99) = 159.47$, $p < .001$, $d = 1.685$. However, and contrary to hypotheses, no group effects occurred, neither with respect to the main effect of group: $F(2, 99) = 0.28$, $p = .754$, nor the expected Group \times Condition interaction: $F(2, 99) = 0.21$, $p = .809$.

¹⁶ Results did not change when excluding trials $\pm 2.5 SD$ above/below each participant's mean/standard deviation, regardless of using transformed or untransformed data.

Table 13. Stroop variables by group

Stroop variable	APDs ($n = 31$)		INCs ($n = 32$)		HCs ($n = 39$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Mean RT						
Congruent	748.0	126.2	768.7	110.8	775.2	170.7
Incongruent	821.9	156.1	835.6	138.1	850.7	207.0
Interference	73.9	57.3	66.9	58.9	75.5	56.1

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. RT = reaction time. Time in milliseconds.

Conflict adaptation. In accordance with mean performance analysis, the previous trial type (i, c) \times current trial type (I, C) \times group (APDs, INCs, HCs) mixed ANOVA revealed a significant main effect of current trial type, $F(1, 99) = 173.67, p < .001$. This effect was qualified by a marginally significant interaction of current trial type and previous trial type, $F(1, 99) = 3.50, p = .064$: While participants showed a trend towards a rather unusual pattern of conflict adaptation, $(iI - iC) > (cI - cC)$, $t(101) = 1.77, p = .079, d = .171$, effect size was negligible (for descriptives see Table 14). More importantly, groups did not significantly differ with regard to their post-conflict adjustments, as evident by non-significant effects of group (Group: $F(2, 99) = 0.26, p = .769$, Group \times Previous trial type: $F(2, 99) = 0.67, p = .516$, Group \times Current trial type: $F(2, 99) = 0.52, p = .595$, Group \times Previous trial type \times Current trial type: $F(2, 99) = 0.79, p = .459$).

Table 14. Reaction times for combinations of current and preceding trial type

Trial type combination	<i>M</i>	<i>SD</i>
cI	831.4	170.2
iI	835.0	173.5
cC	764.9	140.7
iC	756.4	140.3
cI - cC	66.5	66.5
iI - iC	78.6	62.9

Note. cI = incongruent trial preceded by a congruent trial. iI = incongruent trial preceded by an incongruent trial. cC = congruent trial preceded by a congruent trial. iC = congruent trial preceded by an incongruent trial.

Taken together, APDs did not differ from INCs and HCs, neither in terms of overall IC performance nor post-conflict adjustments.

4.3.2. Working Memory and Set-Shifting

Regarding the TMT, no significant group differences occurred, neither with respect to the TMT-A, $H(2) = 0.83, p = .660$, the TMT-B, $H(2) = 0.64, p = .725$, nor B-A, $H(2) = 1.81, p = .404$. Hence, all groups showed similar visuoperceptual, working memory and shifting performances. Descriptives are depicted in Table 15.

Table 15. Trail Making Test times by group

Measure	APDs ($n = 31$)		INCs ($n = 32$)		HCs ($n = 39$)	
	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>Mdn</i> (<i>IQR</i>)	95% CI
TMT-A	27.00 (14.00)	[24.50, 28.50]	27.00 (16.50)	[25.00, 29.00]	28.00 (10.00)	[27.00, 28.00]
TMT-B	63.00 (33.00)	[59.00, 81.00]	65.00 (28.00)	[58.00, 73.00]	63.00 (31.00)	[62.00, 63.00]
B-A	38.00 (28.00)	[34.00, 48.00]	34.00 (23.50)	[28.00, 45.00]	36.00 (20.00)	[29.50, 41.00]

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. TMT = Trail Making Test. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval. Time in seconds.

4.3.3. Trait Impulsivity

Regarding self-reported motor impulsivity as measured by the BIS-15, ANOVA yielded a significant effect of group, $F(2, 100) = 6.66, p = .002$. APDs reported increased impulsivity compared to both, INCs, $t(62) = 3.23, p = .002, d = 0.808$, and HCs, $t(68) = 3.19, p = .002, d = 0.768$, while INCs and HCs did not differ from each other, $t(70) = 0.27, p = .789$. Descriptives for this scale, but also for the other scales not interpreted here, can be seen in Table 16.

Table 16. Groups' impulsivity as measured by the short form of the Barratt Impulsiveness Scale

	APDs (<i>n</i> = 31)			INCs (<i>n</i> = 33)			HCs (<i>n</i> = 39)		
	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI
Motor impulsivity	12.74^b (2.62)		[11.79, 13.68]	10.55^a (2.81)		[9.59, 11.48]	10.72^a (2.65)		[10.00, 11.48]
Further scales									
Nonplanning	11.23 (3.38)		[10.10, 12.44]	11.03 (3.53)		[9.91, 12.11]	10.18 (2.96)		[9.32, 11.17]
Attention impulsivity		10.00 ^{a2} (3.00)	[9.00, 12.00]		9.00 ^{a1} (3.00)	[9.00, 10.00]		9.00 ^{a1} (4.00)	[8.00, 10.00]
Total score		34.00^b (9.00)	[31.00, 37.00]		29.00^a (7.00)	[27.00, 31.50]		29.00^a (8.00)	[28.00, 31.50]

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval. Different superscripts indicate significant differences (reported in bold face) at $p < .05$ in pairwise comparisons after significant omnibus test at $p < .05$. The letters a denote smaller sum of ranks/means.

^{a1} and ^{a2} indicate a marginally significant group difference at $p < .10$.

4.3.4. Anger Experience and Physical Aggression

Concerning the AQ, groups significantly differed with respect to anger, $H(2) = 23.58$, $p < .001$, and physical aggression, $H(2) = 42.42$, $p < .001$. Follow-up tests revealed that APDs reported higher anger experience than both, HCs, $U = 241.50$, $z = 4.31$, $p < .001$, $d = 1.195$, and INC, $U = 203.50$, $z = 4.16$, $p < .001$, $d = 1.209$, while HCs and INCs reported comparable anger levels, $U = 603.50$, $z = 0.46$, $p = .648$. With respect to physical aggression APDs reported increased aggression as evident by group differences to INCs, $U = 129.00$, $z = 5.15$, $p < .001$, $d = 1.676$ and HCs, $U = 107.50$, $z = 5.89$, $p < .001$, $d = 1.973$. INCs, though reporting less aggression than APDs, indicated marginally significant higher aggression than HCs, $U = 474.00$, $z = 1.92$, $p = .054$, $d = 0.463$. Means (or medians) and 95% BCa CIs, including those of the scales not interpreted here, are listed in Table 17.

Table 17. Groups' anger and aggression as measured by the Aggression Questionnaire

	APDs (<i>n</i> = 31)			INCs (<i>n</i> = 33)			HCs (<i>n</i> = 39)		
	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI	<i>M</i> (<i>SD</i>)	<i>Mdn</i> (<i>IQR</i>)	95% CI
Physical Aggression		3.08^b (0.75)	[2.81, 3.34]		1.75^{a2} (1.00)	[1.75, 1.75]		1.50^{a1} (0.63)	[1.38, 1.50]
Anger		2.96^b (0.83)	[2.70, 3.22]		2.17^a (0.92)	[1.83, 2.33]		2.00^a (0.83)	[1.92, 2.17]
Further scales									
Verbal Aggression		3.45^b (0.64)	[3.21, 3.68]		2.88^a (0.54)	[2.69, 3.08]		2.72^a (0.57)	[2.53, 2.91]
Hostility		3.00^b (0.65)	[2.79, 3.22]		2.58^a (0.65)	[2.35, 2.83]		2.39^a (0.83)	[2.15, 2.63]

Note. APDs = Inmates with antisocial personality disorder. INCs = Inmate control participants. HCs = Healthy controls. 95% CI = 95% bias corrected and accelerated bootstrap confidence interval. Different superscripts indicate significant differences (reported in bold face) at $p < .05$ in pairwise comparisons after significant omnibus test at $p < .05$. The letters a denote smaller sum of ranks/means.

^{a1} and ^{a2} indicate a marginally significant group difference at $p < .10$.

4.3.5. Associations between Cognitive Inhibitory Control and Antisocial Symptoms

Although no evidence for poor IC has been found among APDs on the between-groups level, we examined whether IC performance (i.e. Stroop RT interference) was related to inmates' self-reported APD symptoms impulsivity, anger experience and aggression ($n = 63$). We further investigated whether IC performance was associated with overall APD symptom severity when looking at APD dimensionally (i.e. SCID-II score). Within inmates, IC performance was neither significantly related to specific APD symptoms (impulsivity, anger, aggression) nor overall APD symptomatology. Results can be seen in Table 18.

Table 18. Spearman's rank correlation coefficients between cognitive inhibitory control and antisocial symptoms in inmates

	Motor impulsivity (BIS-15)	Anger (AQ)	Physical aggression (AQ)	Overall APD symptom severity (SCID-II)
Stroop RT interference	<.01 [-.25, .25]	.22 [-.03, .44]*	.01 [-.24, .25]	-.01 [-.25 .25]

Note. 95% confidence intervals are reported in brackets. After applying a Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) to decrease the false discovery rate¹⁷, no *p*-value was smaller than the critical value, i.e. no significant correlations occurred. BIS-15 = short form of the Barratt Impulsiveness Scale. AQ = Aggression Questionnaire. APD = antisocial personality disorder. SCID-II = Structured Clinical Interview II for DSM-IV. RT = reaction time.

**p* = .046, non-significant after the Benjamini-Hochberg correction.

4.4. Discussion

This study aimed to clarify whether APDs exhibit impairments in IC and if so, whether such deficits are associated with more pronounced symptoms of impulsivity, anger experience and aggression, as well as increased overall APD symptom severity. To assess whether potential deficits are unique among offenders with the psychiatric disorder APD, we recruited not only a HC group but also a group of INCs. To determine whether potential IC deficits are specific to this cognitive control component, we included another measure of cognitive control beyond IC. Contrary to expectations, no deficient IC was revealed within APDs: Neither did they perform worse than INCs or HCs in terms of IC efficiency, nor did they show adverse post-conflict adjustments. Of note, even with regard to more broad cognitive control abilities (working memory and shifting abilities), no group differences occurred. In contrast to the present null findings concerning cognitive performances, APDs indeed reported increased levels of behavioral impulsivity, anger experience and physical aggression – both in comparison to HCs and INCs, while the latter only marginally differed from each other in terms of aggression. Hence, nosologic APD symptoms were empirically confirmed and were able to distinguish inmates with and without APD, while IC and cognitive control performance did not. Furthermore, no consistent associations between a poor IC and the aforementioned symptoms and overall APD symptom severity (within inmates) were revealed.

Upon first glance, our results contradict Schiffer et al.'s (2014) and Roszyk et al.'s (2013) research on IC, as well as Ogilvie et al.'s (2011) meta-analysis on executive functioning. There are more or less three explanatory approaches for these unexpected results: (1) Either

¹⁷ For a critical account on sequential Bonferroni procedures see Nakagawa (2004).

sampling characteristics are responsible for our null findings, i.e. our sample was not representative for the superordinate population of offenders with APD and/or we used inappropriate comparison groups. (2) Or, we may have missed relevant impairments due to methodological reasons. (3) Or else, the assumptions that APDs exhibit poor overall IC and that such a deficient IC underlies the symptom domain of APD, is simply wrong. We will deal with these different interpretations one after the other.

Regarding objection (1), it must actually be questioned whether our sample was representative of the entire German prison population. However, all studies in this field of research have to deal with these problems of external validity: In the sensitive prison environment, inmates often have reservations about participating in a scientific study (e.g. concerns about collaborations with public prosecutors and the like). For legal and ethical reasons, it is neither possible nor desirable to force inmates to participate (which, by the way, could also distort the results). As a matter of fact, certain offender populations unfortunately remain hidden from research. Nonetheless, we are confident, that our sample was quite representative (see also inmates' reports of various committed offences in Figure 8), since there was little preselection by the prison staff, and we were often able to approach inmates during prison routine. This allowed us to persuade inmates to participate in this study, who originally had concerns. Regarding the suitability of the control groups, we consider the recruitment of two comparison groups and the matching of HCs as a strength of this work. However, despite matching HCs for age and education, APDs showed diminished verbal intelligence (i.e. MWT-B) compared to INCs and HCs, who did not differ from each other. Correspondingly, we cannot rule out that low verbal skills led to a less automated reading condition in APDs as compared to the other groups. This, would have produced less conflict and in turn might have resulted in reduced Stroop interference, thus overestimating APDs' true IC abilities and masking potential group differences. Yet, since we assured sufficient German language skills prior to study enrollment, this explanation is not very likely. Furthermore, no group differences have been found in the TMT either, a measure of executive functions which is largely free of language skills/reading levels. This finding also indicates that there are no (severe) cognitive control deficits (working memory, set shifting, IC) among APDs as opposed to INCs or HCs. Furthermore, it is assumed that SUD, AUD and ADHD are associated with IC deficits (for a recent meta-analysis on post-error slowing in SUDs see Sullivan, Perlman, & Moeller, 2019; for a meta-analysis regarding ADHD see Lansbergen, Kenemans, & van Engeland, 2007). Hence, existing group differences regarding SUD/AUD (APDs > INCs > HCs) and ADHD symptomatology (APDs > HCs) should even have overemphasized IC deficits in the APD

sample – and yet no impairments were revealed. However, one limitation that should be considered when interpreting the present findings is that we did not assess personality disorders other than APD. Although prior findings indicate similar IC performances among psychopathic and non-psychopathic offenders in standard versions of the Stroop task (Dvorak-Bertsch, Sadeh, Glass, Thornton, & Newman, 2007; Hiatt, Schmitt, & Newman, 2004), future work might consider assessing psychopathy as well. Moreover, we focused on male inmates only, thus we are not able to make any conclusion about female APDs or more “successful” APDs in society who do not show up in the “bright field” of legal authorities. Nonetheless, we are confident that the current sampling characteristics (i.e. thorough psychiatric assessment, two control groups) speak more in favor of this study. In fact, this might be one of the reasons for contradicting results compared to Roszyk et al. (2013) and Schiffer et al. (2014) but also Ogilvie et al.’s (2011) meta-analysis – studies, whose sampling in part deviated substantially from that presented here and should be questioned.

As far as objection (2) is concerned, the present study, like previous studies, must deal with methodological criticism: For example, it might be argued that our task was rather undemanding to solve. Evidence for this comes from the very low number of errors across groups. As a consequence, existing deficits in APDs might not have come to bear due to ceiling effects. However, it is important to note that we did find strong interference effects across groups regarding RTs, indicating the expected Stroop pattern. Nevertheless, we cannot rule out that deficits within APDs would have been revealed in a more challenging task. To address this issue, future research might, for example reduce the interstimulus interval to enhance task difficulty. Another, presumably more promising, approach would be to vary the ratio of congruent to incongruent trials, since it is known, that higher proportion of incongruent trials leads to reduced interferences (e.g. Bugg, Jacoby, & Chanani, 2011; Bugg, Jacoby, & Toth, 2008; Dvorak-Bertsch et al., 2007). Our balanced congruency ratio might also explain why no clear behavioral adjustment effect was found across participants. Another methodological problem which makes it difficult to integrate the current results into the existing state of research is the fact of methodological variety in Stroop tasks. Regarding the Stroop Color Word Tasks, there are, for example, paper-pencil and computerized versions, variations regarding the response mode, the number of colors, the already mentioned length of interstimulus interval, the proportion of congruency and list-wise or item-wise condition presentation. Such task variations – even when subtle in nature – might lead to changes in the cognitive processes engaged to successfully complete the task. Therefore, slight methodological differences might in part be responsible for divergent results (see also Braver, 2012). Hence, contradicting results

between the current study and Zeier et al. (2012) with Schiffer et al. (2014) and Roszyk et al. (2013) might not be a contradiction in the narrower sense but only reflect the fact that different groups of people completed different tasks by using different underlying skills. Hence, more research with the same methods is necessary to create a broader database. However, we are also in need of research with different measures of IC or other cognitive control abilities to overcome the task impurity problem (Miyake & Friedman, 2012).

In the current study we considered not only mean performance level but also conflict adaptation effects in an attempt to gain a more comprehensive picture of the potential IC impairments in APDs. Despite that, it may be the case that even more specific cognitive (inhibitory) control deficits in APDs have been overlooked: So, for example, our methodology did not allow us to draw conclusion regarding reactive and proactive control, as distinguished in Braver, Gray, and Burgess' (2007) Dual Mechanisms of Control account. While proactive control refers to a sustained preparatory mode of control prior to anticipated conflict, reactive control embraces a rather late, stimulus-driven, correction of interference (Braver, 2012). In order to optimize the behavioral response, a mixture of both modes of control is likely to be beneficial (Braver, 2012). Hence, APDs' deficits in IC may be more subtle and only detected in tasks that require an interplay of several control skills, for example a mixture of proactive and reactive control. Prior research on adolescents and young adults suggests that offenders rely less on proactive control than non-incarcerated controls (Iselin & DeCoster, 2009). However, this pattern of results was dependent on participants' age (adolescents vs. young adults). Correspondingly, APDs might have a deficit particularly in proactive control that is subject to developmental change. Future research is needed to examine whether adult APDs exhibit specific deficits in one mode of control and to determine the developmental course of such deficits.

Despite the aforementioned objections and open research questions, our study clearly indicates that (3) APDs are not as affected by deficient cognitive control as prior research in the field of antisocial behavior might suggest (e.g. Ogilvie et al., 2011). They were not more prone to interference than INCs or HCs, contradicting a deficit in IC. Besides, they showed a comparable pattern of conflict adaptation, i.e. their continuous adjustment of IC was unremarkable. In addition to that, our findings challenge the assumption of a poor IC being subject to APDs' symptom domain. While – in light of these results – it seems plausible to assume that APDs profit more from interventions targeting ER or masculine norms instead of cognitive control trainings, it would nevertheless be premature to reject this form of training in principle. We are clearly in need of further improving our understanding of APD in order to

finally draw empirically founded conclusions regarding promising treatment programs. Given the scarcity of resources in the prison context, it is particularly important to make full use of the limited possibilities. Ultimately, not only would those who are directly affected benefit from well-founded interventions, but it would also contribute to victim protection and is therefore highly relevant for society as a whole.

In sum, this is the first study to investigate APDs' IC that recruited not only a HC group or a INC group, but both. The current findings regarding IC were based on one task only and should therefore be treated carefully. Yet, our results rule out the simple explanation that APDs suffer from an *overall* deficit in IC: If APDs had troubles in IC, they were quite able to compensate for this in a slightly demanding Stroop task. However, more research is needed to determine whether APDs suffer from more nuanced deficits in the sense of the Dual Mechanisms of Control account.

5. General Discussion

The main goal of the present thesis was to examine whether offenders with APD suffer from impairments in self-regulation as compared to HCs and whether these deficits are specific for this offender population or also apply to INCs. For this purpose, different aspects of ER (habitual as well as spontaneous ER, strategy use and overall emotion dysregulation), aggression (behavioral measure of resource aggression and self-reports on physical aggression) and cognitive control performances (IC efficiency, conflict adjustments, broader cognitive control abilities) were assessed. Both APDs and INCs showed abnormalities with respect to ER and aggression in comparison to HCs, but interestingly to a different degree or in opposite ways. Contrary to expectations, neither APDs nor INCs exhibited general impairments in cognitive (inhibitory) control. In the following, the main results of the current work will be presented and discussed in more detail. Since the preliminary studies were only a means to an end for the main study, their results are not directly addressed in this chapter. Limitations of the present study are demonstrated and implications for future research and practice are derived. The chapter closes with a brief conclusion.

5.1. Integration of Results

The most important findings are discussed consecutively, starting with habitual ER, followed by spontaneous AR, aggressive behavior and finally cognitive (inhibitory) control.

5.1.1. Antisocial Personality Disorder – a Disorder of Habitual Emotion Regulation

APDs' impairments in habitual ER were particularly apparent with respect to anger: they indicated increased trait anger alongside a rather maladaptive AR when compared to HCs but also INCs. Furthermore, APDs, but not INCs, suffered from emotion dysregulation in the context of distressing emotions not limited to anger. Of note, ER impairments predicted antisocial symptom severity in the overall inmate sample. This is remarkable, given that comorbid SUD/AUD, age and verbal IQ have been controlled for. Overall, the current results clearly suggest that APD is not only characterized by purely behavioral abnormalities as might be indicated by diagnostic criteria (American Psychiatric Association, 2013), but also by deficits in ER, which distinguishes them from INCs.

APDs' reports of increased dispositional anger experience and anger impulse are consistent with prior research (Timmermann et al., 2017; Yavuz et al., 2016). Since only APDs, but not INCs, indicated increased trait anger, this finding did not simply reflect APDs'

situational circumstances due to the incarceration. Instead, it appears as if only those offenders with APD suffer from a chronic anger pattern, which goes back to childhood (Hawes et al., 2016), and continues into adulthood. The validity of this finding is further underlined by the fact that it was shown in two instruments (STAXI-2 and AQ), with overall strong effect sizes. Unsurprisingly, APDs' chronic anger pattern was not limited to anger experience but also manifested itself in a different way of dealing with these angry emotions: APDs indicated a disturbed AR compared to both, HCs *and* INCs, who again did not differ from each other. However, APDs' AR could not be clearly assigned to a specific pattern of over- or underregulation (e.g. see Low & Day, 2015). Although APDs indicated increased expression of angry feelings, which, upon first glance, points to underregulation, they did not report an overall decrease in AR effort, which would have been expected from an underregulated type. Instead, they reported an *increased* use of the (generally) maladaptive strategy anger suppression, while no abnormalities in the use of the rather adaptive strategy anger control were found. Altogether, APDs reported a comparable extent of regulation effort compared to HCs and INCs, but indicated less success. Hence, the concept of over-/underregulation is perhaps somewhat oversimplified. The current findings rather suggest that APDs' increased anger expression may be mainly due to an excessive use of maladaptive AR strategies. Given that APDs not only reported disturbed AR, but also the highest levels of habitual physical aggression (APDs > INCs > HCs as assessed with the AQ), the current results suggest not only the significance of anger experience per se, but also AR, for the origin of aggressive behavior (Anderson & Bushman, 2002; Roberton et al., 2012). However, for INCs' increased physical aggression, other factors have to be crucial, since INCs neither reported abnormalities in trait anger nor habitual regulation of angry affect. Hence, different mechanisms might contribute to aggression in APDs and INCs. This aspect will be taken up again later.

Interestingly, APDs' pattern of increased use of rather maladaptive strategies alongside an unremarkable use of adaptive strategies continued when broadening to other unpleasant emotions: Here, APDs again reported an over-engagement in generally maladaptive strategies. However, all but one of the (generally) maladaptive strategies, which were indicated to be used more frequently by APDs as compared to HCs, were also reported to be used more often by INCs. Hence, the question arises whether the increased use of catastrophizing, self-blame and acceptance (of the situation)¹⁸ merely reflected adjustments to the prison environment (see chapter 3.4). Future research needs to clarify this issue, for example by additionally recruiting

¹⁸ The strategy acceptance (of the situation) is not unambiguously classifiable. If the situation cannot be changed, then non-acceptance is rather maladaptive.

offenders on parole. Although both inmate groups reported to have been engaged in illegal activities, APDs reported to blame others more frequently than INCs (and marginally more frequently compared to HCs). This finding is in line with the diagnostic criterion of lack of remorse, as evident by rationalizations (American Psychiatric Association, 2013). So this symptom indeed appears to be specific for those offenders with APD. Of note, INCs reported an increased use of reappraisal as compared to HCs and APDs – though the latter group difference was only marginally significant. It is possible that this result mirrors INCs' compensation for their otherwise increased use of rather dysfunctional ER strategies and therefore reflects a somewhat more balanced appraisal of the current situation (e.g. with regard to detention) among INCs as opposed to APDs. If this is the case, this would again indicate a slightly more functional ER among INCs compared to APDs. However, future research is needed to replicate this finding. For example, this result could also indicate a slight tendency for overregulation of unpleasant feelings among INCs. With regard to APDs' choice of rather adaptive strategies, results were similar to habitual AR: Again, no differences were found between APDs and HCs. These results are consistent with prior research that emphasizes the significance of increased *maladaptive* strategy use relative to a decreased use of adaptive strategies for psychopathology (Aldao, Nolen-Hoeksema, & Schweizer, 2010).

Looking at ER abilities other than strategy use, APDs, but not INCs, reported overall difficulties in ER as compared to HCs. This also underlines ER impairments as a distinguishing feature between offenders with and without APD. Moreover, it provides an explanation as to why previous research failed to detect emotion dysregulation in offender populations (Garofalo et al., 2018; Gillespie et al., 2018). By neglecting APD diagnosis, existing ER deficits in this (more severely) impaired subgroup have probably been overlooked. Despite APDs' increased overall emotion dysregulation, at scale level, only group differences for impulse control were revealed: APDs, as compared to HCs and INCs, reported problems with behavioral control when distressed (i.e. impulse control difficulties). This might indicate that APDs are particularly vulnerable for impulsive behavior when they are in an unpleasant emotional state, thus from a therapeutic point of view, improving APDs' ER skills appears critical (for treatment recommendations see also subsequent chapter 5.3.2). In view of the analysis' rather low power to detect small to intermediate group differences¹⁹ and considering groups' (hardly or not overlapping) CIs, lacking group differences with regard to acceptance of own emotional

¹⁹ In a sensitivity power analysis using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) the (minimum) effect size required to be detected with a probability of $1-\beta \geq .80$ was calculated. However, as no power calculation was available for the Kruskal-Wallis test, the analysis for the parametric equivalent, the one-way ANOVA, was conducted. Based on $n = 3$ groups, a total sample size of $N = 103$, and a power of $1-\beta = .80$, critical effect size for the group effect was $d = .621$ for a significance value of $\alpha = .05$, and $d = .554$ for a significance value of $\alpha = .10$.

experiences, recognizing, describing and differentiating emotions (i.e. scale clarity), difficulties engaging in goal-directed behavior and overall confidence in one's own ER success (i.e. scale limited access) have to be interpreted with caution. Although an attempt was made to compensate the rather low power by also considering marginally significant effects, this approach does not rule out the possibility that existing deficits of small or intermediate effect size were nevertheless overlooked. This is problematic because – given the timeline of the emotion generation process (see chapter 1.2.2) – particularly emotional clarity, acceptance but also goal shielding (and flexibility) are crucial for a functional ER. Difficulties in these abilities could lead to a chain effect of emotion dysregulation. For example, a diminished emotional clarity leads to a lack of (or a “wrong”) goal activation, which in turn results in situationally inappropriate strategy choice or no ER effort at all. Emotion dysregulation would be highly likely to accumulate, which in turn might result in more severe outcomes. Hence, future research with increased sample size is necessary to reinvestigate whether or not APDs also suffer from impairments in basic ER skills. Regarding emotional clarity it might be advisable to use a different measure than the DERS in order to obtain more detailed information on the construct, for example the Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994). Using this instrument, previous research yielded deficits in APDs as compared to controls, at least for the very specific group of patients of a military hospital with APD as compared to other soldiers (Sayar, Ebrinc, & Ak, 2001). In view of the CIs, it might be also interesting to re-examine whether INCs are affected by overall emotion dysregulation after all – though maybe to a lesser degree than APDs (see CIs for overall emotion dysregulation in the DERS).

Remarkably, the present findings support the assumption of APD being a disorder of ER not only when dichotomously assessing APD, but also when looking at it dimensionally (i.e. antisocial symptom severity) and when exclusively sampling incarcerated offenders. Even after controlling for basic variables, which are known to be associated with APD (age, IQ, SUD/AUD comorbidity), ER impairments were a robust predictor of antisocial symptom severity. ER alone accounted for an additional 18% of the variance within antisocial symptom severity, which corresponds to a strong effect. However, only an increased use of expressive (anger) suppression and a decreased use of reappraisal (marginally) contributed to antisocial symptom severity. Nonetheless, this finding is consistent with prior research emphasizing negative outcomes of suppression and underlining the (overall) functionality of reappraisal (e.g. John & Gross, 2004).

5.1.2. Intact Spontaneous Anger Regulation or Overlooked Deficits due to Methodological Shortcomings?

Results regarding habitual AR and ER strategy use were not transferable to the actual regulation situation in the lab. Contrary to expectations, no group differences were revealed either for (1) angry feelings and arousal prior to the AI, (2) AR success (i.e. anger and arousal reactivity) or (3) the specific ER strategies used during the CAT. Although these differing findings between habitual and spontaneous ER may seem surprising at first glance, they are not unusual from an empirical point of view (e.g. see Schreiner et al., 2020). Diverging results do not need to be a contradiction in terms, considering that an (alleged) social interaction during an online game (the Cyberball environment used for assessing spontaneous ER) is *one* very specific situation, hardly comparable to overall, and more authentic, everyday interactions (reference for the assessment of habitual ER). However, it is also possible that methodological restrictions account for the unexpected findings. Therefore, the current results regarding spontaneous ER must be discussed in the context of the methods applied.

With respect to the null findings of (1) angry emotions prior to the AI, it has to be considered that only one component of the emotional reaction (see chapter 1.2.2) was measured – the subjective experience, i.e. the “feeling”. By contrast, state anger as assessed with the STAXI-2, for which group differences have been found, goes beyond the mere feeling. Here, a second component of the emotional response is assessed, the urge to act, i.e. behavioral response tendencies. And indeed, different results occurred in terms of behavioral data: APDs showed increased resource aggression before the AI, i.e. during baseline rounds of the CAT (see following chapter 5.1.3). Hence, apparently contradictory results between anger ratings prior to the CAT and in the STAXI-2 should not be overrated, on the one hand because of the different constructs measured and on the other hand because of the observed floor effects in state anger as assessed with the STAXI-2 (see chapter 3.3.1).

As far as (2) anger reactivity is concerned, it must first be noted that our AI actually worked, which was already suggested by the preliminary studies (see chapters 2.2.2 and 2.3.2) and has been proven again by a significant increase in anger experience across groups in the main study. However, and contrary to expectations, no group differences in reactivity were revealed. Hence, this study found no evidence for diminished AR success in APDs. This is quite astonishing given the above described results on habitual AR. Nonetheless, the current findings are in line with the only comparable study known to me (Lobbestael et al., 2009), and even underline these results by now using a measure with high internal validity (for criticism on Lobbestael et al.’s (2009) stress-induction interview see chapter 3.1.1). However, the current

methodology also had several shortcomings, some of which may have partly caused the present results. Therefore, alternative explanations must be considered when interpreting the current null findings: First, it should be noted that the anger assessments – particularly those after the CAT – may have been prone to bias. It is quite possible that participants who were coded as “deceived”, saw through the deception *after* the last round of the Cyberball game and therefore “embellished” their anger ratings (however, the current study did not capture credibility for this point in time). Second, a general incentive to report low(er) anger ratings could have been the attempt to stabilize own self-esteem by not admitting that the previous conversation had an emotional effect on oneself. Masculine norms, such as “emotions make vulnerable”, “showing emotions is a weakness” and the like might be even more pronounced in prison populations (Laws & Crewe, 2015). Hence, biases in angry affect ratings could have masked potential group differences (even independent of the credibility of the cover story). Given that psychophysiological measures have been previously shown to be more sensitive to changes in affect/arousal (Lobbestael et al., 2008) and are less prone to deliberate bias, subsequent research may consider to additionally conduct psychophysiological measures such as cardiovascular reactivity and skin conductance level. Third, despite these limiting factors which could have underestimated the true impact of the AI, the nature of the AI must certainly be also questioned. People are vulnerable to different types of provocations (Jones, Joyal, Cisler, & Bai, 2017), depending on their values and goals. Given the fact that our provocation was rather weak and impersonal (which was indeed intended due to legal constraints, ethical concerns and approval procedures), existing abnormalities in APDs that occur when there is stronger provocation from significant others with more personal impact may have been overlooked. However, this objection represents a more general problem regarding AIs and is not easy to solve – especially when studying prison populations (conflicting ethical, legal and practical issues). The ideal AI/aggression paradigm has not yet been developed (McCarthy & Elson, 2018; Ritter & Eslea, 2005) and it is to be assumed that this issue cannot be entirely resolved in the foreseeable future – if only in view of the fact that increased external validity is inevitably accompanied by reduced internal validity.

With regard to (3) strategy use during the spontaneous ER, no group differences have been found either. As indicated above, the regulation attempt in the lab is a very specific snapshot. Consequently, it may map everyday regulation attempts more or less well – depending on how representative the ER situation in the lab is for participants’ everyday life. Hence, the above mentioned concerns regarding the nature of the AI (less suitable for APDs?) also apply here. Second, the possibility of response bias due to seeing through the cover story

is particularly true here. Moreover, the choice of our items has to be criticized, as, for example, indicated by quite low internal consistencies. Future research faces the challenge of selecting ER strategy items that are not only already validated by prior research but which are also appropriate to the respective regulation context in the lab. Next, and perhaps most importantly, participants' behavior during the strategy assessment as well as their fast RTs questions their motivation in having answered the items to the best of their ability, in the sense of a cognitive effortful optimizing (Krosnick, 1991). Instead, it seems likely that they tried to finish the questionnaire as fast as possible. Hence, problems with response quality due to so-called satisficing behavior cannot be ruled out (for a review see Matjašič et al., 2018). It has been previously shown that for example non-differentiation in the use of rating scales is more common in less educated samples (Krosnick, 1991). An attempt was made to counter these problems of strong satisficing by applying a lower cutoff for RTs: overly short RTs were excluded based on considerations of minimum reading time and visual inspection of RT distribution (see chapter 3.3.3). However, it is likely that (strong to) weak satisficing behavior was nevertheless still present in the current data set. Nonetheless, the alternative to applying a more conservative cutoff seemed even more disadvantageous due to the exclusion of "good data".

Overall, it is possible that the current results on spontaneous ER are flawed in terms of methodology. Nevertheless, and despite the challenges and obstacles involved, it is right and important to address APDs' spontaneous ER anew. Habitual measures of ER neglect the specific situational context, the intensity of experienced emotions and the success of ER attempts. These aspects are better captured in an actual ER situation in the lab. Only by combining both approaches can we obtain the most comprehensive picture of APDs' ER. Hence, future research should readdress spontaneous ER in APDs with a slightly different methodological approach.

5.1.3. Too Much and Too Little – Miscellaneous Abnormalities in Aggressive Behavior among Offenders with and without Antisocial Personality Disorder

The current study found evidence for deviating aggression patterns among both APDs and INCs, though different in nature, and depending on the measures used or rather the specific aggression form assessed: When looking at *habitual physical* forms of aggression (as assessed with the AQ), APDs and INCs reported increased aggressive behavior compared to HCs. This result is predominantly in line with previous findings (e.g. Garofalo et al., 2018; Graña et al., 2014; Timmermann et al., 2017). Furthermore, it expands prior research by showing increased

aggression in APDs as compared to INCs, thus again confirming diagnostic criteria of APD (American Psychiatric Association, 2013). Importantly, this study not only assessed inmates' self-reports regarding past aggression, but was the first to also use a behavioral measure to examine APDs' aggressive behavior during an actual, standardized provocation scenario. Remarkably, deviating patterns of *resource* aggression were found among APDs and INCs, suggesting a different aggression proneness and possibly different mechanisms contributing to their "real-life" aggression. Both groups' results will be discussed one after the other.

APDs behaved more aggressively than HCs and INCs in a situation where there were no clear incentives for aggressive behavior (i.e. during the baseline). However, and contrary to expectations, APDs as compared to HCs showed a normal pattern of reactive aggression (i.e. during the AI). Hence, APDs exhibited an increased spontaneous aggression, while the presence of provocations (insults, social exclusion and unfair treatment) led to an alignment of APDs' and HCs' aggressive behavior, resulting in comparable levels of reactive aggression. Different explanations come to mind for APDs' unremarkable reactive aggression. In the current study there was no face-to-face interaction between the aggressor (the participant) and the (alleged) aggrieved party (the other players), with the participant being able to remain unidentified. Hence, a rather distant, indirect form of aggression was assessed (Parrott & Giancola, 2007). Given that distance is associated with a decreased threshold for aggressive behavior (Haslam, Loughnan, & Perry, 2014) and the fact that no severe harm was inflicted (mild form of resource aggression), it is likely that particularly HCs' *inhibition* (see I³ Model, chapter 1.1) was reduced. That is, their motivation to override the proclivity to aggress was probably low. However, during baseline there was no "objective" reason to aggress due to the lack of *instigators* (see I³ Model), only during the AI environmental stimuli were present that normatively increase aggressive urges. Hence, due to the interactive effects of instigator and inhibition (see Perfect Storm theory, chapter 1.1), HCs' presumably low inhibition had no impact on their behavior in the absence of instigators (i.e. during the baseline) but was only behaviorally controlling during the simultaneous presence of instigators (i.e. during the AI). This might have resulted in the observed alignment of APDs' and HC' aggressive behavior during the AI. This assumption is in line with the fact that HCs showed the strongest increase in aggressive behavior due to the AI. However, it seems plausible that HCs' inhibition would have been greater when the punishment decision had been about a more serious form of harm and/or the distance to the (alleged) aggrieved party had been less. Hence, the current results on reactive aggressive behavior are likely to have underestimated the true differences between APDs and HCs

regarding more direct and more severe forms of “real-life” aggression. This is also suggested by APDs’ heightened reports of *habitual physical* aggression.

With respect to spontaneous aggression, it is all the more remarkable, that an increased aggression proneness of APDs as compared to HCs and INCs was evident even for such a mild form of aggressive behavior (where the inhibition in *all* groups should be relatively low). In view of the I³ Model, no clear instigating triggers were present. Hence, the current results could point to increased dispositional *impellers* in APDs as compared to HCs. With respect to the current data, APDs’ (a) increased dispositional aggressiveness (see AQ) might be one impelling factor contributing to the observed spontaneous aggression. Another impellor might be APDs’ (b) increased trait anger, i.e. their dispositional anger experience and anger impulse (see STAXI-2). Hence, APDs may have exhibited more unpleasant emotions during baseline as compared to the other groups, which therefore was their incentive to punish. Unfortunately, the methodology does not allow to draw definite conclusions regarding this matter, since there is no information on participants’ angry emotions at that specific point of time. Though it seems questionable whether these self-reports would have been sensitive enough to actually reveal (existing) group differences at all (see criticism in the preceding chapter 5.1.2). Given that emotions are sometimes only experienced in the urge to act, the current measure most likely assessed the behavioral correlate of anger, aggression, and had the advantage of being independent of emotional awareness and/or biased response styles. Further, it is possible that APDs, but also the other participants, experienced some kind of frustration already during the baseline: Due to the (faked) bug in the chat function that prevented participants’ messages from being sent, frustration and helplessness could have resulted. Accordingly, (c) a misinterpretation of arousal (see excitation transfer theories, chapter 1.1) and (d) a hostile attribution bias (see information processing theories, chapter 1.1) could then explain, why (particularly) APDs reacted with aggressive behavior. Especially with simultaneous (e) increased behavioral impulsivity (see BIS-15) this could have caused the (rash) punishment decision. Another impelling factor that could have influenced APDs’ behavior is (f) an increased psychopathic tendency. This could have led to increased punishment during baseline (e.g. due to callous-unemotional traits and sensation seeking), potentially alongside reduced aggression during the AI (for reduced proneness to punish others in response to unfair behavior see Osumi et al., 2012). However, the study’s design is not able to infer on APDs’ psychopathic personality traits. Due to prison sessions’ time limits and the additional permissions required, no measure assessing psychopathy was conducted (the best validated instrument for assessing psychopathy, the Hare Psychopathy Checklist-Revised (Hare, 1991) requires a 60-90 minute

interview as well as access to penitentiary files). Hence, it cannot be ruled out, that psychopathic traits indeed biased results. However, it is important to note, that APD, but not psychopathy, is a psychiatric disorder and far more prevalent (Mokros, Hollerbach, Nitschke, & Habermeyer, 2017). Therefore, from a socio-political point of view, it seems more relevant to first broaden our understanding of APD. Nonetheless, future research should consider to additionally assess these personality characteristics.

Certainly, methodological reasons have also to be taken into account when explaining APDs' increased spontaneous aggression. For example, it is possible that APDs did not understand the response mode at first: The default setting of the slider used to set the amount of money that should be deducted from the other player was in the middle of the visual analogue scale (for an illustration of the slider see Appendix B). This corresponded to a punishment of 50 out of 100 possible. If the participants did not notice this peculiarity at the beginning, but thought the mid-point would represent a neutral response, then increased "aggression" values would have resulted during baseline, thus biasing results. Although this objection cannot be completely ruled out, it does not seem very plausible. If there were any misunderstandings regarding the end points of the visual analogue scale, this should have affected all groups in a similar way, thereby not producing a systematic effect.

With respect to INCs, a completely different pattern of aggression occurred: Compared to APDs, their aggression was consistently diminished. Compared to HCs, they showed similar spontaneous aggression – at a generally low level – and a *reduced* reactive aggression. These observations contradict the above-mentioned findings on increased habitual physical aggression compared to HCs. Hence, the question arises, as to how these discrepant findings can be explained. It is possible that INCs' diminished reactive aggressive behavior mirrors a reduced assertiveness. This is somewhat in line with their increased use of (some but not all) cognitive habitual ER strategies (e.g. reappraisal), thus reflecting aspects of overregulation – though this pattern is not completely consistent (see AR results). Future research is needed to clarify whether INCs, due to low assertiveness and overregulation, generate cumulated stressors that lead to high incentives (current stressor) and high impellers (exhausted self), while in the meantime their own resources are depleted, so that inhibition is low, which eventually leads to an aggressive outbreak (see Perfect Storm Theory; chapter 1.1).

In view of the many divergent results between APDs and INCs, the current study clearly suggests different mechanisms contributing to APDs' and INCs' aggressive behavior. APDs' seem to exhibit an increased aggression proneness, which can be explained by existing aggression theories. However, the influence of such potential mediators still has to be clarified

empirically. Group differences regarding reactive aggression are likely to have been underestimated by HCs' possibly low *inhibition* due to the specifics of the methodology. Clearly, future research has to readdress this issue by examining different forms of aggression.

5.1.4. Have We Been Overestimating the Importance of Cognitive Inhibitory Control?

Although it is assumed that APDs suffer from poor executive functioning (Ogilvie et al., 2011), the current study found no evidence for diminished IC among APDs as compared to INCs or HCs. Neither with respect to their overall (mean) performance level, nor their strategic top-down control after conflict (conflict adjustment effects). Deficits in other cognitive control functions (working memory and set shifting) could not be identified either. Moreover, no associations were found between a low IC and APDs' symptom domains trait anger, physical aggression, impulsivity and their overall symptomatology.

The current findings fall into a number of different results regarding APDs' IC, none of which is fully consistent with the others (the current results vs. Zeier et al., 2012 vs. Roszyk et al., 2013 vs. Schiffer et al., 2014). In view of the heterogeneity of the samples and measurements between these studies, the different results are not surprising and not contradictory in and of themselves. From a methodological point of view, these mixed findings rather highlight the dependency of research results on the specific samples assessed, the comparison group(s) chosen and the respective tasks conducted. Given that each task is in need for (slightly) different underlying skills, it is not even unusual for research to obtain different results for different measures of executive functions within one study, i.e. within the same sample (see Chamberlain et al., 2016; Pasion et al., 2018). Hence, when interpreting the current IC results, this, again, must be based on the specifics of the methodology. In terms of sampling, it has to be assumed that (a) the current diagnosis of APD was more objective and thus more reliable than that of Zeier et al. (2012), that (b) the entire sample was psychiatrically more thoroughly assessed than that of Zeier et al. (2012) and Roszyk et al. (2013), that (c) the APD sample was more representative than that of Roszyk et al. (2013), and that (d) the control groups were more appropriately selected than by Schiffer et al. (2014). The relative frequency of comorbid SUD/AUD and/or ADHD is not considered as a limiting factor of the sample, as these disorders are relatively characteristic for APDs (Black et al., 2010). Furthermore, this comorbidity should have led to group differences rather than preventing them, which gives additional weight to the current null findings. Consequently, the main point of criticism seems not to be the study's sampling, but the applied measures. As already outlined in chapter 4.4 the task measuring IC (a version the Stroop task) may have lacked the sensitivity to detect group

differences due to low task difficulty. However, it seems unlikely that this fact alone was responsible for the current null findings, as no group differences were found for the original version of the TMT either (no task variation conducted here). A reduced statistical power seems to be excluded for the Stroop task as well, whereas at first glance it might explain the null finding for the TMT²⁰. However, groups' CIs for the different TMT outcomes do not support this assumption but also suggest similar performances among groups. Nevertheless, future research should on the one hand increase the sample size to ensure sufficient power and on the other hand, the tasks should be more demanding (see also recommendation by Pruessner, Barnow, Holt, Joormann, & Schulze, 2020), for example by reducing the interstimulus interval and increasing the proportion of congruency. This could reduce the probability of ceiling effects, increase the discriminating power and, if effective, enable an analysis of accuracy, so that not only performance efficiency (i.e. regarding RTs) but also performance effectiveness (i.e. regarding accuracy) can be assessed.

The question remains, how the present results can be interpreted despite the methodological shortcomings. Since research to date, including the present work, provides no clear evidence that APDs suffer from deficits in IC, it is reasonable to assume that (1) overall IC is not as important for APD symptomatology as could theoretically be deduced.

Alternatively, it is possible that (2) IC *was* decisive for APDs' developmental course, but only until adolescence (Moffitt, 1993). Perhaps APDs' IC was only impaired during a critical time frame (note that all APDs have by definition suffered from conduct disorder during childhood/adolescence), but then became balanced during adolescence, when IC continues to mature (Diamond, 2013). From this point on other factors could have been relatively more decisive in maintaining or aggravating APDs' symptomatology, for example broken biographies, or, as suggested by the present work, emotion dysregulation (for the "chain of cumulative continuity" see Moffitt, 1993, p. 12). Slight indications for such a presumption indirectly comes from Ogilvie et al.'s (2011) meta-analysis: It is striking, that a high proportion of the evidence that revealed significant differences in executive functioning between ASBs and control groups comes from studies sampling children or adolescents only (see

²⁰ For the Stroop analysis, a sensitivity power analysis was carried out for the between-within interaction of the repeated measures ANOVA (i.e. group differences in Stroop interference scores). This analysis yielded a critical effect size required to be detected with a probability of $1-\beta = .80$ of $d = .096$, corresponding to a non-effect according to Cohen's (1988) classification (analysis based on the following parameters: $n = 3$ groups, a total sample size of $N = 102$, a significance value of $\alpha = .05$, $n = 2$ number of measurements, correlation among repeated measures $r = .95$, nonsphericity correction $\epsilon = 1$). For the TMT, a sensitivity power analysis was again conducted for the parametric equivalent of the Kruskal-Wallis test, the one-way ANOVA. Based on $n = 3$ groups, a total sample size of $N = 102$, a significance value of $\alpha = .05$, and a power of $1-\beta = .80$, critical effect size for the group effect was rather high with $d = .624$.

Supplementary Material of Ogilvie et al., 2011). Correspondingly, IC impairments might be a deficit which is subject to developmental change. Clearly, more research is needed to clarify the developmental course of APDs' IC.

Further, it is possible that (3) APDs exhibit deficits in IC, but only with regard to specific components: According to Diamond (2013), IC is rarely needed without working memory demands, both typically co-occur. However, the Stroop task used here was a relatively pure measure of IC and selective attention (Diamond, 2013), hardly including any requirements on working memory. Besides, the applied Stroop task did not provide information on proactive or reactive control mechanisms²¹ (see also chapter 4.4). Given that Iselin and DeCoster's (2009) work suggests that adolescents and young adult ASBs rely more on externally guided control mechanisms (reactive control) as opposed to internally guided control mechanisms (proactive control), it may be assumed that APDs show specific deficits in *proactive* control. This could result in an increased risk for goal-conflicting behavior (e.g. unlawful behavior) in situations where there is no external source indicating a conflict. Hence, both aspects – working memory and different modes of control – may be worth investigating in APDs.

Another explanation is that (4) IC deficits only emerge in situations with simultaneous emotional processing. Current research considers emotion and cognition not only as separable processes but emphasizes an integrative perspective (e.g. Gray, 2004). So, on the one hand, it is assumed that impaired cognitive control contributes to maladaptive ER (Ochsner & Gross, 2005; Pruessner et al., 2020; Tang & Schmeichel, 2014). However, the current study challenges this assumption by revealing ER deficits but no IC impairments in APDs. Furthermore, no associations between inmates' IC (Stroop interference) and their trait anger (AQ) have been found. On the other hand, it is also plausible that ER engagement exhausts cognitive resources, so that simultaneously operating cognitive (inhibitory) control abilities are impaired. And indeed, within healthy adults, consistent impairments in cognitive control have been found

²¹ Comparing reactive and proactive control would have required varying congruency proportion: A high proportion of incongruent and a low proportion of congruent trials leads to a relatively high goal support (i.e. indicating the font color, not the word), since conflicts (i.e. incongruent trials) occur frequently. Therefore, a high degree of proactive control is likely to be recruited (i.e. using previously observed context information to anticipate conflict and prepare for the appropriate behavioural response; Braver, 2012). As a result, conflict processing is facilitated compared to high congruency Stroop tasks, resulting in less interference overall (Bugg et al., 2011; Bugg et al., 2008). By contrast, with a low proportion of incongruent trials and a high proportion of congruent trials, the task goal is unlikely as actively maintained as in the previous example. Here, reactive control is likely to be used to a greater extent. In other words, participants react predominantly on the basis of the immediate environmental context. Participants usually show stronger interference in high congruency Stroop tasks, since it is hardly possible to prepare for conflicts, i.e. using proactive control (Bugg et al., 2011; Bugg et al., 2008). Although in our task incongruency was not excessively frequent, it was not rare either (50 : 50 probability). This may have contributed to the low level of task difficulty. However, since congruency proportion (e.g. 80 : 20 vs. 20 : 80 probability) was not varied, it cannot be determined, whether groups differed regarding the preferred mode of control.

during simultaneous, task-irrelevant aversive emotional processing (for a review see Mueller, 2011). Hence, APDs' cognitive (inhibitory) control may be specifically impaired when experiencing angry affect – to which they have a general tendency. Such an association would be extremely unfavorable, as control of goal-oriented behavior is particularly important when it comes to emotionally impressive information (e.g. not running away from a grizzly bear, not attacking the wife's lover; cf. Mueller, 2011). Hence, the interaction of emotion (particularly anger) and cognition (particularly IC) seems to be of special interest within APDs (particularly regarding their aggressive behavior). Despite the high relevance, there are, to the best of my knowledge, no findings on emotion-specific cognitive control in offenders with APD. Although an attempt was made to examine emotion on cognition during the present AI/aggression paradigm (CAT), the efforts unfortunately did not prove to work²². Therefore, it is up to subsequent research to further explore the interplay of angry affect and cognitive (inhibitory) control in APDs.

Overall, and despite the methodological limitations as well as the possible alternative explanations, it has to be noted that there is still no compelling evidence indicating IC deficits among APDs. Obviously, IC deficits are not as characteristic for adult APDs as, for example, emotion dysregulation.

5.2. Limitations

Besides the restrictions already mentioned, other limitations have to be considered when interpreting the results of the present thesis. In addition to natural conceptual restrictions, these limitations mainly concern sampling features as well as methodological shortcomings.

²² Aggressive behavior (the punishment decision) was coupled with a potential consequence (no consequence vs. risk for own monetary loss), in order to measure the inhibition of aggression, when this behavior is associated with a goal-conflicting negative consequence. It was assumed that cognitive control mechanisms lead to reduced punishing behavior when this behavior is associated with a negative consequence, as compared to a decision without a negative consequence (i.e. cognitive control inhibits the goal-threatening behavioral response). However, to represent a measure of cognitive control, it would have required participants' desire to maximize their financial compensation (i.e. goal activation). Otherwise, the "negative" consequence would have been no actual negative consequence and there would have been no need for behavioral adjustments due to the factor consequence (i.e. no conflict). An effect of consequence was only expected during the AI (the anger rounds), as only here, an incentive to punish was assumed (expected floor effects during baseline).

Unfortunately, manipulation checks revealed that participants hardly cared about their reward. Hence, the intended negative consequence (i.e. risk for own monetary loss) had no influence on participants' punishment decisions (see chapter 3.3.3) and our measure was not able to assess cognitive control.

5.2.1. *Beyond the Scope of the Current Work*

While the object of the investigation was relatively clearly outlined with regard to IC by definition, this was not the case for the broader areas ER and aggression. Here, the current study was inevitably faced with the challenge of limiting the scope of the research to a practicable level. Therefore, there are some areas mentioned below, which this study cannot offer further information on. It would however still be interesting to investigate these areas further.

First, and according to Gross' (1998) classification (see chapter 1.2.2), the current work was about intrinsic, explicit ER only. Hence, extrinsic or implicit ER was not addressed. Second, in view of the multi-aspect approach of emotions it has to be noted that the current work focused on the subjective experience (the "feeling") and in parts on perceptual-cognitive processes (e.g. ER skills such as emotional acceptance) and behavioral response tendencies (e.g. aggression). However, neurobiological (e.g. brain activation) and psychophysiological processes (e.g. cardiovascular reactions) were beyond the scope of the present work. Third, the study mainly assessed *cognitive* ER strategies and mainly those that can be assigned to attentional deployment, cognitive change and response modulation. Hence, future research might particularly examine behavioral strategies and/or strategies of the situation selection and situation modification family, i.e. strategies before the emotion generation process has fully run its course (see chapter 1.2.2). Such strategies are among the most foresighted and (co-) determine whether an individual is at all faced with an emotionally critical situation (Gross, 2015). Therefore, it might be interesting to study how APDs contribute to finding themselves in anger-evoking situations by examining their use of proactive ER strategies. Given that experience in cognitive behavioral therapy suggests that the acquisition of rather behavioral forms of ER is relatively efficient (e.g. as compared to purely cognitive ER strategies; cf. Jacobson et al., 1996), addressing these strategies might offer a great potential for the adaptive regulation of one's own emotions. Fourth, and as already pointed out, research investigating aggression in the lab, including the current study, is only able to study mildly harmful forms (McCarthy & Elson, 2018). Our measure referred exclusively to indirect, active resource aggression (theft), neglecting other forms and subtypes such as physical aggression, verbal aggression, postural aggression (i.e. non-verbal acts such as making threatening faces or invading personal space) or damage to property (Parrott & Giancola, 2007). Fifth, since only spontaneous and reactive aggressive behavior of APDs was assessed, it is not possible to draw conclusions about proactive forms of aggression (a common fact in aggression research, see Tedeschi & Quigley, 1996).

5.2.2. *Sampling Issues*

There are several limitations related to our sampling, some of which have already been addressed in chapters 3.4 and 4.4. First, only male, incarcerated APDs were recruited. Thus, generalizability of the results to other APD populations such as women, non-incarcerated offenders or more “successful” APDs, who operate in the “dark field” of crime, is questionable. However, women as compared to men appear to be a less socially significant study sample in this field of research, due to their reduced criminal behavior (Statistisches Bundesamt, 2019b) and their reduced prevalence of APD (Fazel & Danesh, 2002), while “successful” APDs are unfortunately difficult to approach. Second, and despite the relatively low preselection by prison staff, the current sample was by no means representative of the prison population. Many potential participants refused to participate, for example due to distrust in the investigator (fear of cooperation with law enforcement authorities), reservations regarding psychologists in general (“manipulating”), gang rules (refusal of signatures in prison, not even regarding informed consent of study participation), lacking motivation or masculine norms – i.e. samples, that would have been of particular interest. Unfortunately, this is a problem that future research must also face. Third, our APD diagnosis was based on an interview method only. Clearly, it is advantageous to additionally use third-party anamnestic information and/or include prison files, to which the current study unfortunately had no access. However, I am confident that our well-validated semi-structured interview (SCID-II) is preferable to a file-based diagnosis or the mere adoption of an earlier diagnosis without an own diagnostic assessment (such as in Roszyk et al., 2013). This ensured that APD was a current diagnosis, which is important due to the high remission rates of this personality disorder (American Psychiatric Association, 2013). The lack of a clearly specified reference period for APD diagnosis is unfortunately not unusual in research, but significantly complicates integration of results across studies. Fourth, it could be criticized that participants with a former (but not current) diagnosis of APD were included in the INC group. As a result, group differences between inmates (APDs vs. INCs) could have been underestimated. However, this strengthens the significance of group differences that *have* been found. Fifth, it might be objected that the APD sample differed from INCs in terms of age, and from HCs in terms of citizenship. However, these group differences were to be expected: Decreased age in APDs as compared to INCs is in line with previous research (Black et al., 2010; Graña et al., 2014) and is possibly due to symptom reduction of APD with increasing age (see American Psychiatric Association, 2013). The relatively large proportion of non-German participants among the inmate sample (approximately one-third) reflected the distribution in German prison populations (Statistisches Bundesamt, 2019b). Importantly, sufficient language

skills were assured prior to study enrollment and participants with low verbal IQ were excluded (see chapter 3.2.1). Sixth, personality disorders other than APD were not assessed. Hence, although unlikely, it is possible that some of the results were due to comorbid personality disorders, such as borderline personality disorder, which is determined by severe emotion dysregulation (e.g. Daros & Williams, 2019), or narcissistic personality disorder, considering that aggression can result from threats to high self-esteem (Anderson & Bushman, 2002). Therefore, future research should not only gather information about psychopathy (see chapter 4.4), but also about personality disorders other than APD. Seventh, this study's design (the CAT) was in need for naïve participants (use of a cover story). However, participants were fully debriefed during the session. Hence, it is possible that former inmate participants informed interested inmates about the objectives of this study. However, precautions have been taken to avoid such exchanges between inmates as far as possible: On the one hand, participants were asked not to discuss study details with other interested inmates during the following days. They were given suggestions as to what they could tell about the study instead. Due to the careful shaping of relationship (respectful interactions, empathic interest) and participants' frequently observed masculine norms and values (e.g. "helping the young lady", not being a "snitch", keeping one's "mouth shut") and the foreseeable period of time for which they were asked not to disclose the study, it seemed realistic that (the majority of) participants corresponded to this request. On the other hand, a short screening, intended to find out whether participants were already briefed about sensitive study contents, was conducted before study enrollment. Furthermore, the credibility check following the CAT also addressed this issue. In addition, all sessions within a prison were conducted on consecutive days to reduce the likelihood of (former) participants being able to interact with interested inmates. Additionally, within the Bavarian prison, where most inmates were recruited, the sessions were conducted according to location aspects (first one prison wing/work area and then the next), once again to limit the time window for the exchange of information.

5.2.3. Further Methodological Criticism

Further limitations concern procedural, methodological and statistical aspects. Regarding the procedure, disturbances during the sessions with inmates could not be fully avoided due to the specifics of a prison environment (see also Velotti et al., 2017). As a consequence, there have been slight procedural differences between the sessions of inmates as compared to those of HCs. While it is unlikely that disturbances had an influence on inmates' self-reports in the questionnaires, it might have reduced their performance during the Stroop

tasks. And yet, no impairments have been found compared to HCs – which contradicts this assumption. The fact that the investigators have not been blind to the studies hypotheses, might have indeed biased results (Rosenthal effect; Rosenthal & Fode, 1963). However, the dependent variables themselves were quite objective, i.e. they did not depend on the assessment of the investigators.

In terms of the methods used, it could be argued that the use of self-report instruments may have led to a distortion of the results by the presence of awareness biases, social desirability and confirmation biases. Regarding awareness bias, however, no group differences have been revealed with respect to emotional clarity, thus indicating comparable levels of emotional awareness across groups. Although it is certainly questionable to what extent someone with little introspective capacity is aware of his or her lack of awareness. Hence, and as already outlined, when assessing spontaneous ER in the lab, future studies might consider to additionally use psychophysiological measures in order to assess changes in arousal more objectively. However, it has to be noted that the most significant aspect of an emotion is the experience component, which is, by definition, subjective. Another approach would be to assess introspective capacity more thoroughly to be able to take into account a possible lack of awareness for further interpretations (e.g. by using the Toronto Alexithymia Scale; see also chapter 5.1.1). With respect to social desirability, the current results indicated that APDs were less prone to socially desirable response tendencies. It can therefore be assumed that group effects, particularly between APDs and HCs, may have been underestimated rather than overestimated. However, it cannot be completely ruled out that demand characteristics biased results during the CAT. This issue was (only) indirectly addressed with our credibility check.

Furthermore, there are several limitations that apply to our experimental methods, the CAT and the Stroop task. Some criticism has already been mentioned in chapters 3.4 and 4.4, and at the beginning of this chapter (chapters 5.1.2, 5.1.3, and 5.1.4) and shall therefore not be repeated in detail (brief overview: only minor increases in angry affect due to the AI, anger assessment's proneness to bias, questionable generalizability to other forms of aggression, low task difficulty in the Stroop task and accordingly no analysis of accuracy, assessment of an isolated IC component only). However, there are further limitations regarding the CAT that concern construct validity: First, and as in prior aggression research (cf. e.g. Tonnaer et al., 2019), it was omitted to assess participants' intentions for the punishment behavior. However, this would have been crucial to ascertain whether participants' behavior was indeed an aggressive act (McCarthy & Elson, 2018; Ritter & Eslea, 2005; Tedeschi & Quigley, 1996). Thus, although unlikely, it cannot be completely ruled out that some participants had other

intentions than harming the other players (e.g. to teach a lesson, i.e. communicating that they should not behave the way they did). Hence, it is essential for future studies to assess participants' underlying motivation for and goals of the (alleged) aggressive behavior. Second, it might be argued as to whether or not the punishment itself was actually a harmful act. Clearly, the aggressive behavior assessed in the present study was at the very low end of the range of harmful behavior and generalization to other forms of "real-life" aggression is questionable (McCarthy & Elson, 2018). Nonetheless, money deduction is obviously a harmful act, albeit only monetary. Certainly, the current results regarding the behavioral measure of aggression are preliminary and have to be interpreted with caution. Nevertheless, they are the first to *show* abnormalities in APDs as compared to HCs and INCs – for an identical triggering situation (achieved by using a standardized, thus internally valid, instrument), in terms of a relatively objective assessment of aggression (examining *behavior*, not just self-reports) and even with respect to a very mild form of aggression (where group differences are probably less pronounced). It is now up to future research to replicate or refute the current findings.

Finally, there are some restrictions concerning the statistical approaches. Our overall sample size was limited, thus decreasing the power to detect effects^{19,20}. This was a problem specifically for the analysis of the CAT, where the additional between-group factor credibility (deceived, not deceived) was included (unfortunately, G*Power 3 does not allow a power analyses for the four-way mixed design ANOVA). It was therefore decided to not only report effects on the $\alpha \leq .05$ significance level, but also marginally significant effects with $\alpha \leq .10$. Further, no alpha-level adjustments were conducted for hypothesis testing. Whereas this approach can be clearly criticized, it has, however, to be considered that the conventional use of the 5% level for determining statistical significance is arbitrary (overstated: it is used because Ronald Fisher suggested it, cf. Hackshaw & Kirkwood, 2011). When the standard error is quite high (probably due to low sample sizes as in this study), even moderate effects will result in a borderline *p*-value (Hackshaw & Kirkwood, 2011). Hence, given the specificities of this study, it seemed more appropriate to use an increased significance level and to additionally consult effect sizes. Interpretations were then based on both, significance values, effect sizes and, if available, CIs. The interpreted marginally significant effects were relevant effects (Cohen, 1988) without exception and thus support the current approach (all interpreted *ds* $\geq .336$). A clear limitation of the current work is that despite violated assumptions, parametric approaches were used in some cases. However, transformation of data was not successful in approximating normality more closely and no alternative procedures were available in SPSS (e.g. for the four-way mixed ANOVA). Hence, future research is needed to replicate the current findings. This

applies in particular to regression analysis, where a large number of predictors was used ($n = 10$ predictors, risk for overfitting). As mentioned before, the current results are only preliminary and no definite evidence, but they are an important first step in ER and aggression research among inmates with and without APD and provide useful future directions for following research.

5.3. Implications and Future Perspectives

Implications for future research arise mainly from aforementioned limitations and have been largely outlined above (see chapter 5.1 or chapter 5.2.1). However, in order to provide a better overview, the most important research implications are again coherently summarized below. Afterwards initial suggestions for treatment options are derived.

5.3.1. Future Research

Considering the assumed link between unpleasant emotions and reoffending (for a review see Day, 2009), it seems essential to further improve the understanding of ER among offenders – particularly within the high-risk group of APDs (Shepherd et al., 2016). For one thing, APDs' (1) pattern of emotion misregulation (see chapter 5.1.1) should be readdressed by examining a broader set of ER strategies, not limited to the attentional deployment and cognitive change family. Moreover, it is necessary (2) to determine whether APDs only show performance deficits, i.e. they choose the “wrong” strategy, or whether they are also impaired by deficits in strategy implementation, i.e. they report less success when using the same adaptive strategies. As outlined above, (3) more objective psychophysiological (to detect changes in arousal) and neurobiological measures (to detect possible downregulation of amygdala activation and possibly deviating patterns of prefrontal activations during different regulation attempts) could be assessed in addition to self-reports. Furthermore, it would be advisable (4) to consider the specific contexts, in which ER strategies are applied. Given that ER strategies are not per se adaptive or maladaptive (e.g. Gross, 2013; McRae & Gross, 2020) it is not only the mere frequency of strategy use and its correct implementation that determines ER success but rather the flexible, context-sensitive application of strategies (Gratz & Roemer, 2004). Further, with respect to ER failures and overall emotion dysregulation, (5) APDs' impairments need to be specified, whereas INCs' (lack of) more general ER abilities needs further confirmation (see also chapter 5.1.1).

Regarding (behavioral) aggression, the current results are in need for replication by using slightly modified measures (outlined in chapters 5.1.3 and 5.2.3). It will remain a challenge for future research to develop appropriate AI and aggression paradigms that are both, ecologically valid, standardized and adequately intense, but still suitable for a prison context. Studies using different paradigms assessing various forms of aggressive behavior could then, in the aggregate, add up to a more comprehensive understanding of APDs' aggression (McCarthy & Elson, 2018). While specific impellers among APDs (increased trait aggressiveness, chronic anger, habitual impulsivity, misinterpretation of arousal, hostile attribution bias and psychopathic personality traits) and INCs (reduced assertiveness, overregulation) have been suggested, the current study is by no means able to infer on causal factors. Hence, future research needs to investigate the role of these potential mediators in order to clarify, *why* APDs and INCs showed an abnormal aggression pattern. Only when a clear (causal) relationship between these (probably diagnosis-specific) risk factors and aggression is established, would it be possible to make empirically grounded decisions for treatment programs that not only address APD symptoms but also tackle the underlying causes of aggressive behavior. Unfortunately, longitudinal studies that accompany a cohort from childhood to adulthood are very costly (e.g. very large sample sizes required). Another interesting alternative approach, could be to conduct longitudinal studies on released offenders by using ecological momentary assessment. This way, risk factors of reoffending could be identified (however, for challenges see Burke et al., 2017).

Given that research thus far has yielded no clear evidence for (isolated) IC deficits in APDs, it seems appropriate to take this research field to the next step. It is clearly recommended to (1) examine IC with higher task demands (see also Pruessner et al., 2020). Furthermore, it (2) should not only be assessed in isolation but for example in combination with working memory (see chapter 5.1.4). Such an approach seems promising, in that corresponding results could point to specific interventions (Hoorelbeke, Koster, Vanderhasselt, Callewaert, & Demeyer, 2015; Iselin, DeCoster, & Salekin, 2009). Moreover, it would be interesting to address the continuous adjustment of cognitive (inhibitory) control not only by examining conflict adaptation effects, but by also (3) looking at post-error slowing as a measure for corrective behavior (Sullivan, Perlman, & Moeller, 2019). However, an appropriate level of task difficulty would be a prerequisite for this. Further, it is recommended to investigate (4) specific components of IC such as reactive and proactive control. A promising approach would be to use the AX-Continuous Performance Task (Rosvold, Mirsky, Sarason, Bransome Jr, & Beck, 1956), which not only distinguishes failures of reactive control from those of proactive

control, but also allows for variations on working memory demand. In addition, (5) the interplay between cognitive (inhibitory) control and ER may be addressed (see chapter 5.1.4).

Finally, in addition to the new research areas mentioned above, a general recommendation must also be made: Unfortunately, to date, it is very common in psychological research within offender populations to sample subgroups based on offender type only. While at first sight this approach appears attractive due to its relatively high practicability – file inspection is easier than carrying out a costly diagnostic assessment – it bears several disadvantages: Not only is such an approach often inaccurate due to incorrect assignments²³, but could also lead to false negative decisions, as deficits in APDs and/or INCs could neutralize each other and thus be masked. Such overlooked impairments could in the worst case prevent proper treatment programs. If offender samples are still only considered as a whole or if classification procedures are only carried out on the basis of offender type, researchers must, at the latest now, be aware that their results could be distorted by neglecting APD diagnosis (note that APD does not exclusively occur in violent offenders, and that not all INCs are non-violent offenders; see also self-reported lifetime offences of the current sample, depicted in Figure 8). This is particularly true for the field of ER and aggression, since here the study's most explicit differences between offenders with and without APD occurred. Hence, it is important for future research to explicitly address the psychiatric diagnosis of APD. Given that APD itself is still a heterogeneous group (Poythress et al., 2010), future research at best also assesses psychopathic personality traits, in order to be able to attribute potential abnormalities to APD itself and not only to APD's increased comorbidity with psychopathy.

5.3.2. Preliminary Treatment Recommendations

Since the current work is basic research, it is difficult to infer definitive treatment indications. Nonetheless, the current study underlines prior recommendation to individually tailor interventions to subgroups of offenders (e.g. Low & Day, 2015). Since inmates with and without APD indicated different deficits and thus different potential risk factors for offending, a disorder-specific therapy approach might be appropriate in prisons. Certainly, and as usual in cognitive behavioral therapy, interventions should not only focus on the patients' diagnoses but should also be biographically individualized to the patients' specific background conditions (Gall-Peters & Zarbock, 2012). However, it has to be noted that research results, including the current, are only informative at a superordinate group level and not for the individual case.

²³ By only considering extracts from the Federal Central Register or, worse, the index offence (as e.g. in Gillespie et al., 2018; Seruca & Silva, 2016) when categorizing participants, only (aspects of) the “bright field” of crime is assessed, reflecting only the tip of the iceberg and thus providing incomplete information.

Hence, the following suggestions for treatments represent disorder-specific treatment options on the group level only.

Although, to date, there is no evidence for decreased AR success among APDs based on laboratory results (the current study and Lobbestael et al., 2009), there are strong indications for impaired AR in their *usual* environment (the current study; Timmermann et al., 2017; Yavuz et al., 2016). Hence, the present findings clearly suggest to target the functional regulation of anger in APDs' everyday life by using anger management trainings (e.g. Steffgen & Dusi, 2006). However, these treatments should not only intend to reduce the behavioral expression of anger (aggression), or increase anger control, but also decrease an excessive use of maladaptive strategies such as anger suppression (see also Chambers, Ward, Eccleston, & Brown, 2008).

Given that APDs also indicated problematic ER beyond AR, interventions should not exclusively focus on the regulation of anger affects but also address overall ER. Since APDs' regulation pattern was predominantly characterized by increased use of maladaptive strategies – consistent with prior research emphasizing abnormalities in the use of maladaptive relative to adaptive strategies for psychopathology (Aldao et al., 2010) – it might be promising to target the decrease of these strategies. However, from a therapeutic point of view, it is not advisable to simply unlearn things. Instead, alternative and purposeful ways of coping must be learned in their place – if possible before “unlearning” the previous (dysfunctional) strategies. Hence, it seems desirable that interventions also focus on adaptive strategies, such as reappraisal, but also more proactive strategies (i.e. the situation selection/modification family). This would be particularly effective should it prove to be true that APDs suffer from a competence deficit in applying these adaptive strategies.

APDs reported to be vulnerable for impulsive behavior particularly when they are in an unpleasant emotional state. Therefore, it might not only be effective to directly improve their subjective emotional experience, but also to increase the acceptance of unpleasant emotions – even though the current study revealed no significant impairments in this respect (but see chapter 5.1.1): Increasing the general acceptance of emotional states could lead to a decrease in the subjective need to immediately terminate the emotional experience by dysfunctional impulsive behavior. Hence, therapy modules of Dialectical Behavior Therapy (Linehan, 2015) might be also suitable for APDs (however, for a general lack of randomized control trial studies for offenders see Tomlinson, 2018).

Further, APDs' indications for increased use of blaming others suggests the need for cognitive restructuring. By learning to take responsibility for one's own actions and taking the perspective of other's, aggressive urges could be reduced.

As far as INCs are concerned, future research needs to clarify, if pure anger control interventions might even be counterproductive for this subgroup, in that the tendencies of reduced assertiveness that could lead to aggression in the long-term, might be reinforced. If the hypothesis of reduced assertiveness for INCs proves true, it can be assumed that especially this subgroup of offenders could benefit from Social Skills Trainings, potentially with a focus on the situation types "assert one's rights" and "relationships" (e.g. Hinsch & Pflingsten, 2015).

The lack of cognitive (inhibitory) control deficits indicates both good and bad news. The apparent good news is that there are no severe cognitive impairments in either APDs or INCs that raise considerable doubts about the general effectiveness of other treatment programs, such as those mentioned above. However, the bad news is, that, at this point in time, there is no legitimate reason to expect that cost-effective and easily disseminated cognitive trainings (e.g. Pased Auditory Serial Addition Task-based training procedures, see Hoorelbeke & Koster, 2017) might be a promising prospect for offenders.

Clearly, further research is needed to specify APDs' impairments and risk factors in order to be able to infer definitive treatment recommendations. Next, randomized controlled trial studies could be conducted to explore the efficacy of the suggested interventions. As Tomlinson (2018) outlined, this might finally "lead criminal justice policy into an era of prison reform that has the unprecedented luxury of standing upon empirically supported approaches to offender rehabilitation" (Tomlinson, 2018, p. 91). In view of the high social relevance of criminality and recidivism, it should be worth it to us to (a) intensify research in APDs and to (2) intervene (as early as possible), i.e. take preventive and corrective actions. This would not only help the individual, but society as a whole, and lead to significant cost savings in the long term.

5.4. Conclusion

The overall purpose of this thesis was to identify deficits in self-regulation among APDs as compared to HCs. Furthermore, it was aimed to explore whether potential impairments in APDs merely reflect offending per se (APDs = INCs) or rather the psychiatric diagnosis (APDs ≠ INCs). Different aspects of self-regulation were assessed: ER, aggressive behavior and IC.

(1) In order to examine APDs' spontaneous AR and their aggression, a new AI and aggression paradigm was developed and carefully pretested prior to its application in the main

study (see chapter 2). These two preliminary analyses among male community participants provided initial support for the general suitability of the newly developed instrument, the CAT: the paradigm was able to induce angry emotions and achieve a sufficient variation of a mildly harmful form of resource aggression, depending on the presence or absence of the AI. After slight modifications, the instrument was subsequently used in the main study.

(2) Part I of the main study aimed to specify APDs' impairments in ER by not only considering their pattern of *habitual* ER (strategy use, emotion dysregulation and AR) but also their *spontaneous* AR following an AI (see chapter 3). Additionally, for the first time, abnormalities in the spontaneous and reactive aggression of APDs and INCs were investigated using an experimental approach. Results indicated that APDs, but not INCs, suffer from a chronic anger pattern, determined by increased trait anger and maladaptive AR (heightened habitual anger suppression and expression). In contrast to these impairments in their usual environment, APDs (and INCs) reported no abnormalities in AR success or strategy use following the AI in the lab, though methodological limitations have to be considered when interpreting these null findings. However, when considering the behavioral correlate of anger, i.e. aggressive behavior, different results occurred: In the absence of instigating triggers that normatively increase aggressive urges, APDs showed increased *spontaneous* aggression. Several mediators could have been responsible for APDs' heightened aggression proneness (e.g. increased trait anger, impulsivity, hostile attribution bias), but have yet to be confirmed empirically. INCs, on the contrary, showed reduced reactive aggression as compared to both, APDs and HCs. This lack of an appropriate behavioral response in reaction to unfair treatment may reflect a reduced ability in INCs to assert themselves, which, through an accumulation of stressors, may in turn increase the likelihood for aggressive behavior in the long term. Although the current results on behavioral aggression are only preliminary and provide no definitive evidence for "real world" aggression, they are the first to suggest divergent abnormalities in aggression patterns in APDs and INCs. With respect to habitual ER beyond anger, APDs, but not INCs, indicated overall ER difficulties, particularly with respect to impulse control. Although the current study yielded no evidence for ER failures (e.g. lack of clarity, reduced emotional acceptance) among both APDs and INCs, potential deficits might have been overlooked due to the study's limited sample size. Regarding habitual ER strategy use not limited to anger, APDs indicated more severe emotion misregulation than INCs. Again, this pattern was mainly characterized by an increased use of generally maladaptive strategies (instead of a reduced use of generally adaptive strategies). It is noteworthy that impairments in

habitual ER explained variability in inmates' antisocial symptom severity even above and beyond the influence of comorbid SUD/AUD, age and verbal intelligence.

(3) The key objective of part II of the main study (see chapter 4) was to clarify whether or not APDs suffer from deficits in IC, and if so, whether these deficits are specific to APDs (as opposed to INCs) and to IC (as opposed to other cognitive control abilities). Furthermore, it aimed to assess whether deficits in IC might underlie the symptom domain of APD, i.e. whether poor IC is associated with specific APD symptoms and an increased symptom severity among inmates. No evidence for a deficient IC was found as assessed by a computerized Stroop task, neither in APDs nor INCs, either with respect to overall performance level (RTs) or the strategic top-down control after conflict (post-conflict adjustments). No impairments were found with regard to more broad cognitive control abilities as measured by the TMT either. Furthermore, poor IC was not associated with inmates' level of antisocial symptom severity or specific symptoms of APD (increased trait anger, physical aggression and impulsivity). Despite the comparatively low task demands, these findings clearly challenge the assumption that particularly a diminished IC underlies APD. It is important to take this field of research to the next level not only by increasing task difficulty, but also by looking at specific components of IC, examining IC with simultaneous demands on working memory, and specifically assessing IC during an unpleasant emotional state. Corresponding results could then suggest further treatment methods.

In sum, the current work emphasizes that offenders are by no means a homogenous group. Instead, impairments in ER are a distinctive feature among offenders exhibiting APD. Consequently, APD should be considered as a disorder of ER, particularly, but not limited to, AR. The mechanisms that contribute to inmates' increased physical aggression may be quite diverse and may (partly) depend on the presence of APD. Future research is needed to replicate the current findings and expand on them to further improve our understanding of APD and its underlying causes. Only then will we succeed in deriving appropriate, empirically based treatment decisions for offenders with and without APD.

6. References

- Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review, 30*(2), 217-237. <https://doi.org/10.1016/j.cpr.2009.11.004>
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- Anderson, C. A., & Bushman, B. J. (2002). Human Aggression. *Annual Review of Psychology, 53*(1), 27-51. <https://doi.org/10.1146/annurev.psych.53.100901.135231>
- Babcock, J. C., Green, C. E., Webb, S. A., & Yerington, T. P. (2005). Psychophysiological profiles of batterers: Autonomic emotional reactivity as it predicts the antisocial spectrum of behavior among intimate partner abusers. *Journal of Abnormal Psychology, 114*(3), 444-455. <https://doi.org/10.1037/0021-843X.114.3.444>
- Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23-32. [https://doi.org/10.1016/0022-3999\(94\)90005-1](https://doi.org/10.1016/0022-3999(94)90005-1)
- Baliousis, M., Duggan, C., McCarthy, L., Huband, N., & Völlm, B. (2019). Executive function, attention, and memory deficits in antisocial personality disorder and psychopathy. *Psychiatry Research, 278*, 151-161. <https://doi.org/10.1016/j.psychres.2019.05.046>
- Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Barbour, K. A., Eckhardt, C. I., Davison, G. C., & Kassinove, H. (1998). The experience and expression of anger in maritally violent and maritally discordant-nonviolent men. *Behavior Therapy, 29*(2), 173-191. [https://doi.org/10.1016/S0005-7894\(98\)80001-4](https://doi.org/10.1016/S0005-7894(98)80001-4)
- Barnow, S., Reinelt, E., & Sauer, C. (2016). *Emotionsregulation: Manual und Materialien für Trainer und Therapeuten*. Berlin/Heidelberg: Springer.
- Baumeister, R. F., Vohs, K. D., DeWall, N. C., & Liqing, Z. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review, 11*(2), 167-203. <https://doi.org/10.1177/1088868307301033>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

- Berke, D. S., Reidy, D., & Zeichner, A. (2018). Masculinity, emotion regulation, and psychopathology: A critical review and integrated model. *Clinical Psychology Review*, 66, 106-116. <https://doi.org/10.1016/j.cpr.2018.01.004>
- Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin*, 106(1), 59-73. <https://doi.org/10.1037/0033-2909.106.1.59>
- Berkowitz, L., Corwin, R., & Heironimus, M. (1963). Film violence and subsequent aggressive tendencies. *Public Opinion Quarterly*, 27(2), 217-229. <https://doi.org/10.1086/267162>
- Black, D. W., Gunter, T., Loveless, P., Allen, J., & Sieleni, B. (2010). Antisocial personality disorder in incarcerated offenders: Psychiatric comorbidity and quality of life. *Annals of Clinical Psychiatry* 22(2), 113-120.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652. <https://doi.org/10.1037/0033-295X.108.3.624>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49-59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106-113. <https://dx.doi.org/10.1016/j.tics.2011.12.010>
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. Conway, C. Jarrold, A. Kane, A. Miyake, & J. Towse (Eds.), *Variation in Working Memory*. New York: Oxford University Press.
- Bugg, J. M., Jacoby, L. L., & Chanani, S. (2011). Why it is too early to lose control in counts of item-specific proportion congruency effects. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 844-859. <https://doi.org/10.1037/a0019957>
- Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition*, 36(8), 1484-1494. <https://doi.org/10.3758/MC.36.8.1484>
- Bundeskriminalamt. (2019). *Polizeiliche Kriminalstatistik Jahrbuch 2018, Band 4, Version 3.0*. Wiesbaden: Bundeskriminalamt.
- Burgess, R. L., & Akers, R. L. (1966). A differential association-reinforcement theory of criminal behavior. *Social Problems*, 14(2), 128-147. <https://doi.org/10.2307/798612>

- Burke, L. E., Shiffman, S., Music, E., Styn, M. A., Kriska, A., Smailagic, A., ... Rathbun, S. L. (2017). Ecological momentary assessment in behavioral research: Addressing technological and human participant challenges. *Journal of Medical Internet Research*, *19*(3), e77. <https://doi.org/10.2196/jmir.7138>
- Buss, A. H. (1961). *The psychology of aggression*. New York: Wiley.
- Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, *63*, 452-459. <https://doi.org/10.1037/0022-3514.63.3.452>
- Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 367-379. <https://doi.org/10.3758/cabn.7.4.367>
- Chamberlain, S. R., Derbyshire, K. L., Leppink, E. W., & Grant, J. E. (2016). Neurocognitive deficits associated with antisocial personality disorder in non-treatment-seeking young adults. *Journal of the American Academy of Psychiatry and the Law*, *44*(2), 218-225. <https://doi.org/10.17863/CAM.4940>
- Chambers, J. C., Ward, T., Eccleston, L., & Brown, M. (2008). The Pathways Model of Assault: A Qualitative Analysis of the Assault Offender and Offense. *Journal of Interpersonal Violence*, *24*(9), 1423-1449. <https://doi.org/10.1177/0886260508323668>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Daros, A., & Williams, G. (2019). A meta-analysis and systematic review of emotion regulation strategies in borderline personality disorder. *Harvard Review of Psychiatry*, *27*, 1. <https://doi.org/10.1097/HRP.0000000000000212>
- Davison, G. C., Robins, C., & Johnson, M. K. (1983). Articulated thoughts during simulated situations: A paradigm for studying cognition in emotion and behavior. *Cognitive Therapy and Research*, *7*(1), 17-39. <https://doi.org/10.1007/bf01173421>
- Day, A. (2009). Offender emotion and self-regulation: Implications for offender rehabilitation programming. *Psychology, Crime & Law*, *15*, 119-130. <https://doi.org/10.1080/10683160802190848>
- DeWall, C. N., Anderson, C., & Bushman, B. J. (2012). Aggression. In I. B. Weiner, H. Treannen, & J. Susls (Eds.), *Handbook of Psychology, Personality and Social Psychology* (2 ed., Vol. 5, pp. 449-466). New York: Wiley.

- DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *Journal of Personality and Social Psychology, 96*(1), 45-59. <https://doi.org/10.1037/a0013196>
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*(1), 135-168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*(1), 162-170. <https://doi.org/10.2307/1129603>
- Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT: Yale University Press.
- Donahue, J. J. j. g. c., Goranson, A. C., McClure, K. S., & Van Male, L. M. (2014). Emotion dysregulation, negative affect, and aggression: A moderated, multiple mediator analysis. *Personality & Individual Differences, 70*, 23-28. <https://doi.org/10.1016/j.paid.2014.06.009>
- Doyle, M., & Dolan, M. (2006). Evaluating the validity of anger regulation problems, interpersonal style, and disturbed mental state for predicting inpatient violence. *Behavioral Sciences & the Law, 24*(6), 783-798. <https://doi.org/10.1002/bsl.739>
- Dvorak-Bertsch, J. D., Sadeh, N., Glass, S. J., Thornton, D., & Newman, J. P. (2007). Stroop tasks associated with differential activation of anterior cingulate do not differentiate psychopathic and non-psychopathic offenders. *Personality and Individual Differences, 42*(3), 585-595. <https://doi.org/10.1016/j.paid.2006.07.023>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Perception & Psychophysics, 16*, 143-149.
- Etzler, S. L., Rohrmann, S., & Brandt, H. (2014). Validation of the STAXI-2: A study with prison inmates. *Psychological Test and Assessment Modeling, 56*(2), 178-194. Retrieved from <https://pdfs.semanticscholar.org/a892/faa76a000a974a001898ca28343019a0ccc8.pdf>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. <https://doi.org/10.3758/BF03193146>
- Fazel, S., & Danesh, J. (2002). Serious mental disorder in 23 000 prisoners: A systematic review of 62 surveys. *The Lancet, 359*(9306), 545-550. [https://doi.org/10.1016/S0140-6736\(02\)07740-1](https://doi.org/10.1016/S0140-6736(02)07740-1)

- Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th ed.). Los Angeles: Sage.
- Finkel, E. J. (2014). Chapter one - the I3 model: Metatheory, theory, and evidence. *Advances in Experimental Social Psychology*, 49, 1-104. <https://doi.org/10.1016/B978-0-12-800052-6.00001-9>
- Finkel, E. J., DeWall, C. N., Slotter, E. B., Oaten, M., & Foshee, V. A. (2009). Self-regulatory failure and intimate partner violence perpetration. *Journal of Personality and Social Psychology*, 97(3), 483-499. <https://doi.org/10.1037/a0015433>
- Finkel, E. J., & Hall, A. N. (2018). The I3 Model: A metatheoretical framework for understanding aggression. *Current Opinion in Psychology*, 19, 125-130. <https://doi.org/10.1016/j.copsyc.2017.03.013>
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (1996). *Structured Clinical Interview for DSM-IV*. Washington, DC: American Psychiatric Press.
- Fossati, A., Barratt, E. S., Borroni, S., Villa, D., Grazioli, F., & Maffei, C. (2007). Impulsivity, aggressiveness, and DSM-IV personality disorders. *Psychiatry Research*, 149(1), 157-167. <https://doi.org/10.1016/j.psychres.2006.03.011>
- 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>
- Gall-Peters, A., & Zarbock, G. (2012). *Praxisleitfaden Verhaltenstherapie: Störungsspezifische Strategien, Therapieindividualisierung, Patienteninformationen*. Lengerich: Pabst.
- Garnefski, N., Kraaij, V., & Spinhoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Personality and Individual Differences*, 30(8), 1311-1327. [https://doi.org/10.1016/S0191-8869\(00\)00113-6](https://doi.org/10.1016/S0191-8869(00)00113-6)
- Garofalo, C., Velotti, P., & Zavattini, G. C. (2018). Emotion regulation and aggression: The incremental contribution of alexithymia, impulsivity, and emotion dysregulation facets. *Psychology of Violence*, 8(4), 470-483. <https://doi.org/10.1037/vio0000141>
- Gillespie, S. M., Garofalo, C., & Velotti, P. (2018). Emotion regulation, mindfulness, and alexithymia: Specific or general impairments in sexual, violent, and homicide offenders? *Journal of Criminal Justice*, 58, 56-66. <https://doi.org/10.1016/j.jcrimjus.2018.07.006>
- Gottfredson, M. R., & Hirschi, T. (1990). *A General Theory of Crime*. Stanford, CA: Stanford University Press.

- Gottfried, E. D., & Christopher, S. C. (2017). Mental disorders among criminal offenders: A review of the literature. *Journal of Correctional Health Care*, 23(3), 336-346. <https://doi.org/10.1177/1078345817716180>
- Graña, J. L., Redondo, N., Muñoz-Rivas, M. J., & Cantos, A. L. (2014). Subtypes of batterers in treatment: Empirical support for a distinction between type I, type II and type III. *PLoS ONE*, 9(10), e110651. <https://doi.org/10.1371/journal.pone.0110651>
- Gratz, K. L., Moore, K. E., & Tull, M. T. (2016). The role of emotion dysregulation in the presence, associated difficulties, and treatment of borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment*, 7(4), 344-353. <https://doi.org/10.1037/per0000198>
- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, 26(1), 41-54. <https://doi.org/10.1023/b:Joba.0000007455.08539.94>
- Gray, J. R. (2004). Integration of emotion and cognitive control. *Current Directions in Psychological Science*, 13(2), 46-48. <https://doi.org/10.1111/j.0963-7214.2004.00272.x>
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2(3), 271-299. <https://doi.org/10.1037/1089-2680.2.3.271>
- Gross, J. J. (2013). Emotion regulation: Taking stock and moving forward. *Emotion*, 13(3), 359-365. <https://doi.org/10.1037/a0032135>
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26(1), 1-26. <https://doi.org/10.1080/1047840X.2014.940781>
- Gross, J. J., & Jazaieri, H. (2014). Emotion, emotion regulation, and psychopathology: An affective science perspective. *Clinical Psychological Science*, 2(4), 387-401. <https://doi.org/10.1177/2167702614536164>
- Gross, J. J., Sheppes, G., & Urry, H. L. (2011). Cognition and Emotion Lecture at the 2010 SPSP Emotion Preconference. *Cognition and Emotion*, 25(5), 765-781. <https://doi.org/10.1080/02699931.2011.555753>
- Gruber, J., Harvey, A. G., & Gross, J. J. (2012). When trying is not enough: Emotion regulation and the effort-success gap in bipolar disorder. *Emotion*, 12(5), 997-1003. <https://doi.org/10.1037/a0026822>
- Hackshaw, A., & Kirkwood, A. (2011). Interpreting and reporting clinical trials with results of borderline significance. *BMJ*, 343, d3340. <https://doi.org/10.1136/bmj.d3340>

- Hare, R. D. (1991). *The Hare Psychopathy Checklist - Revised*. Toronto, Ontario: Mulit-Health Systems.
- Harmon-Jones, E., Amodio, D. M., & Zinner, L. R. (2007). Social psychological methods of emotion elicitation. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of Emotion Elicitation and Assessment* (pp. 91-105). New York: Oxford University Press.
- Haslam, N., Loughnan, S., & Perry, G. (2014). Meta-Milgram: An empirical synthesis of the obedience experiments. *PLoS ONE*, *9*(4), 1-9. <https://doi.org/10.1371/journal.pone.0093927>
- Hawes, S. W., Perlman, S. B., Byrd, A. L., Raine, A., Loeber, R., & Pardini, D. A. (2016). Chronic anger as a precursor to adult antisocial personality features: The moderating influence of cognitive control. *Journal of Abnormal Psychology*, *125*(1), 64-74. <https://doi.org/10.1037/abn0000129>, <https://doi.org/10.1037/abn0000129.supp> (Supplemental)
- Herzberg, P. Y. (2003). Faktorstruktur, Gütekriterien und Konstruktvalidität der deutschen Übersetzung des Aggressionsfragebogens von Buss und Perry. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *24*(4), 311-323. <https://doi.org/10.1024/0170-1789.24.4.311>
- Hiatt, K. D., Schmitt, W. A., & Newman, J. P. (2004). Stroop tasks reveal abnormal selective attention among psychopathic offenders. *Neuropsychology*, *18*(1), 50-59. <https://doi.org/10.1037/0894-4105.18.1.50>
- Hinnant, J. B., & Forman-Alberti, A. B. (2019). Deviant peer behavior and adolescent delinquency: Protective effects of inhibitory control, planning, or decision making? *Journal of Research on Adolescence*, *29*(3), 682-695. <https://doi.org/10.1111/jora.12405>
- Hinsch, R., & Pfingsten, U. (2015). *Gruppentraining sozialer Kompetenzen GSK : Grundlagen, Durchführung, Anwendungsbeispiele*. Weinheim: Beltz.
- Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, *32*(1), 17-22. <https://doi.org/10.1080/00031305.1978.10479237>
- Holley, S. R., Ewing, S. T., Stiver, J. T., & Bloch, L. (2015). The relationship between emotion regulation, executive functioning, and aggressive behaviors. *Journal of Interpersonal Violence*, *32*(11), 1692-1707. <https://doi.org/10.1177/0886260515592619>

- Home Office. (2005). The economic and social costs of crime against individuals and households 2003/04. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/191498/Green_Book_supplementary_guidance_economic_social_costs_crime_individuals_households.pdf
- Hoorelbeke, K., & Koster, E. H. W. (2017). Internet-delivered cognitive control training as a preventive intervention for remitted depressed patients: Evidence from a double-blind randomized controlled trial study. *Journal of Consulting and Clinical Psychology, 85*(2), 135-146. <https://doi.org/10.1037/ccp0000128>
- Hoorelbeke, K., Koster, E. H. W., Vanderhasselt, M.-A., Callewaert, S., & Demeyer, I. (2015). The influence of cognitive control training on stress reactivity and rumination in response to a lab stressor and naturalistic stress. *Behaviour Research and Therapy, 69*, 1-10. <https://doi.org/10.1016/j.brat.2015.03.010>
- Iselin, A.-M. R., & DeCoster, J. (2009). Reactive and proactive control in incarcerated and community adolescents and young adults. *Cognitive Development, 24*(2), 192-206. <https://doi.org/10.1016/j.cogdev.2008.07.001>
- Iselin, A.-M. R., DeCoster, J., & Salekin, R. T. (2009). Maturity in adolescent and young adult offenders: The role of cognitive control. *Law and Human Behavior, 33*(6), 455-469. <https://doi.org/10.1007/s10979-008-9160-x>
- Izadpanah, S., Barnow, S., Neubauer, A. B., & Holl, J. (2019). Development and validation of the Heidelberg Form for Emotion Regulation Strategies (HFERST): Factor structure, reliability, and validity. *Assessment, 26*(5), 880-906. <https://doi.org/10.1177/1073191117720283>
- Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review, 2*(4), 363-370. <https://doi.org/10.1177/1754073910374661>
- Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., ... Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology 64*(2), 295-304. <https://doi.org/10.1037//0022-006x.64.2.295>
- Jehle, J. M., Albrecht, H. J., Hohmann-Fricke, S., & Tetel, C. (2013). *Legalbewährung nach strafrechtlichen Sanktionen. Eine bundesweite Rückfalluntersuchung 2007 bis 2010 und 2004 bis 2010*. Berlin: Bundesministerium der Justiz.

- John, O. P., & Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *Journal of Personality*, 72(6), 1301-1334. <https://doi.org/10.1111/j.1467-6494.2004.00298.x>
- Jones, D. N., & Paulhus, D. L. (2010). Different provocations trigger aggression in narcissists and psychopaths. *Social Psychological and Personality Science*, 1(1), 12-18. <https://doi.org/10.1177/1948550609347591>
- Jones, S., Joyal, C. C., Cisler, J. M., & Bai, S. (2017). Exploring emotion regulation in juveniles who have sexually offended: An fMRI study. *Journal of Child Sexual Abuse*, 26(1), 40-57. <https://doi.org/10.1080/10538712.2016.1259280>
- Joormann, J., & Vanderlind, W. M. (2014). Emotion regulation in depression. *Clinical Psychological Science*, 2(4), 402-421. [10.1177/2167702614536163](https://doi.org/10.1177/2167702614536163)
- Katsiyannis, A., Whitford, D. K., Zhang, D., & Gage, N. A. (2018). Adult recidivism in United States: A meta-analysis 1994–2015. *Journal of Child and Family Studies*, 27(3), 686-696. <https://doi.org/10.1007/s10826-017-0945-8>
- Kolla, N. J., Meyer, J. H., Bagby, R. M., & Brijmohan, A. (2017). Trait anger, physical aggression, and violent offending in antisocial and borderline personality disorders. *Journal of Forensic Sciences*, 62(1), 137-141. <https://doi.org/10.1111/1556-4029.13234>
- Koole, S. L. (2009). The psychology of emotion regulation: An integrative review. *Cognition and Emotion*, 23(1), 4-41. <https://doi.org/10.1080/02699930802619031>
- Krakowski, M. I., Foxe, J., de Sanctis, P., Nolan, K., Hoptman, M. J., Shope, C., . . . Czobor, P. (2015). Aberrant response inhibition and task switching in psychopathic individuals. *Psychiatry Research*, 229(3), 1017-1023. <https://doi.org/10.1016/j.psychres.2015.06.018>
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3), 213-236. <https://doi.org/10.1002/acp.2350050305>
- Lansbergen, M. M., Kenemans, J. L., & van Engeland, H. (2007). Stroop interference and attention-deficit/hyperactivity disorder: a review and meta-analysis. *Neuropsychology*, 21(2), 251-262. <https://doi.org/10.1037/0894-4105.21.2.251>
- Laws, B., & Crewe, B. (2015). Emotion regulation among male prisoners. *Theoretical Criminology*, 20(4), 529-547. <https://doi.org/10.1177/1362480615622532>
- Lehrl, S. (2005). *Manual zum MWT-B* (5th ed.). Balingen: Spitta-Verlag.
- Lenhard, W., & Lenhard, A. (2014). *Hypothesis tests for comparing correlations*. Retrieved from: <https://www.psychometrica.de/correlation.html>. Bibergau (Germany): Psychometrica. <https://doi.org/10.13140/RG.2.1.2954.1367>

- Lenhard, W., & Lenhard, A. (2016). *Calculation of effect sizes*. Retrieved from https://www.psychometrica.de/effect_size.html Dettelbach (Germany): Psychometrica. <https://doi.org/10.13140/RG.2.1.3478.4245>
- Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, 25(5), 331-348. [https://doi.org/10.1002/\(sici\)1098-2337\(1999\)25:5<331::Aid-ab2>3.0.Co;2-1](https://doi.org/10.1002/(sici)1098-2337(1999)25:5<331::Aid-ab2>3.0.Co;2-1)
- Linehan, M. M. (2015). *DBT® skills training manual* (2nd ed.). New York, NY: Guilford Press.
- Lobbestael, J., Arntz, A., Cima, M., & Chakhssi, F. (2009). Effects of induced anger in patients with antisocial personality disorder. *Psychological Medicine*, 39(4), 557-568. <https://doi.org/10.1017/S0033291708005102>
- Lobbestael, J., Arntz, A., & Wiers, R. W. (2008). How to push someone's buttons: A comparison of four anger-induction methods. *Cognition and Emotion*, 22(2), 353-373. <https://doi.org/10.1080/02699930701438285>
- Low, K., & Day, A. (2015). Toward a clinically meaningful taxonomy of violent offenders: The role of anger and thinking styles. *Journal of Interpersonal Violence*. <https://doi.org/10.1177/0886260515586365>
- Matjašič, M., Vehovar, V., & Manfreda, K. L. (2018). Web survey paradata on response time outliers: A systematic literature review. *Advances in Methodology & Statistics / Metodoloski zvezki*, 15(1), 23-41. Retrieved from <http://www.redi-bw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d131141707%26site%3dehost-live>
- McCarthy, R. J., & Elson, M. (2018). A conceptual review of lab-based aggression paradigms. *Collabra: Psychology*, 4(1), 1-12. <https://doi.org/10.1525/collabra.104>
- McMurrin, M. (2011). Emotions and antisocial behaviour: An introduction to the special issue. *Journal of Forensic Psychiatry & Psychology*, 22(5), 629-634. <https://doi.org/10.1080/14789949.2011.617533>
- McRae, K. (2013). Emotion regulation frequency and success: Separating constructs from methods and time scale. *Social and Personality Psychology Compass*, 7(5), 289-302. <https://doi.org/10.1111/spc3.12027>
- McRae, K., & Gross, J. J. (2020). Emotion regulation. *Emotion*, 20(1), 1-9. <https://doi.org/10.1037/emo0000703>
- McRae, K., Jacobs, S. E., Ray, R. D., John, O. P., & Gross, J. J. (2012). Individual differences in reappraisal ability: Links to reappraisal frequency, well-being, and cognitive control. *Journal of Research in Personality*, 46, 2-7. <https://doi.org/10.1016/j.jrp.2011.10.003>

- Miller, D. J., Vachon, D. D., & Aalsma, M. C. (2012). Negative affect and emotion dysregulation: Conditional relations with violence and risky sexual behavior in a sample of justice-involved adolescents. *Criminal Justice and Behavior*, *39*, 1316-1327. <https://doi.org/10.1177/0093854812448784>
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8-14. <https://doi.org/10.1177/0963721411429458>
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*(4), 674-701. <https://doi.org/10.1037/0033-295X.100.4.674>
- Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, *13*(02), 355-375. <https://doi.org/10.1017/S0954579401002097>
- Mokros, A., Hollerbach, P., Nitschke, J., & Habermeyer, E. (2017). *PCL-R: Hare Psychopathy Checklist – Revised*. Göttingen: Hogrefe.
- Moran, P. (1999). The epidemiology of antisocial personality disorder. *Social Psychiatry & Psychiatric Epidemiology*, *34*(5), 231. <https://doi.org/10.1007/s001270050138>
- 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](https://doi.org/10.1016/S0272-7358(98)00096-8)
- Mueller, S. (2011). The influence of emotion on cognitive control: relevance for development and adolescent psychopathology. *Frontiers in Psychology*, *2*. <https://doi.org/10.3389/fpsyg.2011.00327>
- Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, *15*, 1044-1045. <https://doi.org/10.1093/beheco/arh107>
- 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>
- Novaco, R. W. (2011). Anger dysregulation: driver of violent offending. *Journal of Forensic Psychiatry & Psychology*, *22*(5), 650-668. <https://doi.org/10.1080/14789949.2011.617536>

- Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242-249. <https://doi.org/10.1016/j.tics.2005.03.010>
- 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>
- Ogloff, J. R. P. (2006). Psychopathy/antisocial personality disorder conundrum. *Australian and New Zealand Journal of Psychiatry*, 40, 519-528. <https://doi.org/10.1080/j.1440-1614.2006.01834.x>
- Osumi, T., Nakao, T., Kasuya, Y., Shinoda, J., Yamada, J., & Ohira, H. (2012). Amygdala dysfunction attenuates frustration-induced aggression in psychopathic individuals in a non-criminal population. *Journal of Affective Disorders*, 142(1), 331-338. <https://doi.org/10.1016/j.jad.2012.05.012>
- Parrott, D. J., & Giancola, P. R. (2007). Addressing “The criterion problem” in the assessment of aggressive behavior: Development of a new taxonomic system. *Aggression and Violent Behavior*, 12(3), 280-299. <https://doi.org/10.1016/j.avb.2006.08.002>
- 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, 1713-1713. <https://doi.org/10.3389/fpsyg.2018.01713>
- Poythress, N. G., Edens, J. F., Skeem, J. L., Lilienfeld, S. O., Douglas, K. S., Frick, P. J., . . . Wang, T. (2010). Identifying subtypes among offenders with antisocial personality disorder: A cluster-analytic study. *Journal of Abnormal Psychology*, 119(2), 389-400. <https://doi.org/10.1037/a0018611>
- Pruessner, L., Barnow, S., Holt, D. V., Joormann, J., & Schulze, K. (2020). A cognitive control framework for understanding emotion regulation flexibility. *Emotion*, 20(1), 21-29. <https://doi.org/10.1037/emo0000658>, <https://doi.org/10.1037/emo0000658.supp> (Supplemental)
- Reitan, R. M. (1992). *Trail Making Test. Manual for administration and scoring*. Tucson, Arizona: Reitan Neuropsychological Laboratory.
- Ritter, D., & Eslea, M. (2005). Hot Sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior*, 31, 407-419. <https://doi.org/10.1002/ab.20066>
- Robertson, T., Daffern, M., & Bucks, R. S. (2012). Emotion regulation and aggression. *Aggression and Violent Behavior*, 17(1), 72-82. <https://doi.org/10.1016/j.avb.2011.09.006>

- Robertson, T., Daffern, M., & Bucks, R. S. (2014). Maladaptive emotion regulation and aggression in adult offenders. *Psychology, Crime & Law*, 20(10), 933-954. <https://doi.org/10.1080/1068316X.2014.893333>
- Robertson, T., Daffern, M., & Bucks, R. S. (2015). Beyond anger control: Difficulty attending to emotions also predicts aggression in offenders. *Psychology of Violence*, 5(1), 74-83. <https://doi.org/10.1037/a0037214>
- Rosenthal, R., & Fode, K. L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science*, 8(3), 183. <https://doi.org/10.1002/bs.3830080302>
- Rösler, M., Retz-Junginger, P., Retz, W., & Stieglitz, R.-D. (2008). *HASE - Homburger ADHS-Skalen für Erwachsene*. Göttingen: Hogrefe.
- Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr, E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, 20(5), 343-350. <https://doi.org/10.1037/h0043220>
- Roszyk, A., Izdebska, A., & Peichert, K. (2013). Planning and inhibitory abilities in criminals with antisocial personality disorder. *Acta Neuropsychologica*, 11(2), 193-205. <https://doi.org/10.5604/17307503.1073475>
- Rottenberg, J., Ray, R. D., & Gross, J. J. (2007). Emotion elicitation using films. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 9-28). London: Oxford University Press.
- Rotter, M., Way, B., Steinbacher, M., Sawyer, D., & Smith, H. (2002). Personality disorders in prison: Aren't they all antisocial? *Psychiatric Quarterly*, 73(4), 337-349. <https://doi.org/10.1023/A:1020468117930>
- Russell, G. W., Arms, R. L., Loof, S. D., & Dwyer, R. S. (1996). Men's aggression toward women in a bungled procedure paradigm. *Journal of Social Behavior & Personality*, 11(4), 729-738. Retrieved from <https://www.redibw.de/db/ebsco.php/search.ebscohost.com/login.aspx%3fdirect%3dtrue%26db%3daph%26AN%3d9703272122%26site%3dehost-live>
- Sánchez-Cubillo, I., Periáñez, J. A., Adrover-Roig, D., Rodríguez-Sánchez, J. M., Ríos-Lago, M., Tirapu, J., & Barceló, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, 15, 438-450. <https://doi.org/10.1017/S1355617709090626>

- Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica*, 45(8), 1989-1996. <https://doi.org/10.2307/1914122>
- Sayar, K., Ebrinc, S., & Ak, I. (2001). Alexithymia in patients with antisocial personality disorder in a military hospital setting. *The Israel Journal of Psychiatry and Related Sciences*, 38(2), 81-87. Retrieved from <https://www.questia.com/library/journal/1P3-75340373/alexithymia-in-patients-with-antisocial-personality>
- Schiffer, B., Pawliczek, C., Müller, B., Forsting, M., Gizewski, E., Leygraf, N., & Hodgins, S. (2014). Neural mechanisms underlying cognitive control of men with lifelong antisocial behavior. *Psychiatry Research*, 222(1-2), 43-51. <https://doi.org/10.1016/j.psychresns.2014.01.008>
- Schreiner, E., Wolkenstein, L., & Joormann, J. (2020). *Can't stop me now – cognitive-emotional impairments in euthymic bipolar disorder*. Manuscript in preparation.
- Sedgwick, O., Young, S., Baumeister, D., Greer, B., Das, M., & Kumari, V. (2017). Neuropsychology and emotion processing in violent individuals with antisocial personality disorder or schizophrenia: The same or different? A systematic review and meta-analysis. *The Australian and New Zealand Journal of Psychiatry*, 51(12), 1178-1197. <https://doi.org/10.1177/0004867417731525>
- Seruca, T., & Silva, C. F. (2015). Recidivist criminal behaviour and executive functions: A comparative study. *Journal of Forensic Psychiatry & Psychology*, 26(5), 699-717. <https://doi.org/10.1080/14789949.2015.1054856>
- Seruca, T., & Silva, C. F. (2016). Executive functioning in criminal behavior: Differentiating between types of crime and exploring the relation between shifting, inhibition, and anger. *International Journal of Forensic Mental Health*, 15(3), 235-246. <https://doi.org/10.1080/14999013.2016.1158755>
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., . . . Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59(Supplement 20), 22-33. Retrieved from <https://www.psychiatrist.com/JCP/article/Pages/1998/v59s20/v59s2005.aspx>
- Shepherd, S. M., Campbell, R. E., & Ogloff, J. R. P. (2016). Psychopathy, antisocial personality disorder, and reconviction in an Australian sample of forensic patients. *International Journal of Offender Therapy and Comparative Criminology*, 62(3), 609-628. <https://doi.org/10.1177/0306624X16653193>

- Skeem, J. L., Schubert, C., Odgers, C., Mulvey, E. P., Gardner, W., & Lidz, C. (2006). Psychiatric symptoms and community violence among high-risk patients: A test of the relationship at the weekly level. *Journal of Consulting and Clinical Psychology, 74*(5), 967-979. <https://doi.org/10.1037/0022-006X.74.5.967>
- Spielberger, C. D. (1999). *State-Trait Anger Expression Inventory-2 (STAXI-2)*. Professional Manual. Tampa, FL: Psychological Assessment Resources.
- Spinella, M. (2007). Normative data and a short form of the Barratt Impulsiveness Scale. *International Journal of Neuroscience, 117*(3), 359-368. <https://doi.org/10.1080/00207450600588881>
- Statistisches Bundesamt (2014). *Rechnungsergebnisse der öffentlichen Haushalte – Fachserie 14, Reihe 3.1 – 2011*. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt (2019a). *Bestand der Gefangenen und Verwahrten in den deutschen Justizvollzugsanstalten nach ihrer Unterbringung auf Haftplätzen des geschlossenen und offenen Vollzuges*. Wiesbaden: Statistisches Bundesamt (Destatis).
- Statistisches Bundesamt (2019b) *Strafverfolgung – Fachserie 10, Reihe 3 – 2018*. Wiesbaden: Statistisches Bundesamt (Destatis).
- Steffgen, G., & Dusi, D. (2006). Ärgerbewältigungstraining. In F. J. Schermer & A. Weber (Eds.), *Methoden der Verhaltensänderung: Komplexe Interventionsprogramme* (pp. 37-64). Stuttgart: Kohlhammer.
- Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17). *European Journal of Psychological Assessment, 17*(3), 222-232. <https://doi.org/10.1027//1015-5759.17.3.222>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. <https://doi.org/10.1037/h0054651>
- Sukhodolsky, D. G., Golub, A., & Cromwell, E. N. (2001). Development and validation of the Anger Rumination Scale. *Personality and Individual Differences, 31*(5), 689-700. [https://doi.org/10.1016/S0191-8869\(00\)00171-9](https://doi.org/10.1016/S0191-8869(00)00171-9)
- Sullivan, R. M., Perlman, G., & Moeller, S. J. (2019). Meta-analysis of aberrant post-error slowing in substance use disorder: implications for behavioral adaptation and self-control. *European Journal of Neuroscience, 50*(3), 2467-2476. <https://doi.org/10.1111/ejn.14229>
- Sullivan, T. N., Helms, S. W., Kliewer, W., & Goodman, K. L. (2010). Associations between sadness and anger regulation coping, emotional expression, and physical and relational aggression among urban adolescents. *Social Development, 19*(1), 30-51. <https://doi.org/10.1111/j.1467-9507.2008.00531.x>

- Tager, D., Good, G. E., & Brammer, S. (2010). "Walking over 'em": An exploration of relations between emotion dysregulation, masculine norms, and intimate partner abuse in a clinical sample of men. *Psychology of Men & Masculinity*, *11*(3), 233-239. <https://doi.org/10.1037/a0017636>
- Tang, D., & Schmeichel, B. J. (2014). Stopping anger and anxiety: Evidence that inhibitory ability predicts negative emotional responding. *Cognition and Emotion*, *28*(1), 132-142. <https://doi.org/10.1080/02699931.2013.799459>
- Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, *35*, 297-310. <https://doi.org/10.1111/j.1467-6494.1967.tb01430.x>
- Tedeschi, J. T., & Quigley, B. M. (1996). Limitations of laboratory paradigms for studying aggression. *Aggression and Violent Behavior*, *1*(2), 163-177. [https://doi.org/10.1016/1359-1789\(95\)00014-3](https://doi.org/10.1016/1359-1789(95)00014-3)
- Timmermann, M., Jeung, H., Schmitt, R., Boll, S., Freitag, C. M., Bertsch, K., & Herpertz, S. C. (2017). Oxytocin improves facial emotion recognition in young adults with antisocial personality disorder. *Psychoneuroendocrinology*, *85*, 158-164. <https://doi.org/10.1016/j.psyneuen.2017.07.483>
- Tombaugh, T. N. (2004). Trail Making Test A and B: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*(2), 203-214. [https://doi.org/10.1016/S0887-6177\(03\)00039-8](https://doi.org/10.1016/S0887-6177(03)00039-8)
- Tomlinson, M. F. (2018). A theoretical and empirical review of Dialectical Behavior Therapy within forensic psychiatric and correctional settings worldwide. *International Journal of Forensic Mental Health*, *17*(1), 72-95. <https://doi.org/10.1080/14999013.2017.1416003>
- Tonnaer, F., Cima, M., & Arntz, A. (2019). Explosive matters: Does venting anger reduce or increase aggression? Differences in anger venting effects in violent offenders. *Journal of Aggression, Maltreatment & Trauma*, 1-17. <https://doi.org/10.1080/10926771.2019.1575303>
- Tonnaer, F., Siep, N., van Zutphen, L., Arntz, A., & Cima, M. (2017). Anger provocation in violent offenders leads to emotion dysregulation. *Scientific Reports*, *7*(1), 3583. <https://doi.org/10.1038/s41598-017-03870-y>
- Vazsonyi, A. T., Mikuška, J., & Kelley, E. L. (2017). It's time: A meta-analysis on the self-control-deviance link. *Journal of Criminal Justice*, *48*, 48-63. <https://doi.org/10.1016/j.jcrimjus.2016.10.001>

- Velotti, P., Garofalo, C., Callea, A., Bucks, R. S., Robertson, T., & Daffern, M. (2017). Exploring anger among offenders: The role of emotion dysregulation and alexithymia. *Psychiatry, Psychology and Law*, 24(1), 128-138. <https://doi.org/10.1080/13218719.2016.1164639>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070. <https://doi.org/10.1037//0022-3514.54.6.1063>
- Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, 138(4), 775-808. <https://doi.org/10.1037/a0027600>
- Wilkowski, B. M., Crowe, S. E., & Ferguson, E. L. (2015). Learning to keep your cool: Reducing aggression through the experimental modification of cognitive control. *Cognition and Emotion*, 29(2), 251-265. <https://doi.org/10.1080/02699931.2014.911146>
- Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, 38(1), 174-180. <https://doi.org/10.3758/BF03192765>
- Williams, K. D., Yeager, D. S., Cheung, C. K. T., & Choi, W. (2012). Cyberball 4.0 [Software].
- Yavuz, K. F., Şahin, O., Ulusoy, S., İpek, O. U., & Kurt, E. (2016). Experiential avoidance, empathy, and anger-related attitudes in antisocial personality disorder. *Turkish Journal Of Medical Sciences*, 46(6), 1792-1800. <https://doi.org/10.3906/sag-1601-80>
- Yoon, J., & Knight, R. A. (2015). Emotional processing of individuals high in psychopathic traits. *Australian Journal of Psychology*, 67(1), 29-37. <https://doi.org/10.1111/ajpy.12063>
- Zajenkowski, M., & Zajenkowska, A. (2015). Intelligence and aggression: The role of cognitive control and test related stress. *Personality and Individual Differences*, 81, 23-28. <https://doi.org/10.1016/j.paid.2014.12.062>
- Zeier, J. D., Baskin-Sommers, A. R., Hiatt Racer, K. D., & Newman, J. P. (2012). Cognitive control deficits associated with antisocial personality disorder and psychopathy. *Personality Disorders: Theory, Research, and Treatment*, 3(3), 283-293. <https://doi.org/10.1037/a0023137>
- Zillmann, D., & Bryant, J. (1974). Effect of residual excitation on the emotional response to provocation and delayed aggressive behavior. *Journal of Personality and Social Psychology*, 30(6), 782-791. <https://doi.org/10.1037/h0037541>

7. Appendix

7.1. Appendix A

Appendix A. Complete Cyberball chat conversation (round 1 – 12), including corresponding condition (baseline vs. anger) and pass rate (participant : other player)

Round no.	Condition	Sender	Chat comment	Pass rate
1	Baseline	Player 1	hey guys, im marc	1 : 1 (50%)
		Player 3	Hi	
		Player 3	Andi here	
		Player 1	whats your name player 2?	
2	Baseline	Player 3	Are you also sitting in one of these musty rooms?	1 : 1 (50%)
		Player 1	yeah	
		Player 1	im crowed in fucking breeding cage...	
		Player 1	what about you, olayer 2?	
3	Baseline	Player 3	Man, I hope the experiment goes fast	1 : 1 (50%)
		Player 1	ya, really got better things to do...	
		Player 3	How about you, player 2?	
4	Baseline	Player 3	Hey player 2	1 : 1 (50%)
		Player 3	What's up?	
		Player 1	whats your name?	
		Player 1	write something!	
5	Anger	Player 3	Player 2??	1 : 2 (33%)
		Player 3	Why don't you wirte?????	
		Player 1	what the fuck is this?	
		Player 3	Think you're really phat, right?	
		Player 1	such an ass!	
6	Anger	Player 3	...?!	1 : 2 (33%)
		Player 1	think you can ignore us, you arrogant wannabe?	
		Player 3	moron..	
		Player 1	such an idiot...	
7	Anger	Player 3	Wow Player2	1 : 2 (33%)
		Player 3	You throw like a chick...	
		Player 1	hehe ☺	
		Player 1	gross^^...	
		Player 3	You must be such a wimp!	

(Continued)

Appendix A. Complete Cyberball chat conversation (round 1 – 12), including corresponding condition (baseline vs. anger) and pass rate (participant : other player) (continued)

Round no.	Condition	Sender	Chat comment	Pass rate
8	Anger	Player 1	ey, the guy gets more bucks the more we give him the ball!	2× (13.3%)
		Player 1	just got info	
		Player 3	Eh?	
		Player 1	just dont give him the bal lanynmore!!!!	
		Player 1	he will get less money, ok?!	
		Player 3	That sucks	
		Player 3	Ok	
9	Anger	Player 1	remember, dont give him	0 : 1 (0%)
		Player 1	the ball!	
		Player 3	Boah, not up for this guy anymore...	
		Player 1	think im keen for the jerk?	
10	Anger	Player 1	shit man, the moron is still here	1× (6.7%)
		Player 3	I was hoping he finally got lost...	
		Player 1	hey! not him!	
		Player 3	Sorry	
11	Anger	Player 3	Come on, player 2, just beat it and don't bother us any more, ok!??	0 : 1 (0%)
		Player 1	no shit	
		Player 1	wont get a ball anyway	
		Player 3	Right! Such an ass...	
		Player 1	for real: you do know youre a pain int he ass, dont ya?	
12	Anger	Player 3	..nicely (!!) put..	0 : 1 (0%)
		Player 1	just beat it!	

Note. All chat comments were translated from German to English for this manuscript. The original chat conversation is available upon request. Typos are intentional to enhance credibility.

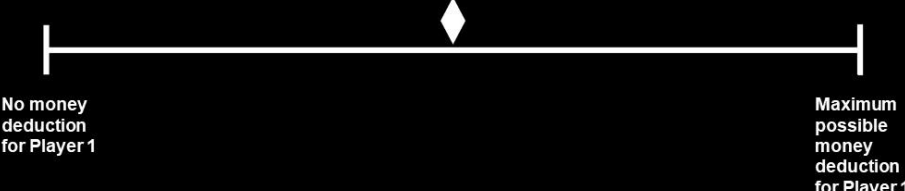
7.2. Appendix B

a) No consequence

Shall Player 1 lose a part of his payment?

Decide for or against a money deduction for Player 1 by adjusting the slider with the mouse and then clicking accordingly.

Your decision has no consequence regarding your own reward.



No money deduction for Player 1

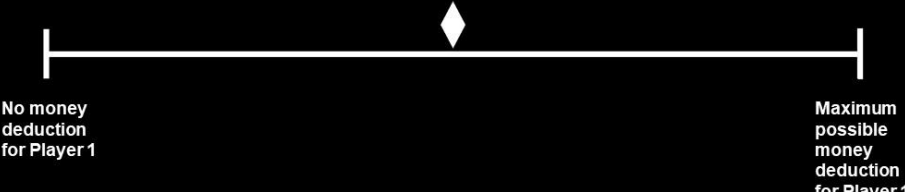
Maximum possible money deduction for Player 1

b) Negative consequence

Shall Player 1 lose a part of his payment?

Decide for or against a money deduction for Player 1 by adjusting the slider with the mouse and then clicking accordingly.

Caution: If you deduce money from Player 1, there is the risk of losing a proportion of your own reward. The more money you deduce, the more likely it is that you will suffer a reduction in financial compensation yourself.



No money deduction for Player 1

Maximum possible money deduction for Player 1

Appendix B. Screenshots of the punishment slide within the Cyberball Aggression Task depending on consequence (no consequence, negative consequence). The slides were translated from German to English for this manuscript. The starting position of the slider is in the middle of the visual analogue scale (50 of 100). When the slider has been adjusted, the participant is asked to confirm his decision by pressing on an (then appearing) “accept” button. The decision can be revised as often as desired.

7.3. Appendix C

Credibility Coding

Credibility was dichotomized into “deceived” and “not deceived”. When coded as “deceived”, participants had no idea up to the last round of Cyberball that (a) the players were faked, nor did they believe (b) the other players were instructed to insult them. Furthermore, they did not express an understanding of (c) the money deduction not being real. Hence, only diffuse doubts about the CAT were allowed. Accordingly, a participant who was assigned “not deceived” either had an understanding of (a) the faked players, thought (b) the players were asked to provoke him or expressed any idea that doubted (c) the announced money deduction. The coding was based on the below mentioned interview, while the investigators’ impression was the decisive component. This is important to note, as some participants were convincingly annoyed during the game (as evident by their behavior during the CAT, e.g. calling the instructor, soliloquizing, shouting swear words, writing insulting chat comments on their own or similar behaviors), but later, during the interview, they behaved as if they had looked through the cover story from the very beginning. Obviously, they did not see through the deception until the last round of Cyberball but came to an understanding during the ER strategy assessment or the interview. They however tried to hide this understanding. In this case, participants were nonetheless classified as „deceived“.

Credibility Assessment

The CAT was followed by an unstandardized interview. The investigator tried to subtly assess credibility of the paradigm, without overtly revealing the deception. This was ensured by asking predominantly open questions while still sticking to the cover story, unless the participant clearly got the deception figured out. No leading questions were asked. Notes about participants’ answers and investigators’ impressions were written down after the interview to enable discussions of ratings between investigators following the sessions.

After the participant gave the signal to be finished with the CAT, the investigator terminated the computer program and asked – quite casually – what it had been like. This served as a starting point to introduce the following, more specific, questions. In several cases, there were also introductory questions regarding participants’ comments and his behavior during the CAT (e.g. “earlier you said that...”, “you mentioned that they insulted you?”). Depending on participants’ statements, further, initially quite general, questions were asked, always with the goal of finding out, if and what the participant had seen through during the CAT. In order to maintain the cover story, the investigator tried to appear authentic and to take a naïve but

empathic role while asking (e.g. showing that she is sorry or ashamed for the other players' treatment of him). Reassurances (e.g. "what do you mean by that?") and summarizing (e.g. "you think...?") were frequently made to ensure the participant was understood correctly and to invite him to add further feedback. With increasing duration, the questions became more concrete, while still not revealing the cover story (e.g. "why did they insult you?", "how did you react?", "(why) did you deduct money from player 1?"). Finally, participants were directly asked about their beliefs (e.g. "what do you think the task was about?") and at what point in time they might have seen through the deception (e.g. "when did you notice, that something was weird?", "Did I understand you correctly: You began to wonder that something was weird when nobody passed you the ball anymore. And then, after those questions about your emotions appeared, you figured out the game was pre-programmed?"). The interview was followed by an oral and written debriefing (see chapter 3.2.3).

7.4. Appendix D



Fachbereich Psychologie Schleichstr. 4 · 72076 Tübingen

**Mathematisch-
Naturwissenschaftliche
Fakultät**

Universität Tübingen
Fachbereich Psychologie
Arbeitsbereich Klinische Psychologie

Ansprechpartner für Rückfragen:
Dipl.-Psych. Elena Schreiner
Telefax +49 7071 29-5219

Aufklärungstext

Universität Tübingen/LMU München

Titel der Studie: Kognitive Fertigkeiten bei verschiedenen Personengruppen

Lieber Teilnehmer,

im Folgenden möchten wir Ihnen einige Informationen über die vorangegangene Aufgabe und den tatsächlichen Zweck der Untersuchung zukommen lassen.

Entgegen unserer Ankündigung geht es in Cyberball nicht darum, die Fähigkeit der mentalen Visualisierung einzuüben. Stattdessen wollen wir untersuchen, ob Sie durch die Äußerungen von Spieler 1 und Spieler 3 einen Anstieg ärgerlicher Gefühle erfuhren. Zudem interessiert uns, ob sich die Provokationen der Mitspieler in Ihrem Bestrafungsverhalten bemerkbar machten. In den Fällen, in denen eine Bestrafung an eine mögliche negative Konsequenz gekoppelt war (die Wahrscheinlichkeit, mit der Sie eine eigene Entlohnungsminderung befürchten mussten), brauchte man für das Widerstehen dieser Bestrafungstendenz die Fähigkeit der sogenannten kognitiven Kontrolle. In der vorliegenden Studie wollen wir untersuchen, ob sich Strafgefangene mit und ohne antisoziale Persönlichkeitsstörung in ihrer kognitiven Kontrollleistung von nicht-inhaftierten gesunden Kontrollprobanden unterscheiden. Um diese Fragestellung untersuchen zu können, war es jedoch erforderlich, Sie an einigen Stellen des Experiments falsch zu informieren: So handelte es sich bei Cyberball nicht um ein Online-Spiel, sondern um ein vorprogrammiertes Experiment. Die vermeintlichen Mitspieler

waren keine echten Probanden, sondern wurden vom Computer gesteuert. Ebenso waren die Chat-Kommentare, die Sie erhielten, vorprogrammiert. Die darin enthaltenen Äußerungen waren also nicht persönlich auf Sie bezogen. Wir haben im Rahmen einiger Vorstudien die Erfahrung gemacht, dass die präsentierten Kommentare imstande sind, ärgerliche Gefühle hervorzurufen. Genau das war Sinn und Zweck des Chats. Auch die Fehlermeldung beim Versenden Ihrer eigenen Nachrichten wurde absichtlich so programmiert. Möglicherweise wäre Ihnen sonst aufgefallen, dass Ihre vermeintlichen Mitspieler gar nicht auf die von Ihnen verfassten Beiträge eingehen. Vielleicht hätte Sie das stutzig und die Aufgabe somit weniger glaubwürdig gemacht. **Selbstverständlich erhalten Sie nach Beendigung der Studie Ihre vollständige Entlohnung.** Anders als während der Aufgabe angekündigt gibt es keine Abzüge aufgrund Ihres Bestrafungsverhaltens. Sie erhalten für Ihre Teilnahme an der Studie nach wie vor eine Vergütung von 8€ pro Stunde, bis zu einem Maximalbetrag von 25€.

Unser Ziel war es keinesfalls, Sie durch die anfänglichen Fehlinformationen in irgendeiner Weise auszutricksen oder gar vorzuführen. Wir wollten mit der genannten Täuschung vielmehr sicherstellen, dass Sie sich bei der vorangegangenen Aufgabe möglichst authentisch verhalten. Nur dann kann es gelingen, gültige Studienergebnisse zu erhalten. Wir glauben, dass diese Studie wichtig ist, weil sie uns die Chance bietet, besser zwischen verschiedenen Straftätergruppen zu unterscheiden. Langfristiges Ziel wäre es, die in der JVA angebotenen Behandlungsmaßnahmen entsprechend anzupassen und zu verbessern. Fänden sich Defizite in der kognitiven Kontrollleistung bei einer bestimmten Untergruppe von Straftätern, ließen sich daraus spezifische Trainings ableiten. Diesbezüglich konnten bei depressiven Patienten schon erste Erfolge erzielt werden.

Möglicherweise waren einige der oben genannten Aspekte für Sie irreführend. Dafür möchten wir uns entschuldigen. Wir hoffen auf Ihr Verständnis für die von uns angewandte Methode, waren wir doch nur so in der Lage, der uns interessierenden Fragestellung nachzugehen. Selbstverständlich werden alle von Ihnen erhobenen Daten weiterhin vertraulich und wie in den Teilnehmerinformationen beschrieben, behandelt. Wir sind nicht an individuellen Ergebnissen einzelner Studienteilnehmer interessiert, sondern schauen uns vielmehr die Antworten bestimmter Teilnehmergruppen an, und kombinieren hierzu deren Werte.

Wenn Sie sich in irgendeiner Art und Weise unwohl fühlen, bitten wir Sie, sich nun an den Versuchsleiter zu wenden. Gerne beantwortet dieser Ihnen alle weiteren Fragen.

Nochmals vielen Dank für Ihre Mithilfe!

7.5. Appendix E

Appendix E. Non-significant effects regarding aggressive behavior during the Cyberball Aggression Task

Effect	ANOVA F	p
Consequence	$F(1, 91) = 0.15$.702
Consequence \times Credibility	$F(1, 91) = 0.43$.513
Consequence \times Group	$F(2, 91) = 0.24$.790
Consequence \times Condition	$F(1, 91) = 2.25$.137
Credibility \times Group	$F(2, 91) = 0.39$.677
Consequence \times Credibility \times Group	$F(2, 91) = 2.23$.114
Consequence \times Credibility \times Condition	$F(1, 91) < 0.01$.996
Consequence \times Group \times Condition	$F(2, 91) = 0.74$.480
Credibility \times Group \times Condition	$F(2, 91) = 1.05$.356
Consequence \times Credibility \times Condition \times Group	$F(2, 91) = 0.76$.470

7.6. Appendix F

Appendix F. Non-significant effects regarding emotion regulation strategy use during and after the Cyberball Aggression Task

Dependent variable and effect	ANOVA F	p
Positive refocusing		
Credibility	$F(1, 91) = 1.73$.191
Group	$F(2, 91) = 0.12$.888
Credibility \times Group	$F(2, 91) = 0.73$.485
Putting into perspective		
Credibility	$F(1, 91) = 0.51$.479
Group	$F(2, 91) = 0.79$.455
Credibility \times Group	$F(2, 91) = 0.13$.883
Positive reappraisal		
Credibility	$F(1, 91) = 0.41$.522
Group	$F(2, 91) = 0.19$.829
Credibility \times Group	$F(2, 91) = 1.99$.143
Acceptance (of the situation)		
Credibility	$F(1, 91) = 0.51$.477
Group	$F(2, 91) = 1.37$.260
Credibility \times Group	$F(2, 91) = 1.36$.262
Experience suppression		
Credibility	$F(1, 91) = 0.12$.730
Group	$F(2, 91) = 0.49$.614
Credibility \times Group	$F(2, 91) = 2.09$.129
Understanding of causes		
Group	$F(2, 91) = 1.92$.152
Credibility \times Group	$F(2, 91) = 0.67$.515
Angry afterthoughts		
Group	$F(2, 91) = 1.17$.314
Credibility \times Group	$F(2, 91) = 1.07$.348