UNIVERSITÄT ZU LÜBECK

Understanding aggression episodes: novel experimental approaches

Dissertation

Submitted by

Macià Buades-Rotger

Lübeck 2017

UNIVERSITÄT ZU LÜBECK

From the Department of Neurology
of the University of Lübeck
Director: Prof. Dr. Thomas F. Münte

Understanding aggression episodes: novel experimental approaches

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Natural Sciences

Submitted by:

Macià Buades-Rotger
from Algaida (Spain)

Lübeck 2017

First referee: Prof. Dr. rer. nat Ulrike M. Krämer

Second referee: Prof. Dr. phil. So Young Park

Date of oral examination: 25/10/2017

Approved for printing. Lübeck, 2017

Table of contents

Acknowledgements **7**

Abstract **8**

Zusammenfassung **10**

Chapter 1: Introduction **12**

1.1. Theoretical background **12**

1.2. Models of aggression **14**

1.2.1. Early models 14

1.2.2. The General Aggression Model 15

1.2.3. Situational influences on aggression 17

1.2.4. Dispositional influences on aggression 18

1.3. Measuring aggression **21**

1.3.1. The Taylor Aggression Paradigm 21

1.3.2. The Point Subtraction Aggression Paradigm 21

1.3.3. The Buss Aggression Machine 22

1.3.4. The Ultimatum Game 23

1.3.5. The Preemptive Strike Game 23

1.3.6. The Voodoo Doll Task 24

1.3.7. The Hot Sauce Paradigm 24

1.3.8. Evaluation-based paradigms 25

1.3.9. Other laboratory measures 25

1.3.10. Self-reports 27

1.3.11. Observational and experience sampling methods 28

1.4. Evolutionary basis of aggressive behavior **29**

1.5. Brain structure and function involved in reactive aggression **31**

1.5.1. Neurobiological models of reactive aggression 31

1.5.2. Animal studies 32

1.5.3. Lesion studies in humans 34

1.5.4. Stimulation studies in humans 36

1.5.5. Gray matter structure and aggression 38

1.5.6. White matter structure and aggression 40

1.5.7. Electroencephalographic studies 41

1.5.8. Positron emission tomography studies 42

1.5.9. Functional magnetic resonance imaging studies 44

1.5.10. Genetics of aggression 48

1.5.11. Serotonin and reactive aggression 49

1.5.12. Dopamine and reactive aggression 51

1.5.13. Hormonal influences on aggression 52

1.5.14. Hormonal influences on aggression-related neural structures 55

1.6. Outline of the thesis **56**

Chapter 2: Endogenous testosterone is associated with lower amygdala reactivity to angry faces and reduced aggressive behavior in healthy young women **56**

2.1. Introduction **58**

2.2. Methods **62**

2.2.1. Participants 62

2.2.2. Procedure 63

2.2.3. Saliva and blood collection 64

2.2.4. Social Threat Aggression Paradigm (STAP) 65

2.2.5. Neuroimaging data acquisition 66

2.2.6. Behavioral data analysis 66

2.2.7. Questionnaire data analysis 67

2.2.8. Hormone assays 67

2.2.9. Hormone data analysis 68

2.2.10. Neuroimaging data analysis 69

2.2.11. Mediation analyses 71

2.3. Results **71**

2.3.1. Behavioral results 71

2.3.2. Hormonal results 72

2.3.3. Hormone-behavior relationships 74

2.3.4. Neuroimaging results: task effects 75

2.3.5. Neuroimaging results: brain-behavior relationships 77

2.3.6. Neuroimaging results: brain-hormone relationships 78

2.3.7. Mediation results 78

2.4. Discussion **80**

2.4.1. Aggressive behavior and hormones 81

2.4.2. Neural reactivity to anger expressions and aggressive behavior 82

2.4.3. Hormonal effects on neural reactivity to anger expressions 85

2.4.4. Limitations 86

2.5. Conclusion **87**

Chapter 3: Introducing the Fight-or-Escape (FOE) paradigm **88**

3.1. Introduction **88**

Experiment 1 **91**

3.2. Methods **91**

3.2.1. Participants and procedure 91

3.2.2. Fight-or-Escape (FOE) paradigm 92

3.2.3. Behavioral measures 94

3.2.4. Personality questionnaires 95

3.2.5. Measurement of Fear Potentiation 95

3.2.6. EMG measurement and analysis 96
3.2.7. Correlational analyses 96
3.3. Results 97
3.4. Discussion 98
Experiment 2 99
3.5. Introduction 99
3.6. Methods 101
3.6.1. Participants and procedure 101
3.6.2. FOE-paradigm 101
3.6.3. Approach-Avoidance-Task 101
3.6.4. Dot-Probe-Task 102
3.6.5. Correlational analyses 103
3.7. Results 104
3.8. Discussion 105
3.9. General discussion 107
3.10. Limitations and future directions 108
Chapter 4: Avoidant responses to interpersonal provocation are associated with increased amygdala and decreased mentalizing network activity 110
4.1. Introduction 110
4.2. Materials and methods 112
4.2.1. Participants 112
4.2.2. Procedure 113
4.2.3. The Fight-or-Escape (FOE) Paradigm 113
4.2.4. Computerized and self-report measures 115
4.2.5. Heart rate data acquisition 118
4.2.6. Neuroimaging data acquisition 119
4.2.7. Behavioral data analysis 119
4.2.8. Heart rate data analysis 120
4.2.9. Neuroimaging data analysis 121
4.3. Results 124
4.3.1. Behavioral and self-report results 124
4.3.2. Heart rate results 126
4.3.3. Neuroimaging results 127
4.4. Discussion 131
4.4.1. Behavioral and cardiac responses to provocation 134
4.4.2. Neural responses to provocation 136
4.4.3. Neural activation underlying fight decisions 137
4.4.5. Neural activation underlying avoid decisions 138
4.4.6. Brain activity in the outcome phase 141

4.4.7. Limitations 141
4.5. Conclusions 142
Chapter 5: Ventral striatum connectivity during physical aggression 143
5.1. Introduction 143
5.2. Materials and methods 146
5.2.1. Participants 146
5.2.2. Procedure 146
5.2.3. Task design 146
5.2.4. Data acquisition 147
5.2.5. Data analysis 148
5.3. Results 150
5.4. Discussion 152
5.5. Conclusion 157
Chapter 6: General discussion 158
6.1. Summary of findings 158
6.2. Strengths and limitations 160
6.3. Implications for neurobiological models of aggression 162
6.4. Conceptual implications 165
6.5. Future directions 168
References 171

Acknowledgements

First of all I would like to thank everyone at the Department of Neurology for allowing me to conduct my studies in such a supportive atmosphere. Special thanks go to Thomas Münte for building the bridge between the Baltic and the Mediterranean through which I ended up in Lübeck. On the Mediterranean side, I would like to extend my gratitude to the people at the Faculty of Psychology at Mundet, especially David Gallardo-Pujol, for getting me started in this business while making a couple dear friends in the process.

It gets cheesier. Heartfelt thanks to Ulrike for her awesome advice, but mostly for believing in me. To Frederike, for her patience and help and zombies. To all people involved in the data collection: Marie Claes, Christin Engelke, Matthias Liebrand, Christian Erdmann, Susanne Schellbach, and to the confederates and participants. To all my colleagues over the years -including those who came and went- for stoically enduring my lame jokes. They have of course also provided helpful feedback and much needed distractions for which I am also grateful.

Gràcies als meus pares, Antònia i Pep, per ser i deixar-me ser. A les padrines i padrins, al Conco, a la tia Eva, i a na Llucia per tot l'amor que m'han donat i que espero have retornat, tot sense oblidar la resta de la família d'Algaida i Llucmajor. A na Roser, per ser un ravenet. Als meus amics, per fer-me companyia d'aprop i d'enfora en aquesta aventura. Ja que no puc anomenar a tothom, em limitaré a esmentar algunes comarques on habiten: Migjorn, Pla, Palma, Raiguer, el Bages, el Vallès, i el Barcelonès. Vull agrair també la seva feina als meus mestres i professors del Liceu, el Ramon Llull, i l'Autònoma.

Zu guter Letzt herzlichen Dank an meine Freunde des Roten Sterns Lübeck für den Spaß. Ich hoffe, dass meine Torvorlagen die falschen Einwürfe ausgleichen. Forza RSL!

Abstract

Reactive aggression can be defined as the intent to harm another organism in response to threat or provocation. In the present thesis, we aimed to characterize the neural and cognitive underpinnings of reactive aggressive behavior. To that end, we set participants to compete in a series of reaction time tasks in which the winner can punish the loser with sound blasts. In each study, we manipulated certain aspects of the task to answer specifics questions about reactive aggression.

In chapter 1, I introduce basic concepts and theoretical models of aggression, and review the evidence that leads to the present work. The focus is on studies relating brain structure and function to aggressive behavior, going from animal and lesion studies to neuroimaging.

In chapter 2, we investigated whether the hormones testosterone and cortisol modulate neural reactivity to angry faces and associated aggression in women (n=39). We observed that women with high testosterone were less aggressive and showed lower reactivity to angry faces in the amygdala, a brain region thought to be crucial for the experience of aggressive impulses.

In chapter 3, we developed and validated an experimental paradigm to measure both avoidant and aggressive reactions to provocation. In a first study (n=27), participants with higher emotional reactivity to threat were more avoidant and less aggressive. In a second study (n=34, 13 men), those with a stronger approach motivation toward positive stimuli showed reduced avoidant behavior when provoked.

In chapter 4, we used the paradigm validated in chapter 3 to examine the neural basis of aggressive and avoidant responses to provocation in a sample of healthy young women (n=36). We observed that increased amygdala activity and decreased mentalizing network

engagement were related to avoidance in response to high provocation. On the other hand, the orbitofrontal cortex (OFC) and motor regions were associated with aggressive decisions.

In chapter 5, we explored functional connectivity of the ventral striatum (VS), a brain structure that codes for human rewards, in a sample of healthy men (n=15). Specifically we inspected whether the VS shows differential connectivity with other brain areas when punishing a provoking opponent (active trials) as opposed to avoiding punishment oneself (passive trials). Results showed that coupling between the VS and other areas associated with reward valuation, arousal, attention, and motor control was increased in active compared to passive trials. Furthermore, interindividual differences in connectivity strength were correlated with aggression, reaction times, and trait anger.

Finally, in chapter 6 I discuss how the methods we employed can further the study of reactive aggression in experimental settings, and how our results contribute to the understanding of this phenomenon.

Zusammenfassung

Reaktive Aggression kann definiert werden als die Intention, einer anderen Person in Erwiderung einer Bedrohung oder Provokation zu schaden. Das Ziel der vorliegenden Arbeit war, die neuralen und kognitiven Grundlagen reaktiver Aggression zu charakterisieren. Zu diesem Zweck haben wir Proband*innen in einer Serie kompetitiver Reaktionszeitaufgaben, in denen der Gewinner den Verlierer mit einem unangenehmen Geräusch bestrafen darf, interagieren lassen. In jeder Studie haben wir bestimmte Aspekte des Paradigmas manipuliert, um spezifische Forschungsfragen über reaktive Aggression zu beantworten.

Im Kapitel 1 stelle ich grundlegende Konzepte und theoretische Modelle, die zur vorliegenden Arbeit führen, vor. Der Schwerpunkt liegt auf Tier-, Läsions- und Bildgebungsstudien, die Zusammenhänge zwischen Hirnstruktur, Hirnfunktion, und Aggression untersucht haben.

Im Kapitel 2 untersuchten wir, ob die Steroidhormone Testosteron und Kortisol neurale Reaktivität auf ärgerliche Gesichtsausdrücke und die damit assoziierte Aggression bei Frauen (n=39) beeinflussen können. Wir stellten fest, dass Frauen mit höherem Testosteron weniger aggressiv waren und geringere Reaktivität auf ärgerliche Gesichter in der Amygdala zeigten, einer wichtigen Hirnregion für die Verarbeitung aggressiver Impulse.

Im Kapitel 3 entwickelten und validierten wir ein experimentelles Paradigma zur Messung von sowohl vermeidenden als auch aggressiven Reaktionen auf Provokation. In einer ersten Studie zeigten sich Probandinnen (n=27, alle weiblich) mit höherer emotioneller Reaktivität als stärker vermeidend und weniger aggressiv. In einer zweiten Studie reagierten Versuchspersonen (n=34, 13 Männer), die eine höhere Annäherungsmotivation auf positive Reize aufwiesen, aggressiverer auf Provokation.

Im Kapitel 4 untersuchten wir die neuronale Grundlagen aggressiver und vermeidender Antworten auf Provokation mit dem im Kapitel 3 validierten Paradigma in einer Stichprobe von gesunden jungen Frauen (n=36). Stärkere Aktivität in der Amygdala und reduzierte Aktivität von „Theory of Mind“ Hirnarealen war mit Vermeidungsverhalten in Reaktion auf hohe Provokation verbunden. Der orbitofrontale Cortex und motorische Regionen waren anderseits mit aggressiven Entscheidungen assoziiert.

Im Kapitel 5 untersuchten wir die funktionelle Konnektivität des ventralen Striatums (VS), einer für die Belohnungsverarbeitung relevanten Hirnregion, in einer Stichprobe von gesunden Männern (n=15). Wir überprüften, ob Konnektivität zwischen dem VS und anderen Hirnregionen variiert, wenn ein provozierender Gegner bestraft wird (aktive Bedingung) oder wenn eine Bestrafung nur selbst vermieden wird (passive Bedingung). Die Ergebnisse zeigten, dass Konnektivität zwischen dem VS und Hirnregionen, die mit Belohnung, Erregung, Aufmerksamkeit, und motorischer Kontrolle assoziiert sind, in aktiven im Vergleich zu passiven Durchgängen erhöht wurde. Interindividuelle Unterschiede in der Kopplungsstärke waren überdies mit Aggression, Reaktionszeiten, und Neigung zum Ärger korreliert.

Im Kapitel 6 diskutiere ich, wie die vorgestellten Methoden helfen können, Aggression experimentell zu erforschen, und wie unsere Ergebnisse dazu beitragen, dieses Phänomen besser zu verstehen.

Chapter 1: Introduction

Now, please, don't put up a fight. It releases unpleasant chemicals into the bloodstream.

- Terry Pratchett

1.1. Theoretical background

Human aggression is a rudimentary behavior that can occur in practically any social situation, be it on a football field (R. Williams, 2009), at a university (Pandey, 2017; Wilkowski & Robinson, 2010), or even in the European parliament (Elgot, Halliday, & Rankin, 2016). While aggression can be adaptively expressed in certain contexts such as contact sports, uncontrolled aggression outbursts can have greatly detrimental consequences. Aggression can severely damage one's reputation, which may eventually result in job loss or social rejection (Rachlin, 2004; Tavris, 1984). In serious cases, it can even lead to prosecution and imprisonment, which can cause further problems to reintegrate in society (Edwards & Mottarella, 2014; Ward, Krohn, & Gibson, 2013). To that one must add the long-term consequences of aggression in terms of physical or psychological disability, plus less tangible effects such as detriments in quality of life and general well-being (B. X. Lee, 2016). In the United States, it has been estimated that the monetary costs of interpersonal violence can be as high as 3.3% of the country's gross domestic product (Waters, Hyder, Rajkotia, Basu, & Butchart, 2005). Aggression thus takes an enormous personal and societal toll. Investigating its distal and proximal causes is hence central for the betterment of society, and to understand human nature more generally.

Whether a given behavior will be judged as aggressive depends on the attribution of personal responsibility and on the perception of whom the aggressor and the victim are (e.g. police vs rioters; Bandura, 1978). Moreover, when someone is described as being aggressive,

it can refer to many different aspects of behavior, such as assertiveness or tenacity (R. A. Baron & Richardson, 1994). Hence, a first challenge to study aggression is to find a definition that is relatively impervious to subjectivity. To this end, we adopt the following operational definition of aggression: "any form of behavior directed towards the goal of harming or injuring another living being who is motivated to avoid such treatment" (R. A. Baron & Richardson, 1994). Note that, for this conceptualization of aggression, it is irrelevant whether the perpetrator succeeds in inflicting harm to the target. Rather, this definition places highest importance on an identifiable malevolent intent.

Some scholars have distinguished between *proactive* and *reactive* forms of aggression (Raine et al., 2006). Proactive or instrumental aggression is "cold-blooded" and goal-directed, i.e., it is performed in order to achieve a superordinate objective. Reactive or impulsive aggression, on the other hand, is "hot-blooded" and comprises aggressive acts executed in response to threat or provocation. Therefore, the primary goal of reactive aggression is to harm the target, whereas proactive aggression is performed to attain some other objective. Clinically, proactive aggression is typical of psychopathic individuals, whereas reactive aggression is most commonly observed in persons with antisocial personality disorder (APD) or intermittent explosive disorder (IED), conditions characterized by exaggerated aggressive reactions to mild stressors or ambiguous signals of threat or provocation. This taxonomy is similar to a widely used classification in animal research, namely appetitive and defensive aggression (Albert, Walsh, & Jonik, 1993). The former would be driven by a drive to attain rewards, whereas the latter would be enacted to avoid a punishment (Lang & Bradley, 2013). The distinction between proactive and reactive aggression is not clear cut, because the motives and precipitant factors that lead to aggression are hard to distinguish in practice (Bushman & Anderson, 2001). For instance, impulsive aggression in response to provocation can be driven by a wish to restore damaged honor (i.e. "face-saving"; Vasquez et al., 2013).

In this case, the aggressive act would subserve a larger goal beyond hurting the provoker. For the purposes of the present thesis we will nevertheless focus on the proactive versus reactive terminology because of its usefulness and intuitiveness. There are other classifications of aggression, for example verbal versus physical (Look, McCloskey, & Coccaro, 2015), or direct versus indirect (Card, Stucky, Sawalani, & Little, 2008). We will also use some of these other terms where appropriate, given that they are not mutually exclusive with the reactive versus proactive dichotomy.

1.2. Models of aggression

1.2.1. Early models

One of the first domain-specific models of aggression in modern psychology was Dollard and Miller's frustration-aggression theory (FAT; Dollard, Miller, Doob, Mowrer, & Sears, 1939). According to the FAT, blocking the attainment of goals relevant to the individual leads to aggression. Although this central tenet is still generally accepted, Dollard and colleagues originally affirmed that frustration was the *only* event that can cause aggression, and that *any* kind of frustration can serve as an aggression trigger (Berkowitz, 1989). The FAT has been toned down over the years upon discovering instances in which frustration does not univocally elicit aggression (e.g. when frustration is perceived as justified or reasonable; E. Burnstein & Worchel, 1962), and upon the realization that some stimuli do not involve goal-thwarting but can cause aggression nonetheless (e.g. physical pain; Berkowitz, 1983). A more recent reformulation, the cognitive neo-associationist theory (CNT), suggests that external events cause aggression insofar as they increase an organism's negative affect (Berkowitz, 1989). The CNT also postulates that aggression-related cues (e.g., words semantically linked to aggression, angry faces) increase the accessibility of hostile thoughts and aggressive behavioral scripts. The combination of negative affect, arousal, and hostile cognitions would

therefore result in a heightened readiness for aggression (Berkowitz, 2012). Further, the CNT predicts that whether a given event will elicit aggression depends on how an individual appraises said event, so that "one cannot say in the abstract whether thwartings are weak or strong instigators to aggression" (Berkowitz, 1989). This assumed idiosyncrasy constrains the CNT's predictive power, and is against a wealth of studies showing that many well-characterized situational factors can consistently elicit high aggression (Bettencourt, Talley, Benjamin, & Valentine, 2006).

1.2.2. The General Aggression Model

Due to the limitations of previous proposals, a less inferential and more encompassing theory of aggression was introduced in the early 21st century: the General Aggression Model (GAM). The GAM is a theoretical framework that classifies factors involved in an aggression episode at different stages of the conflict, and establishes temporal relationships between them (C. A. Anderson & Bushman, 2002). This model thus conceptualizes aggression episodes as sequential and cyclical in nature. Though mainly descriptive, the GAM provides a general and accessible point of departure to test more specific hypothesis on how a given set of variables modulate aggression in specific contexts and at given stages. Personal factors are those inherent to the individual, such as aggressiveness or brain structure. Situational factors refer to stimuli, conditions or events in the environment that can have an enhancing or deterring effect on aggression such as provocation, temperature, presence of bystanders, or availability of escape routes. Situational and personal factors interact to determine a person's affect, cognition, and physiological arousal, which in turn influence decision-making processes. A reasoned or an impulsive decision follows, which leads to the social encounter proper. The outcome of the encounter then feeds back to the persons and the situation, modifying subsequent events. Thus, the aftermath of an altercation can either lead to a reduction of the

ongoing hostilities, or, on the contrary, to violence escalation. For instance, after successive encounters with a stronger opponent, one might learn to refrain from further confrontation against him/her, whereas successive victories might positively reinforce aggressive behavior (Caramaschi, de Boer, de Vries, & Koolhaas, 2008; Hillman, 2013).

In the present thesis, we adopt the GAM to delineate different biological and situational aspects that intervene at different stages of an aggression episode. A schematic depiction of the GAM, as well as the focus of each chapter of the present thesis is presented in Figure 1.1.

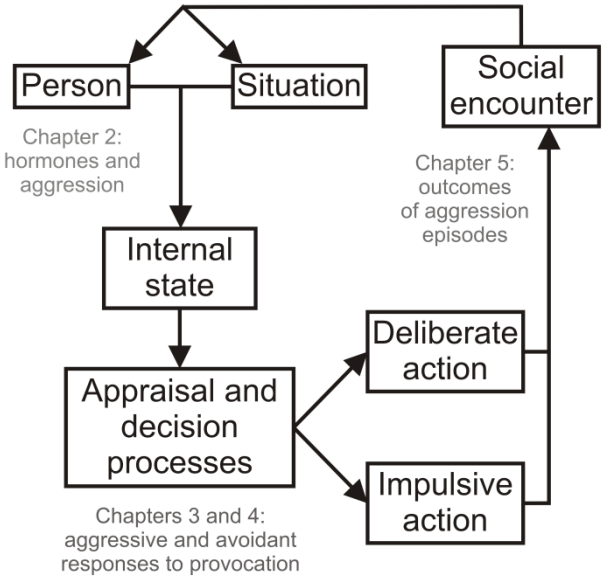

Figure 1.1 Schematic depiction of an aggression episode according the General Aggression Model (adapted from Anderson & Bushman, 2002), and the stage on which each of the chapters focused (gray).

The GAM has been criticized for failing to distinguish between reactive and proactive aggression, among other reasons (Ferguson & Dyck, 2012). Our work is agnostic with regards to the GAM's implicit assumptions. Rather, as already pointed out, we adopt the GAM as a general descriptive taxonomy.

1.2.3. Situational influences on aggression

Following the GAM, we will briefly review some of the most relevant aggression-inducing situational factors as well as personal variables that influence how such events are appraised and responded to. Provocation in the form of physical or verbal insults is a powerful trigger of aggression (Bettencourt, et al., 2006). In the laboratory, physical provocation is typically operationalized as loud noises delivered by an ostensible opponent, whereas verbal provocation is usually administered as negative feedback on an essay written by the participants or direct offensive remarks from a confederate (Ritter & Eslea, 2005). Other, more subtle forms of provocation exist, so that overt offense is not necessary to elicit aggression. Social exclusion, or ostracism, has been shown to increase aggressive behavior in a variety of aggression measures (Twenge, Baumeister, Tice, & Stucke, 2001; K. D. Williams, 2007). Ostracism is thought to threaten affiliative motivation (i.e., "the need to belong"), so that some individuals react aggressively when excluded. Frustration-based provocation, in which a confederate blocks the subjects' goals, is another form of provocation that can cause aggression without entailing direct verbal or physical insults (Goldstein, Davis, & Herman, 1975). Purely physical characteristics of the environment, such as hot temperature or unpleasant odors, can also increase aggression and hostility (C. A. Anderson & Bushman, 2002).

Although the focus of the present thesis lies on proximal or immediate situational inputs, it is also worth mentioning variables related to the learning and development of aggression. Socioeconomic level has been causally linked to aggression (Greitemeyer & Sagioglou, 2016). More specifically, lower perceived social rank leads to increased hostile expectations, which in turn facilitates aggressive behavior (J. R. Davis & Reyna, 2015; Greitemeyer & Sagioglou, 2016; Kraus, Horberg, Goetz, & Keltner, 2011). Indeed, correlational studies have shown that economically disadvantaged individuals are more likely

to be both victims and perpetrators of aggression (Chaux & Castellanos, 2015; Pabayo, Molnar, & Kawachi, 2014). In impoverished environments, aggression displays can be seen as a means to preserve one's reputation, especially in societies that place a high value on honor or in which people are educated to "stand their ground" (Cohen & Nisbett, 1994). Socioeconomic status also has on impact on families, being associated with conditions that ease the development of aggression such as poor and crowded housing, child abuse, neglect, parental criminality and drug addiction among others (Eron, Huesmann, & Zelli, 1991; Farrington, 1993; Haapasalo & Tremblay, 1994; Lewis, 1992; Widom, 1989).

1.2.4. Dispositional influences on aggression

We will now briefly comment on personal variables associated with aggression. Gender is one of the most well-established factors in this regard. Men are generally more physically aggressive than women, presumably as a result of evolutionary pressures (Campbell, 2013) and socialization practices (Nivette, Eisner, Malti, & Ribeaud, 2014). Men are more often involved in fights and inflict more severe harm (Graham et al., 2006; Parks, Osgood, Felson, Wells, & Graham, 2013), but women score higher in self-reports of anger (Gallardo-Pujol, Kramp, García-Forero, Pérez-Ramírez, & Andrés-Pueyo, 2006; Krämer, 2008). Across experimental studies, men have been shown to be more physically aggressive than women, although gender differences are lower for reactive than for proactive aggression (Bettencourt & Miller, 1996). In fact, in this meta-analysis the effect of provocation alone was about four times larger than that of gender. Another meta-analysis found that personality effects on aggression were at least twice as large as those of gender, and did not significantly differ between men and women (Bettencourt, et al., 2006). Hence, personality seems to predict aggression better than gender.

Personality dimensions such as trait anger, irritability, narcissism, and susceptibility have been repeatedly associated with aggression (Bettencourt, et al., 2006). Two important constructs in this regard are approach and avoidance motivation, which are defined by the extent to which an organism's behavior is driven by reward attainment or punishment avoidance, respectively (Carver & Harmon-Jones, 2009; McNaughton & Corr, 2004). Both dimensions are relatively independent and are conceptually related to the broader personality constructs of extraversion and neuroticism (Gable, Reis, & Elliot, 2000). Anger has been suggested to arise from approach motivation, as it is governed by a readiness to approximate and attack the source of frustration or provocation (Carver & Harmon-Jones, 2009). Fear can be considered as the counterpart of anger: it is also a negative, high-arousal emotion, but is characterized by the willingness to avoid the threat (Carver, 2004). Nevertheless, aggression can also be motivated by a desire to drive a threat away (McNaughton & Corr, 2004; Simunovic, Mifune, & Yamagishi, 2013). Hence, proactive or appetitive aggression would be mostly related to approach motivation, whereas reactive or defensive aggression could originate from *both* approach and avoidance motivation. It is however unclear whether the two dispositions might conflict when individuals must choose between retaliation and avoidance.

Another relevant personal factor is executive functioning, a broad set of interrelated cognitive processes involving the intentional control of behavior. Deficits in executive functions have been long suggested as a predisposing factor for aggression (Hawkins & Trobst, 2000). Although this relationship is well-documented in clinical or at-risk populations (Gul & Ahmad, 2014; Micai, Kavussanu, & Ring, 2015; Tonnaer, Cima, & Arntz, 2016), the evidence is weaker in healthy subjects (Krämer, Kopyciok, Richter, Rodriguez-Fornells, & Münte, 2011; Young et al., 2009), so that only specific measures of executive functions are predictive of aggressive behavior. Impulse control, understood as the capacity to suppress

momentary urges, has been put forward as the single most important executive function for the regulation of reactive aggression (Krämer, Kopyciok, et al., 2011; Lewis, 1992; Puhalla, Ammerman, Uyeji, Berman, & McCloskey, 2016; Strüber, Lück, & Roth, 2008). A longitudinal study found that response inhibition, a measure of impulse control, was more strongly related to behavioral disinhibition (i.e. vulnerability to externalizing disorders) than other executive functions such as working memory updating (Young, et al., 2009). In line with this, both implicit and explicit measures of self-control are related to lower self-reported aggression (Keatley, Allom, & Mullan, 2017). Thus, the evidence suggests that impulse control is a better predictor of reactive aggression than general executive functioning, at least in healthy individuals.

One must anyway bear in mind that the commented personal and situational factors might have interactive effects on aggression, as posited by the GAM. For example, narcissistic persons react more aggressively than others when provoked, but need not be necessarily more aggressive than others in the absence of threats (Rasmussen, 2016). Some personal factors can actually be defined by an individual's tendency to react in a specific manner to situational variables (Mischel & Shoda, 1995). In the context of aggression, some people may consistently react aggressively to ambiguous threat signals. The tendency to interpret others' behavior as threatening or provoking is called *hostile interpretation bias*, and it has been shown to be a basic predisposing factor for aggression (Wilkowski & Robinson, 2010). Hence, it is generally advisable to consider social and personal factors in mutual interaction rather than in isolation.

1.3. Measuring aggression

1.3.1. The Taylor Aggression Paradigm

Many behavioral and self-report methods to measure aggression have been developed following the concepts presented in the two previous sections. The Taylor Aggression Paradigm (TAP) is one of the most widely used measures of aggression in experimental settings (Taylor, 1967). In the TAP, participants are led to believe that they will play a competitive reaction time task against a purported opponent. In reality, the game is controlled by the experimenters. The goal of the task is to be quicker than the opponent in pressing a button when a target stimulus appears. In each round, participants must choose the intensity and/or duration of an aversive stimulus (e.g. a sound blast, an electric shock) that will be delivered to the other participant should they win the reaction time task. Since duration and intensity tend to be highly correlated, an aggregate of the two is often used as the dependent variable. The paradigm is set such that the fictitious opponent delivers increasingly intense aversive stimulation, which typically causes participants to retaliate. A concern related to this point is that, by nature of being preprogrammed, the TAP might not realistically recreate spontaneous aggressive encounters. However, in one version of the TAP, the game was actually interactive, such that two opponents played in real time with each other. In this study, aggression escalated similarly as in a non-interactive version, the degree of escalation being dependent on participants' trait aggressiveness (C. A. Anderson, Buckley, & Carnagey, 2008).

1.3.2. The Point Subtraction Aggression Paradigm

In the point subtraction aggression paradigm (PSAP), participants are placed in a fictional multiplayer game with three possible response options, of which at least one is aggressive (Cherek, 1981). The first, standard response consists of pressing a button which yields a small monetary reward (e.g. 0.10 €) after a fixed number of presses (typically 100). The second

response option is to subtract the same amount of money from another player after pressing a second button a lower number of times (typically 10). This response yields greater benefits for less effort at the expense of the other participant's earnings, and it is hence considered an aggressive option. In the original study, the third option was also aggressive (i.e. delivering a sound blast to the opponent after pressing a third button 10 times). In most versions of the task, however, the third option is often implemented as a defensive response, i.e. pressing a third button 10 times to ensure a provocation-free interval (e.g. 250 seconds). In reality, there is no other player, and the task is preprogrammed to provoke participants. Specifically, the fictional co-player uses the aggressive option at random a number of times in the course of the game, usually causing participants to retaliate (Gallardo-Pujol, Andrés-Pueyo, & Maydeu-Olivares, 2013; Gan et al., 2016).

1.3.3. The Buss Aggression Machine

The teacher/learner paradigm, also called Buss Aggression Machine (BAM), is an experimental task in which the subject (the "teacher") is rewarded depending on the performance of another subject (the "learner"), who is a confederate of the experimenters (Goldstein, et al., 1975). Participants are allowed to choose the duration and/or intensity of an electroshock, which is administered to the learner if he/she commits an error. Since the task is easy and the learner often commits many mistakes, the task is set to induce aggression through frustration. The main problem with the BAM is that punishments might be driven by profit maximization, or even by prosocial motives (i.e. helping the learner to improve his/her performance; R. A. Baron & Eggleston, 1972), which questions its validity as an aggression measure.

1.3.4. The Ultimatum Game

The Ultimatum Game (UG) is a well-known economic exchange paradigm in which two players must split a sum of money (e.g. 10 €). One player -the *proposer*- makes an offer on how to divide the money (e.g. 6:4 €), and then the other player –the *receiver*- decides whether to take it or not. If the receiver accepts the offer, the money is split as proposed. If the receiver rejects the offer, nobody gets any money. Although the receiver has no reason to decline offers, people tend to refuse unfair proposals (e.g. 9:1 €). Rejections are often interpreted as aggressive responses to unfairness, although the underlying motives are still hotly debated (Corradi-Dell'Acqua, Civai, Rumiati, & Fink, 2012; White, Brislin, Sinclair, & Blair, 2014). For instance, rejections could be a means to enforce equity, but they could also represent a spiteful revenge to a perceived insult. Another limitation of the UG is that participants do not incur physical harm, but simply lose part of the participant endowment, and thus do not suffer real losses.

1.3.5. The Preemptive Strike Game

In the Preemptive Strike Game (PSG), participants have a limited time to decide whether to press a button or do nothing. If they press the button, they lose a small amount of money and cause an alleged co-player to lose a high amount. If they do nothing, both participant and co-player receive the maximum endowment. The profit-maximizing strategy is to do nothing, but some participants attack preemptively in order to prevent a higher loss, and more so if they are led to believe that the risk of being attacked is high (Halevy, 2017). The PSG has the unique advantage of measuring defensive aggression before any attack ensues, but it has the problem of being a one-shot interaction with a binary outcome, thus potentially failing to capture both intra- and interindividual variability in aggression. Furthermore, just as the UG, it only entails economic but not physical punishments.

1.3.6. The Voodoo Doll Task

In the Voodoo Doll Task (VDT), participants are asked to place needles on a doll that allegedly represents the person to be harmed. This paradigm has been extensively validated against a number of self-report and behavioral aggression measures, and is especially suited to capture aggressive tendencies in close personal relationships (DeWall et al., 2013). The main limitation of the VDP lies in that it is a fundamentally indirect measure of aggression, which renders it uninteresting for research focused on direct verbal or physical responses to provocation. Moreover, scores in this paradigm might be affected by other confounds such as superstitious beliefs. The VDP has the advantage that it can be measured on the field, without the need for computers or electronic measurement devices, and it can be expected to be more sensitive to spontaneous fluctuations in aggressive tendencies than daily diaries. In a study in married couples, the number of pins participants inserted on the doll (representing their partner) was positively correlated with the intensity of sound blasts administered to their spouse in the TAP, but negatively associated with blood glucose levels (Bushman, DeWall, Pond, & Hanus, 2014). The authors affirm that low glucose would lead to an impaired ability to regulate aggressive impulses. Nevertheless, since the pins do not entail direct pain infliction, the VDP can be assumed to measure aggressive inclinations, but not aggression proper.

1.3.7. The Hot Sauce Paradigm

In the hot sauce paradigm (HSP), participants are asked to pour any amount of spicy sauce on a recipient. The sauce is to be allegedly consumed by another person who dislikes hot sauce, and who has in some way provoked participants in a previous experimental situation (J. D. Lieberman, Solomon, Greenberg, & McGregor, 1999). The HSP was inspired by real events in which spicy sauce had been deliberately used to harm other persons. For instance, a New

Hampshire cook who had a negative attitude towards the police reportedly poured hot sauce on the breakfast of two officers, causing them extreme digestive discomfort. The aggressor was subsequently arrested for assault (J. D. Lieberman, et al., 1999). The HSP has the advantage that participants' behavior is unambiguously aggressive, since it does not serve any competitive or performance-related purpose. Nevertheless, participants are not directly responsible for the putative harm inflicted to the target, as the experimenter is allegedly the one serving the hot sauce to the provoker. Furthermore, participants' own food preferences might play a role in the amount of hot sauce allocated, and the paradigm requires a convoluted cover story to be believable.

1.3.8. Evaluation-based paradigms

In this type of task, subjects must rate a person in terms of competence or likeability, with lower scores indicating higher aggression (Greitemeyer & Sagioglou, 2016). In some versions, participants have to rate something that the person has crafted or written (e.g. an essay; Ritter & Eslea, 2005). A cover story is often used to give the impression that these ratings are relevant for the evaluated person. For instance, participants might be told that depending on the obtained scores, the person's contract might be extended. These paradigms are however a measure of verbal rather than physical aggression, since there is no direct confrontation or physical harm involved.

1.3.9. Other laboratory measures

Many other aggression tasks exist. In the Bungled Procedure Paradigm, participants are given the chance to choose among a set of toy guns to shoot at another person wearing protective gear. The power of the gun and the number of pellets shot are then considered a proxy for aggression (Ritter & Eslea, 2005). In the Bobo Modelling Paradigm, subjects -typically children- see an adult punching or ignoring an inflatable clown doll in a room where other

toys are present, and then are allowed to play in the same room (Ritter & Eslea, 2005). Children's aggressiveness towards the doll constitutes the main aggression measure. These tasks can be deemed to have excessive face validity (i.e. it might be too obvious that the outcome variable is aggression), are markedly directive (i.e. aggressive behavior is strongly cued by the setting), and subjects' behavior might ultimately reflect horseplay rather than genuine aggressive intent. See Tedeschi & Quigley (1996) and Ritter & Eslea (2005) for a critical account of classic laboratory aggression paradigms.

More recently, a series of tasks have been devised to measure prosocial relative to antisocial tendencies. In the SoMi paradigm, participants are allowed to choose a product among a range of repeated or unique items. Chosen items allegedly become no longer available for subsequent participants. In reality, however, all participants choose from the same assortment. As single and repeated items have an almost equivalent value (i.e. a red vs three green apples), if participants choose a unique item, this is interpreted as a hostile response towards other subjects (Van Doesum, Van Prooijen, Verburgh, & Van Lange, 2016). Indeed, participants make more selfish decisions when they think that the next subject belongs to a rival group, whereas, when the next player allegedly belongs to the own group, they choose less selfishly. Similar to this paradigm is the Tangram Help/Hurt Task (THHT; Saleem, Anderson, & Barlett, 2015). In the THHT, participants have to assign a number of puzzles to another participant, who can purportedly obtain a reward if he/she solves all puzzles within a time limit. Puzzles are classified as easy, medium, or hard depending on the number of pieces involved. Hence, if participants assign a greater proportion of hard puzzles to the other participants, this can be considered a hostile response, whereas, if they assign a greater proportion of easy puzzles, this can be considered a prosocial behavior. The THHT has shown convergent validity with other self-report and behavioral aggression measures (including the HSP). Its main limitation, shared with other measures such as the SoMi or the

UG, is that it does not entail direct harm to any of the players. A version of the TAP in which participants could choose not to shock the opponent partly overcame this limitation (Zeichner, Frey, Parrott, & Butryn, 1999). However, shock frequency correlated with the usual parameters of intensity and duration, suggesting that the amount of times participants choose not to punish the opponent reflects reduced aggression rather than prosocial behavior.

1.3.10. Self-reports

Another way to measure aggression are self-reports. One of the most used measures is the Aggression Questionnaire (Buss & Perry, 1992), which assesses the general tendency to behave aggressively (example item: "Given enough provocation, I may hit another person"). The AQ has 27 items divided in four scales, though shorter versions have also been developed (Bryant & Smith, 2001; Webster et al., 2014). While the AQ focuses on reactive aggression, the Reactive-Proactive Aggression Questionnaire (RPAQ; Raine, et al., 2006) also allows to tap into proactive aggression. The Life History of Aggression (LHA) quantifies the lifetime occurrence of aggression episodes, which is highly relevant from a developmental perspective, and *a priori* more specific than self-appraisals (Coccaro, Berman, & Kavoussi, 1997). While highly useful in clinical settings, the LHA might suffer from a floor effect in healthy participants, i.e. most will tend to have low or null scores (Soloff et al., 2017). Other questionnaires rather assess personality characteristics that predispose to aggression, namely trait anger (Forgays, Forgays, & Spielberger, 1997) or hostility (Velicer, Govia, Cherico, & Corriveau, 1985). Some other, less commonly utilized instruments such as the Normative Beliefs about Aggression Scale (NOBAGS; Huesmann & Guerra, 1997) attempt to capture subjects' attitudes towards the use of aggression and violence. Indeed, people who endorse aggression might be more likely to enact it, which makes attitudinal questionnaires doubtlessly relevant to study cultural or intergroup influences on aggression. However, these

questionnaires typically tap subjects' opinion towards aggression in general but do not measure whether they would behave aggressively themselves in a given situation. Thus, in many instances it can be preferable to employ measures that assess individual aggressive propensity. Regardless of the exact underlying construct, self-reports might be influenced by social desirability, i.e. the tendency to respond in a manner consistent with acceptable societal practices. The latter bias might be an especially poignant problem in aggression questionnaires because aggression is often seen as an undesirable behavior and is punished by law (Vigil-Colet, Ruiz-Pamies, Anguiano-Carrasco, & Lorenzo-Seva, 2012).

1.3.11. Observational and experience sampling methods

Aggressive behavior can also be measured *in situ* (Graham, et al., 2006; Parks, et al., 2013). In observational studies, researchers go to public places in order to count and characterize aggression episodes (e.g. number and gender of persons involved). This methodology has high external validity, but it does not provide the experimental control of laboratory methods. For instance, in a barroom brawl, the observer might miss out on the event that triggered the hostilities, and, although the degree of inebriation can be inferred from behavior (Parks, et al., 2013), the exact amount of alcohol ingested is hard to determine. Other methodologies to measure aggressive behavior in natural settings involve the use of daily diaries, or, more recently, device-based measurements at random times during the day (Hillbrand, Waite, Miller, Spitz, & Lingswiler, 2000). These measures -often termed experience sampling or ecological momentary assessment- are also free from the artificial constrains of the laboratory, but are subject to the uncontrollability of observational studies. They also suffer from most limitations inherent to questionnaires, but they do limit retrospective bias (i.e. incorrect recall of past events; Barrett & Barrett, 2001). Albeit in the present thesis we only employed experimental and self-report measures of aggression, ideally one should adopt a

multimethod approach, since each measurement technique can provide complementary information.

1.4. Evolutionary basis of aggressive behavior

We will subsequently comment on the biological underpinnings of aggression, starting with its putative evolutionary basis. The idea that aggression can be grounded in evolution at all has been strongly criticized, mostly adducing the danger of biological determinism and the associated notion that aggression is inevitable. The most blatant critique in this regard is arguably the Seville Statement on Violence (SSV), which affirmed that claims of violent behavior being genetically or biologically determined are scientifically inaccurate (Adams et al., 1990). This position can be traced back to Rousseau and his notion of the natural goodness of man, i.e. that humans are born good and are corrupted by society (Elliott, 1988). Such "blank slate" notion has been the official stance in mainstream academic psychology during a good part of the 20th century, partly due to the preponderance of behaviorism (Pinker, 2002). Although the founding authors of behaviorism did not deny the importance of evolution for the development of innate responses (Ludvig, 2003), they did place the explanatory burden of goal-directed behavior on the environment. From the behaviorist perspective, thus, aggressive behavior would be learned by reinforcement, that is, on the basis of its beneficial or detrimental consequences for the organism (Rachlin, 2004). Over the years, the radical behaviorist perspective expanded, including not only environmental contingencies, but also how subjects apprehended and processed those contingencies (Eron, 1987). Social learning theorists added a new dimension to this, arguing that humans can learn to behave aggressively not only by their own experience, but also by third-person accounts. Nevertheless, the blank slate notion persisted: "People are not born with preformed repertoires of aggressive behavior; they must learn them" (Bandura, 1978).

On the opposite end one can find theories emphasizing the role of biology above all. Perhaps the most influential amongst these is the "killer ape theory" (Dart, 1953), which suggests that human evolution selected for aggression over other traits. That is, being aggressive would have provided a competitive advantage that would have been favored in the "evolutionary arms race". In Dart's own words: "[…] man's predecessors differed from living apes in being confirmed killers: carnivorous creatures, that seized living quarries by violence, battered them to death, tore apart their broken bodies, dismembered them limb from limb, slaking their ravenous thirst with the hot blood of victims and greedily devouring livid writhing flesh" (Dart, 1953). These proposals were most famously represented in Stanley Kubrick and Arthur C. Clarke's film *2001: a Space Odyssey*. In the movie, it is implied that it was not only the ability to grasp, but also to *kill* with the grasped object, what allowed apes to evolve into mankind.

The biological perspective was also espoused by a number of ethologists, chiefly Konrad Lorenz. According to this author, aggression is bound to arise no matter the circumstance: "Knowledge of the fact that the aggression drive is a true, primarily species-preserving instinct enables us to recognize its full danger: it is the spontaneity of the instinct that makes it so dangerous." (Lorenz, 1963/2005). This view links with the psychoanalytic notion of the death wish (Slap, 1967) and presupposes the existence of a destructive tendency in the individual that can be only temporarily repressed. These theories have been largely disproven because there is no evidence that aggression has been evolutionarily selected over other behavioral traits. Not only that, but it has been completely overturned, since it appears to be prosocial, rather than antisocial behavior what allowed humans to thrive (Hare, 2017). If anything, human evolution favored the ability to understand others and to cooperate with them, rather than the ability to physically injure them. This conversely implies that aggressive

behavior –especially reactive aggression- is relatively better conserved phylogenetically than other behaviors such as empathy (Albert, et al., 1993; Blanchard & Blanchard, 2003).

Purely biological accounts of aggression as an instinctive impulse, including the Lorenzian and Freudian formulations, are inherently limited, as are those emphasizing the role of the environment, such as social learning theories. Evolution has produced general cognitive and biological *mechanisms* which can give rise to aggressive behavior depending on contextual influences (Buss & Shackelford, 1997). From this point of view, the biological basis of aggression would be partly shared with other cognitive processes such as emotion regulation (Davidson, Putnam, & Larson, 2000) or moral reasoning (Raine & Yang, 2006), and would be highly subject to social and environmental inputs. Aggression, as many complex behaviors, is a biosocial phenomenon, and can only be understood by examining the interaction between these domains (Scarpa & Raine, 2007).

1.5. Brain structure and function involved in reactive aggression

1.5.1. Neurobiological models of reactive aggression

In the following sections we will summarize some of the most relevant neuroscientific literature on aggressive behavior with a focus on reactive aggression. Early neurobiological models understood reactive aggression as a result of impaired prefrontal control over subcortical areas (Raine, Buchsbaum, & Lacasse, 1997; Raine et al., 1998). More specifically, these models suggested that aggressive impulses originate in regions such as the brainstem or the amygdala, which then receive downstream inhibitory input from prefrontal areas. These basic notions have been shown to be overly simplistic and anatomically unspecific, as increased (rather than decreased) activity in different regions of the prefrontal cortex (PFC) has been related to aggression and antisocial behavior. A sequential model was hence put forward which acknowledged the role of top-down control pathways, but also included

reciprocal connections between subcortical and prefrontal areas involved in emotion processing (Coccaro, Sripada, Yanowitch, & Phan, 2011). According to this model, the amygdala and brainstem receive inhibitory input from dorsal and lateral prefrontal regions, but have bidirectional links with more ventral areas of the prefrontal cortex. Crosstalk between these regions would then ultimately determine the consummation of aggressive impulses. Notice that some regions such as the anterior cingulate cortex (ACC) or the medial orbitofrontal cortex (OFC) are considered to act both as a regulatory and a reciprocal influence, suggesting that they might have a flexible role in aggression. The model is depicted in figure 1.2.

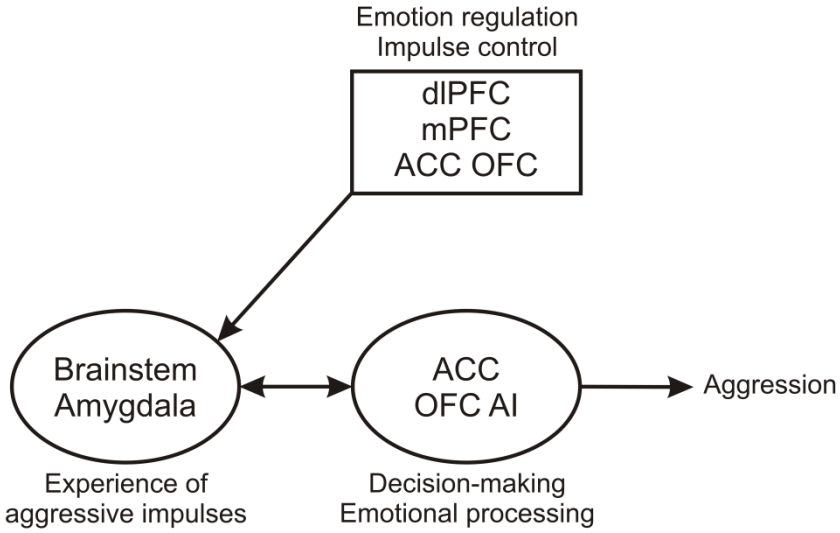

Figure 1.2 Schematic depiction of a sequential model of reactive aggression (modified from Coccaro, Sripada, Yanowitch, & Phan, 2011). DLPFC: dorsolateral prefrontal cortex, DMPFC: dorsal-medial prefrontal cortex, ACC: anterior cingulate cortex. OFC: orbitofrontal cortex, AI: anterior insula.

1.5.2. Animal studies

First, we will turn to the relationship between brain structure and aggression. This has been extensively investigated by means of lesion or stimulation studies in animals, mainly mammals such as cats or mice. Stimulation has been traditionally performed by applying

electrical currents to specific brain regions. In recent years, however, optogenetics has gained popularity as the technique of choice to more accurately target specific neuronal populations. Optogenetic methods permit to genetically modify predefined cell groups in order to make them sensitive to light, and then increase or decrease their activity by applying light of the corresponding wavelength (Deisseroth, 2011). Classic stimulation studies have long revealed a core network of brain regions whose stimulation causes aggression, namely the periaqueductal gray (PAG), the hypothalamus, and the amygdala (Gregg & Siegel, 2001; Siegel, Roeling, Gregg, & Kruk, 1999). These subcortical areas are thought to enhance negative affect and peripheral physiological activation, facilitating attack readiness (Halasz, Liposits, Meelis, Kruk, & Haller, 2002). Both electrical and optogenetic stimulation of the hypothalamus causes aggression in rats, especially in those with stronger preexisting aggressive tendencies (Albert, Nanji, Brayley, & Madryga, 1979; Falkner, Grosenick, Davidson, Deisseroth, & Lin, 2016; Panksepp, 1971; Woodworth, 1971). Conversely, lesions of the anterior hypothalamus can reduce aggression in mice (Shivers & Edwards, 1978). Although the hypothalamus is also involved in sexual behavior, optogenetic stimulation of a specific hypothalamic subregion (ventrolateral subdivision of the ventromedial hypothalamus or VMHvl) causes indiscriminate aggression, and neurons in this area are inactivated during mating in mice (D. Lin et al., 2011). Crucially, genetic or pharmacologic inactivation of this structure has been shown to reduce aggression across a range of studies, suggesting that this is a highly domain-specific structure for aggression (Falkner & Lin, 2014). Electric stimulation of the amygdala (Ma & Kanwal, 2014; Potegal, Hebert, DeCoster, & Meyerhoff, 1996) has a comparable effect as that of the hypothalamus, although the latter structure seems to be necessary for PAG and amygdala activation to trigger aggression outbursts in animals (Halasz, et al., 2002). The orbitofrontal cortex (OFC) has been long suspected to be an important prefrontal region for impulse regulation, including aggressive impulses, due to its

dense reciprocal connections with the medial temporal lobe, in which the amygdala is located (Stalnaker, Cooch, & Schoenbaum, 2015). Indeed, OFC lesions have been reported to slightly increase aggression and impulsivity in rats (Rudebeck et al., 2007). However, OFC lesions in monkeys lead to increased social avoidance, rather than aggression (Butter, Mishkin, & Mirsky, 1968). Furthermore, optogenetic activation of the dorsal medial prefrontal cortex (mPFC), but not the OFC, reduces excessive aggression in mice (A. Takahashi, Nagayasu, Nishitani, Kaneko, & Koide, 2014). Hence, the exact role of the OFC in aggression remains obscure. Animal studies also argue for the involvement of the reward system in aggression (Ferrari, Van Erp, Tornatzky, & Miczek, 2003), such that stimulation of dopaminergic midbrain sites potentiates aggression (Q. Yu et al., 2014). Some findings obtained from stimulation and lesion studies have also been confirmed by immunohistochemistry, a technique in which the c-Fos protein is often employed as a post-mortem index of recent neuronal firing (Dragunow & Faull, 1989). These studies have shown that mouse-killing rats have increased activity in the amygdala and lateral hypothalamus (Tulogdi et al., 2015), and that aggressive encounters enhance activation of these same regions, particularly in rats selected for aggression (Konoshenko, Timoshenko, & Plyusnina, 2013; Nehrenberg, Sheikh, & Ghashghaei, 2013). Summing up, animal studies have highlighted a core aggression network comprising subcortical structures such as the amygdala, hypothalamus, and brainstem –particularly the PAG.

1.5.3. Lesion studies in humans

Animal studies have the advantage of a high experimental control regarding the precise brain site of lesion or stimulation, and permit to delineate the gross anatomy of aggression, as well as its cellular and molecular underpinnings. Nonetheless, the folding of the neocortex -particularly in prefrontal regions- is much greater in humans than in most other mammals

(Kaas, 2013; Zilles, Palomero-Gallagher, & Amunts, 2013). Thus, animal studies mostly allow us to understand the role of subcortical structures, but the information they can provide on cortical regions in inherently limited. Furthermore, despite the commented phylogenetic continuity between animals and humans in reactive aggression, the underlying subjective state of animals is difficult to infer. Even if basic affective experiences, such as fear, do seem to have animal homologues (Panksepp, 2011), complex social emotions that can elicit aggression in humans, such as shame (Thomaes, Bushman, Stegge, & Olthof, 2008) or spite (D. K. Marcus, Zeigler-Hill, Mercer, & Norris, 2014), hardly have a parallel in the animal realm.

Naturally occurring lesions in humans can overcome such limitations and have thus also been used as a valuable source of information. Starting from the well-known but oversimplified case of Phineas Gage in the 19th century (see Kotowicz, 2007 for discussion), lesion studies in humans have proven valuable to understand aggression and its neural basis. In general, there is agreement that prefrontal lesions can increase aggression (Hawkins & Trobst, 2000). Studies in veterans show an interesting dissociation within the PFC, such that dorsolateral prefrontal (dlPFC) lesions are associated with more positive implicit attitudes toward violence (Cristofori et al., 2016), whereas ventrolateral PFC (vlPFC) damage is linked to overt verbal aggression (e.g. tantrums, threats) as assessed by friends and family of the patients (Grafman et al., 1996). Bilateral OFC damage has been reported to cause aggressive and callous behavior with a specific impairment in the recognition and response to angry expressions, without affecting other cognitive functions (Blair & Cipolotti, 2000). However, this report was based on a single case, and later evidence suggests that OFC lesions cause general deficits in associative learning rather than in social information processing specifically (Rolls & Grabenhorst, 2008). Neurosurgery has also yielded important insights into the neurophysiology of aggression. Most prominently, cauterization of the posterior

hypothalamus and the amygdala has been shown to reduce aggression in pathologically aggressive patients (Fountas & Smith, 2007; Sano, Mayanagi, Sekino, Ogashiwa, & Ishijima, 1970). In summary, lesion studies in humans tend to corroborate findings in nonhuman mammals, though highlighting a more prominent role of the PFC in aggression regulation.

1.5.4. Stimulation studies in humans

While highly informative, lesion studies in humans have several disadvantages, namely, that lesions are often spread over several functionally and anatomically distinct regions, and that there might be certain lesion-derived changes in distal areas which might carry their own behavioral consequences. In addition, there is great heterogeneity in the degree of damage that a brain can sustain and still function normally (Forsdyke, 2015; García, Sedeño, Herrera Murcia, Couto, & Ibáñez, 2017). Deep-brain stimulation (DBS), an invasive technique in which electrodes are implanted at specific brain sites, offer a greater anatomical precision than lesion studies, although, as in the case of psychosurgery, its application in psychiatric patients is highly controversial (Faria, 2013). In any case, DBS has also been used to treat severe cases of aggression in humans. DBS in the posterior hypothalamus has been shown to reduce aggression, perhaps by normalizing this structure's activity (Franzini, Broggi, Cordella, Dones, & Messina, 2013; Micieli, Rios, Aguilar, Posada, & Hutchison, 2016), whereas amygdala stimulation causes rage bouts (Faria, 2013). Note anyway that the specificity of locally implanted electrodes does not preclude tissue damage caused by surgery. Furthermore, DBS is used in patients, such that disease- and stimulation-related effects are confounded.

Non-invasive brain stimulation approaches have also been developed to identify brain regions that can modulate aggression. Such studies can be carried out in larger samples of healthy participants, making findings *a priori* more generalizable than those obtained in

patients. Researchers have mostly employed electric currents (transcranial Direct Current Stimulation; tDCS) to stimulate or inhibit aggression-regulatory centers, although magnetic pulses (transcranial magnetic stimulation, TMS) have also been employed. It has been reported that tDCS over the right vlPFC reduces aggression after social exclusion (Riva, Romero Lauro, DeWall, Chester, & Bushman, 2014). Similarly, tDCS over the same region caused a general decrease in aggressive behavior in the TAP, and especially after playing violent videogames (Riva et al., 2016). Such findings dovetail with a study showing that an almost identical stimulation regime (i.e. tDCS over the right vlPFC, though at a slightly lower intensity) reduced impulsivity in a response inhibition task (Jacobson, Javitt, & Lavidor, 2011). Stimulation of the right dlPFC also seems to reduce aggression, but results are more mixed. In one study, dlPFC stimulation reduced proactive, but not reactive aggression in men only (Dambacher et al., 2015b), whereas, in another study, it reduced reactive aggression in both genders depending on state anger (Hortensius, Schutter, & Harmon-Jones, 2011). Comparable results were observed in a TMS investigation in which inactivation of the left dlPFC increased both proactive and reactive aggression in the PSAP (Perach-Barzilay et al., 2013). Across studies, the most consistent finding is that the aggression-reducing effect seems to be stronger when the right hemisphere is upregulated, or conversely, when the left hemisphere is downregulated (Dambacher et al., 2015a). Another TMS study supports this notion by showing that participants with stronger leftward motor-evoked potentials (MEP, i.e. muscular activity elicited by TMS over primary motor areas) had higher trait aggression scores and a more marked attentional bias toward angry faces (Hofman & Schutter, 2009). Similarly, violent offenders show higher left MEP when two TMS pulses are presented consecutively (i.e., reduced cortical inhibition) than healthy controls (Philipp-Wiegmann et al., 2011). This pattern that has also been observed in Attention Deficit Hyperactivity Disorder (ADHD) patients relative to controls, and the effect correlates with symptomatology

scores (Schneider et al., 2007). All in all, non-invasive stimulation studies offer good evidence for a role of the dlPFC and vlPFC in stopping aggressive impulses. The latter region is part of the inferior frontal gyrus (IFG), a lateral cortical area that has been consistently related to response stopping (Aron, Robbins, & Poldrack, 2014), but also to empathy (Liakakis, Nickel, & Seitz, 2011; M. D. Lieberman, 2007). Hence, vlPFC might buffer aggression through enhanced impulse control and, to a lesser extent, by an increased commiseration toward the rival.

1.5.5. Gray matter structure and aggression

A limitation common to lesions and other methods of stimulation is that they might be affecting system-wide function, so that collateral effects on other brain regions cannot be readily determined. Furthermore, aggression-related areas are distributed across the brain, but lesion and stimulation studies must, by necessity, focus on one or few specific regions at a time –in the case of tDCS and TMS, only cortical regions bordering the skull can be targeted.

Since the irruption of Magnetic Resonance Imaging (MRI) in the seventies it has become possible to observe and quantify brain structure *in vivo* across the whole brain. Many structural MRI studies have hence found associations between gray matter volume or cortical thickness and aggression. A classic study on this topic found that individuals with Antisocial Personality Disorder (APD), who typically display exaggerated aggressiveness, have reductions of around 11% in general prefrontal volume relative to healthy controls and substance-dependent subjects (Raine, Lencz, Bihrle, LaCasse, & Colletti, 2000). More recent investigations have narrowed down the focus to specific regions of the PFC, suggesting that volume reductions in APD patients are most pronounced in OFC and dlPFC (Leutgeb et al., 2016; Raine, Yang, Narr, & Toga, 2011; Tiihonen et al., 2008). Furthermore, volume in these regions has been negatively associated with both self- and clinician-rated antisocial behavior

(Raine, et al., 2011). Another study showed that offenders with a diagnosis of Borderline Personality Disorder (BPD), a psychiatric illness characterized by high dispositional impulsivity, have lower gray matter volume in orbitofrontal and frontopolar cortex as compared to healthy controls (Bertsch et al., 2013). Similarly, adolescents with conduct disorder (CD, the term for APD in underage individuals) have reduced orbitofrontal and inferior frontal cortex volume compared to healthy controls *and* ADHD patients (Stevens & Haney-Caron, 2012). Within ADHD patients, cortical thickness of the OFC is negatively related to aggression (Cha et al., 2015), as has been observed in a heterogeneous sample of psychiatric patients (Gansler et al., 2009). Perhaps surprisingly, lower amygdala volume has also been consistently associated to increased self-reported aggression across a range of populations (Driessen, Herrmann, Stahl, & et al., 2000; Gopal et al., 2013; Pardini, Raine, Erickson, & Loeber, 2014; Thijssen et al., 2015; Y. Yang, Raine, Narr, Colletti, & Toga, 2009; Zhang et al., 2013). Other research has shown that volume of the striatum, a brain region involved in reward processing and action initiation (Balleine, Delgado, & Hikosaka, 2007), is associated with greater aggression in children (Ducharme et al., 2011), violent offenders (Schiffer, Müller, Scherbaum, & et al., 2011) and psychopaths (Glenn, Raine, Yaralian, & Yang, 2010). Such findings suggest that a relatively larger striatum might lead to aggression partly due to increased reward dependence, given that striatal volume has been related to heightened novelty seeking and drug consumption (Churchwell, Carey, Ferrett, Stein, & Yurgelun-Todd, 2012; Li et al., 2015) as well as reduced anhedonia (Pizzagalli et al., 2009). This first overview of structural variability associated with aggression reveals an initial pattern, such that volume of the amygdala and of areas involved in cognitive control and emotion regulation are generally associated with lower aggression, whereas relatively larger striatal volumes are related to increased aggressive proclivity.

1.5.6. White matter structure and aggression

Inspecting brain volume and cortical thickness offers an initial glimpse into the basic neural architecture of human aggression. However, a more informative picture can be obtained by additionally studying the connectivity between brain regions, which is subserved by white matter. The most widely used technique to map white matter is diffusion tensor imaging (DTI), which measures the diffusion of water particles across axonal fibers and thereby provides different estimates on the efficiency of a given white matter tract. In this regard, integrity of the uncinate fasciculus (UF), which connects the amygdala and the OFC, has been shown to be compromised in patients with antisocial personality disorder (ASPD; Sundram et al., 2012) and psychopathy (Craig et al., 2009; Motzkin, Newman, Kiehl, & Koenigs, 2011), as well as in aggressive football players with a history of multiple concussions (Goswami et al., 2016). Interestingly, these aggression-related deficits seem to be highly specific for the right UF, since other psychiatric diseases seem to involve rather the left or the bilateral UF (Von Der Heide, Skipper, Klobusicky, & Olson, 2013). However, the UF is not related to aggression in healthy subjects (Beyer, Münte, Wiechert, Heldmann, & Krämer, 2014), suggesting that this anatomical connection is mostly relevant for pathological aggression. Aside from the UF, there is recent converging evidence linking structural properties of the superior longitudinal fasciculus (SLF) to reactive aggression (Karlsgodt et al., 2015; R. Lee et al., 2016). The SLF connects dorsal aspects of the frontal cortex with parietal regions, and it is thus considered central for executive functions, although it has also been implicated in empathy (Díez-Cirarda et al., 2015; Parkinson & Wheatley, 2012). SLF integrity has been shown to mediate the negative relationship between working memory performance and self-reported aggression (Karlsgodt, et al., 2015), is related to impulsive aggression –but not psychopathy-, and is generally reduced in IED relative to healthy controls and non-IED psychiatric patients (R. Lee, et al., 2016). This suggests that, indeed, structural abnormalities

in the SLF might lead to impairments in the regulation of aggressive impulses. Future studies could test the specificity of UF and SLF impairment by directly comparing samples characterized *a priori* by proactive versus reactive aggression.

1.5.7. Electroencephalographic studies

Aggression-related structural anomalies are most marked in pathological conditions, but in healthy subjects aggression is probably more strongly influenced by brain function than by structural variability. Brain function associated with aggressive behavior can be studied using electroencephalography (EEG), which measures cortical electric potentials at the scalp. There is some evidence that resting-state EEG measures can predict aggression, particularly in the alpha frequency band (≈8-15 Hz) over frontal areas (Keune et al., 2012; Peterson, Shackman, & Harmon-Jones, 2008). A study reported that resting frontal alpha power is associated with increased aggression in adolescents (Niv et al., 2015). Dovetailing this finding, latency of alpha phase shifting between frontal and parietal electrodes at rest -indicating a sluggish coordination between these cortical regions- has also been associated with self-reported aggression (Lackner et al., 2014). Furthermore, asymmetry between low- and high-frequency oscillations is related to increased dominance and a heightened attentional bias toward angry faces (Hofman, Terburg, van Wielink, & Schutter, 2013). These findings indicate that lower cortical arousal might reflect deficiencies in aggression regulation and/or a stronger approach motivation. Such studies are anyway difficult to extrapolate to reactive aggression, which should be mostly determined by brain *reactivity* to threat or provocation than by activity at rest. Hence, reactive aggression has been often investigated with event-related potentials (ERP), i.e. changes in electroencephalographic activity in response to specific stimuli. ERP studies with the TAP have shown that an increased frontal negativity in response to provocation is associated with the ability to suppress aggressive impulses, both in healthy

participants and in young violent offenders (Krämer, Büttner, Roth, & Münte, 2008; Wiswede et al., 2011). This effect is thought to arise from medial and lateral prefrontal regions, perhaps reflecting subjective effort to downregulate aggression. In another EEG study with a variant of the TAP, non-aggressive individuals demonstrated an increased amplitude of the P3 component in frontal electrodes when receiving infrequent loud sound blasts, but this effect was not observed in aggressive individuals (Fanning, Berman, & Long, 2014). Since frontal P3 amplitude is enhanced when uncommon or unexpected stimuli appear (Verleger, Jaskowski, & Wauschkuhn, 1994), this result suggests that aggressive individuals expect to be provoked and therefore show little or no surprise if this happens. Complementing these results, aggressive individuals show an enhanced mismatch negativity over temporal and parietal electrodes when hostile expectations are violated (Gagnon, Aubin, Emond, Derguy, Bessette, et al., 2016), and the extent of this negativity is related to trait measures of hostile attributional bias (Gagnon, Aubin, Emond, Derguy, Brochu, et al., 2016). ERP studies thus seem to confirm that the tendency to attribute hostile intentions to others is manifested in reduced frontal brain responses when aggression happens unexpectedly, whereas it is increased in temporal and parietal areas when aggression can be expected but does not ensue. Hence, aggressive individuals might have a default readiness for aggression, and, once confrontation starts, they require a stronger engagement of prefrontal regions to suppress violent impulses.

1.5.8. Positron emission tomography studies

EEG studies offer a very precise temporal resolution, but the source of activation is hard to determine. Furthermore, EEG measured at the scalp cannot directly provide information on subcortical activity. Positron emission tomography (PET), a technique that quantifies concentrations of a given radioactive tracer by gamma ray emission detection, has increased

spatial resolution and permits to determine the amount of specific chemicals across the brain. Generally, PET studies have found reduced relative glucose metabolic rate (rGMR) in the prefrontal cortex of violent offenders, particularly in the medial aspect of this region (Raine, et al., 1997; Raine, et al., 2000; Volkow et al., 1995). In this vein, another PET investigation found that aggressive epileptic children had lower rGMR in the mPFC relative to non-aggressive epileptic controls (Juhász, Behen, Muzik, Chugani, & Chugani, 2001). Similarly, depressive patients with anger attacks show reduced OFC activation than healthy controls during script-induced anger (D. D. Dougherty, Rauch, Deckersbach, & et al., 2004). A PET study using the PSAP showed that individuals with a dual diagnosis of BPD and IED have exaggerated rGMR in the amygdala after provocation (New et al., 2009). Importantly, these patients do not differ from controls in amygdala rGMR at rest, though they do display negative baseline coupling between the amygdala and the OFC (New et al., 2007). Other PET-based research has localized functional abnormalities related to serotonin in aggressive individuals (see also section 1.5.11. Serotonin and reactive aggression). These studies have revealed that low serotonin concentrations in the cingulate cortex are related to lower aggression (da Cunha-Bang et al., 2016; Frankle et al., 2005; though see van de Giessen et al., 2014 for a failure to replicate this effect). Conversely, impulsive-aggressive patients with a diagnosis of personality disorder show blunted metabolic responses to a serotonin receptor agonist in ventrolateral prefrontal and cingulate areas (New, Hazlett, Buchsbaum, & et al., 2002), and patients with comorbid BPD and IED show greater orbitofrontal glucose blood flow after serotonin antagonist administration (New et al., 2004). Overall, PET studies tend to agree with the animal studies reviewed earlier (i.e. enhanced amygdala reactivity) as well as with human structural studies (i.e. reduced PFC engagement), and further suggest that serotonin might underlie some of the neuro-functional features associated with aggression.

1.5.9. Functional magnetic resonance imaging studies

Since the 1990s, human brain function can be also investigated with functional magnetic resonance imaging (fMRI), which measures oxygen uptake in the brain with an excellent spatial resolution. Unlike PET, fMRI is completely non-invasive, and also offers enhanced temporal resolution. Some researchers have used fMRI to identify neurocognitive processes that might influence aggression indirectly (e.g., response inhibition), and/or by comparing neural responses of aggressive and nonaggressive participants to threat or provocation. This work has revealed that individuals with high scores in trait aggression show reduced activity in the supplementary motor area (SMA) when withholding motor responses (Pawliczek, Derntl, Kellermann, Kohn, et al., 2013), as well as in dlPFC and vlPFC during a frustrating anagram-solving task (Pawliczek, Derntl, Kellermann, Gur, et al., 2013). Complementing the latter finding, increasing frustration leads to activation of the core aggression network (R. Yu, Mobbs, Seymour, Rowe, & Calder, 2014). Interestingly, functional activations related to commission errors during response inhibition and reactivity to provocation overlap in the insula (Dambacher et al., 2014). Studies in reactive-aggressive clinical samples have generally found increased OFC activation during response inhibition (Soloff, White, Omari, Ramaseshan, & Diwadkar, 2015; Völlm et al., 2004), as well as increased OFC-amygdala coupling (Soloff, Abraham, Ramaseshan, Burgess, & Diwadkar, 2017). OFC activity during the presentation of negative facial expressions in an affective response inhibition task has been further related to lower lifetime aggression in BPD patients (Soloff, Abraham, Burgess, et al., 2017). Hence, the increased medial OFC response during effortful cognitive tasks in aggressive individuals might represent compensatory activity.

Within the subcortical aggression network, the amygdala is the region most consistently associated with reactive aggression in response to threat signals (Coccaro, et al., 2011). Oppositely, psychopathy is related to amygdala hypoactivation when imagining or seeing

others in pain (Decety, Chen, Harenski, & Kiehl, 2013; Marsh et al., 2013) and in response to emotional faces (Carré, Hyde, Neumann, Viding, & Hariri, 2013; Hyde, Byrd, Votruba-Drzal, Hariri, & Manuck, 2014). This suggests that proactive and reactive aggression can be distinguished by partly opposing patterns of brain activity. In line with this, pathologically aggressive youths with low psychopathic traits show heightened amygdala and PAG reactivity to provocation as compared to those with high psychopathic traits and healthy controls (White et al., 2015). Similarly, decoupling between the amygdala and the OFC during the presentation of angry faces has been observed in IED patients (Coccaro, McCloskey, Fitzgerald, & Phan, 2007; McCloskey et al., 2016) and in participants with high approach motivation –a predictor of aggression (Passamonti et al., 2008). Reduced coupling between these regions correlates as well with heightened negative affect when watching aversive pictures (Banks, Eddy, Angstadt, Nathan, & Phan, 2007). Resting-state functional connectivity studies, which measure temporal correlation between remote brain regions at rest, largely mimic these findings. Amygdala-OFC decoupling at rest has been observed in BPD patients (Baczkowski et al., 2016; New, et al., 2007) and in children with ADHD (Ho et al., 2015), and is associated with self-reported aggression in male schizophrenia patients (Hoptman et al., 2010), as well as with psychopathic traits in children with conduct disorder (Aghajani et al., 2016).

Most of the studies just reviewed used tasks that are only indirectly related to aggression (e.g. frustrating tasks, response inhibition tasks, watching angry faces), or compared brain reactivity to provocation between aggressive and non-aggressive groups. In order to better capture interindividual differences in aggression, there has been increasing interest in adapting the interactive tasks presented in section 1.3 to the fMRI scanner. This in turn permits to establish direct relationships between brain activity and reactive aggression. In agreement with the literature reviewed so far, these studies have generally confirmed that

aggression is associated with increased amygdala reactivity (Chester & DeWall, 2016; Gospic et al., 2011; Lotze, Veit, Anders, & Birbaumer, 2007), and with lower OFC reactivity to threat or provocation (Beyer, Münte, Göttlich, & Krämer, 2015; Mehta & Beer, 2009; White, et al., 2014). Nonetheless, medial OFC activation has also been observed when punishing unfair rivals (Lotze, et al., 2007; Singer et al., 2006). This, together with the known role of this region in reward valuation (Liu, Hairston, Schrier, & Fan, 2011), suggests that the function of the OFC in aggressive interactions might depend on the organism's predominant motivational drive, i.e. to suppress or act out the aggressive impulse. The dlPFC also plays a role in reducing reactive aggression, as participants with stronger dlPFC activation in response to negative social feedback select shorter sound blasts against the purported co-player (Achterberg, van Duijvenvoorde, Bakermans-Kranenburg, & Crone, 2016), and individuals with a history of alcohol dependence are more aggressive and display reduced reactivity to provocation in the dlPFC when playing the PSAP (Kose et al., 2015). The latter results are in agreement with some of the brain stimulation studies commented above. This line of research also suggests a role for the mPFC in aggressive interactions. In the study by Achterberg et al., both positive and negative social feedback elicited mPFC activation, and heightened mPFC reactivity to provocation has been observed in the decision phase of the TAP (Krämer, Jansma, Tempelmann, & Münte, 2007; Lotze, et al., 2007). The mPFC belongs to a set of brain regions typically recruited during social cognition which includes the precuneus, the IFG, or the temporal-parietal junction (TPJ) and is often referred to as the *mentalizing network* (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). Hence, mPFC activity in the TAP has been interpreted as reflecting social-cognitive processes (Krämer, et al., 2007). Indeed, the mPFC seems to be upregulated in response to provocation in aggressive individuals (Lotze, et al., 2007), whereas it is less strongly recruited in participants with high emotional reactivity to threat (Beyer, Münte, Erdmann, & Krämer, 2013). These results would

simplistically suggest that mPFC recruitment is associated with higher aggression, presumably by favoring a better understanding of the opponent's intentions. Nevertheless, it remains to be seen whether mentalizing areas can favor alternative behavioral strategies (e.g. negotiation, avoidance) if these are available.

Neuroimaging studies employing competitive paradigms have also implicated the reward system in aggression, mimicking animal and human structural studies. The ventral striatum (VS), a core reward-processing area entailing the nucleus accumbens and the head of the caudate nucleus, shows increased activity when winning relative to losing (Brunnlieb, Münte, Krämer, Tempelmann, & Heldmann, 2013; Krämer, et al., 2007). Reactivity to provocation in this region has been related to increased aggression (Beyer, et al., 2013; Chester & DeWall, 2016). Similarly, alcohol-induced VS activity has been related to reactive aggression in the TAP (Gan, Sterzer, Marxen, Zimmermann, & Smolka, 2015). An fMRI study using the PSAP found that individuals with IED symptoms had heightened striatal activation when obtaining monetary rewards relative to retaliation, and the extent of this activation correlated with the amount of times they pressed the reward button (Gan, et al., 2016). However, aggressive participants did not use the aggressive option more often than controls, nor was VS activation related to aggression. This could be due to the conflict between monetary and retaliatory rewards in the PSAP, i.e. that pressing the reward button offers slower but safer rewards than the aggressive option, and hence might be highly attractive for approach-driven individuals. A recent study has limited this confound by not allowing participants to keep the subtracted money, thus rendering the subtraction option unambiguously aggressive (da Cunha-Bang, MacDonald Fisher, et al., in press). In that study, violent offenders showed increased reactivity to provocation than controls in amygdala and VS, and the extent of the provocation effect correlated with aggression across both groups. Related to this point, and as commented in the previous paragraph, the medial OFC is

recruited when punishing unfair co-players, which is often considered a reward-driven effect (Seymour, Singer, & Dolan, 2007). Specifically, it has been proposed that the VS and the medial OFC might jointly encode the gratification derived from punishing those who violate social norms such as reciprocity (Seymour, et al., 2007). Taken together, these results confirm the relationship between reward sensitivity and reactive aggression, they agree with the striatal enlargement observed in structural studies and with the enhanced dopaminergic signaling seen in mice, and fit the notion of anger and aggression as approach-driven phenomena. Nonetheless, the circumstances under which the mPFC and the OFC lead to aggression are yet to be fully elucidated. Further interactive paradigms that more clearly dissect the cognitive processes underlying aggression are needed to settle this issue.

1.5.10. Genetics of aggression

Genetic studies suggest that aggression is at least partly heritable, to an extent comparable with personality traits and other complex phenotypes (Waltes, Chiocchetti, & Freitag, 2016). However, genetic factors only account for ≈50% to ≈80% of variance in aggressive behavior (Porsch et al., 2016), and, more importantly, no single gene seems to consistently predispose to aggression by itself (Vassos, Collier, & Fazel, 2014). Rather, as most complex phenotypes, the relative effect of genes on aggression is dependent on environmental input. Many polymorphisms (i.e. gene variants) such as the Catechol-O-methyltransferase (COMT) or the serotonin transporter (5HTTLPR) gene have been investigated in the context of aggression (Vassos, et al., 2014). However, across multiple studies, the monoamine oxidase A (MAOA) gene, located in the X chromosome, is the polymorphism most robustly associated to aggression in interaction with environmental stress (Byrd & Manuck, 2014; Ficks & Waldman, 2014). This gene codes for the homonymous enzyme that catabolizes norepinephrine (NE), dopamine (DA), and serotonin (5-HT). The amount of repetitions of a

specific amino acid sequence within this gene determines relative MAOA activity, with individuals being usually classified as carrying the low (MAOA-L) or high (MAOA-H) activity variants of the gene. The MAOA-L variant seems to confer greater vulnerability to adverse life events, such that MAOA-L carriers are more likely to become aggressive in response to interpersonal insults or maltreatment, presumably because of lower tonic levels of the enzyme in the brain (Buades-Rotger & Gallardo-Pujol, 2014). This is corroborated by several findings. Perinatal disruption of MAOA activity increases aggression during adolescence in mice (Q. Yu, et al., 2014). In humans, lower brain MAOA activity correlates with trait aggression (Alia-Klein et al., 2008), violent offenders with a diagnosis of ASPD have lower MAOA concentrations in OFC and other prefrontal areas (Kolla et al., 2015), and amygdala activity in a face matching task correlates with aggression in MAOA-L men (Holz et al., 2016).

1.5.11. Serotonin and reactive aggression

Among all monoamines catabolized by the MAOA enzyme, serotonin is arguably the most studied neurotransmitter in the context of aggression. 5-HT neurons originate in the raphe nuclei of the brainstem with rostral nuclei projecting to the forebrain and caudal nuclei projecting downstream to the spinal cord (Hornung, 2003). The known involvement of 5-HT in affect regulation has bolstered much investigation on its possible role in aggression. Since increasing extracellular concentrations of 5-HT generally improves mood, serotonin could also buffer frustration- or provocation-induced aggressive impulses (Glick, 2015). Indeed, many studies have found that higher 5-HT concentrations relate to lower aggression, so much so that the negative relationship between 5-HT levels and aggression has been called “perhaps the most reliable finding in the history of psychiatry” (Fishbein, 2001; as cited in Duke, Bègue, Bell, & Eisenlohr-Moul, 2013). Meta-analytic evidence has anyway questioned this

assumption: across studies, the relationship between 5-HT concentrations and aggression is small in size, and strongly affected by methodological variability. Specifically, the relationship is weaker in correlational studies, and strongest in pharmacological studies (Duke, et al., 2013). A further complication is that pharmacological manipulations in humans alter brain-wide availability of 5-HT, and so effects observed in-vivo might be directly due to serotonergic action or a collateral effect of the drug. This suggests the possibility that the effect of serotonin on aggression is dependent on the specific site of action. A recent study in mice has identified a specific subset of serotonergic populations in the raphe nuclei that can increase aggressive behavior when silenced (Niederkofler et al., 2016). One of these neuronal populations projects to medial prefrontal areas, whereas the other rather innervates sensory and associative regions. Both of these neuronal populations have reciprocal axonal connections with midbrain dopaminergic sites, implying a possible interplay between circulating DA and serotonin-mediated aggression. One fMRI study showed that acute tryptophan depletion (ATD, i.e. the transient reduction of central 5-HT) increases rejection of unfair offers in the UG, an effect that concurred with higher dorsal striatum activity (Crockett et al., 2013). Another fMRI study with the TAP oppositely showed that ATD *reduced* aggression in low trait-aggressive participants, and generally dampened insula activity during the decision phase (Krämer, Riba, Richter, & Münte, 2011). Since the insula is also involved in retaliatory behavior (Emmerling et al., 2016), the aforementioned results suggest that normal 5-HT levels might indeed be necessary for reactive aggression in healthy subjects, and further question a simple negative link between serotonin and aggression. Importantly, the effects of serotonin on aggression-related brain regions seem to be mostly dependent on 5-HT1A and B receptors (da Cunha-Bang, Hjordt, et al., in press; Parsey et al., 2002; Witte et al., 2009), but not 5-HT2A (da Cunha-Bang et al., 2013), which narrows down putative targets for serotonergic pharmacological interventions.

1.5.12. Dopamine and reactive aggression

Other neurotransmitters aside from 5-HT are involved in the genesis of aggression, most notably dopamine. DA is produced in the substantia nigra and ventral tegmental area (VTA) of the brainstem, and its mesolimbic pathway is thought to mediate the initiation and maintenance of motivated behavior (Bromberg-Martin, Matsumoto, & Hikosaka, 2010). DA signaling is enhanced in this pathway during aggressive confrontations in rats (Anstrom, Miczek, & Budygin, 2009). In the study by Yu et al. (2014) cited earlier, blockade of the 5-HT transporter during adolescence reduced aggression, whereas blocking the DA transporter increased it. In line with this, mice selected for high aggressiveness show increased prefrontal dopamine turnover, but decreased serotonin, after repeated aggressive encounters with an intruder (Caramaschi, et al., 2008). Importantly, they also show increased aggressiveness over time, suggesting that repeatedly experiencing victory might reinforce aggressive behavior by plastically changing prefrontal neurotransmission (Caramaschi, et al., 2008). Indeed, 5-HT levels measured *in vivo* in the mPFC decrease after aggressive confrontations (van Erp & Miczek, 2000). This effect is accompanied by increased DA concentrations in the nucleus accumbens, which rise during, and peak after an aggressive encounter (Ferrari, et al., 2003). However, as in the case of 5-HT, the effect of DA on human aggression is not so straightforward. In a PET study in humans, MAOA-H subjects showed higher dopamine release in the striatum than their MAOA-L counterparts when watching a violent movie (Schlüter et al., 2016). This is counterintuitive since MAOA-H carriers should have *less* circulating dopamine. Also puzzling is another PET study of the PSAP in healthy men, which found that dopamine synthesis capacity in the midbrain and striatum was related to *lower* aggression (Schlüter et al., 2013). These results suggest a complex interplay of DA and 5-HT on aggression-related brain centers, but highlight the involvement of the reward-processing circuitry in aggression.

1.5.13. Hormonal influences on aggression

Aggression is also subject to endocrine influences. The steroid hormones cortisol (C) and testosterone (T) have been long considered to play a particularly prominent role in this regard (Carré, McCormick, & Hariri, 2011). C is synthesized in the adrenal glands and secreted in a cascading process initiated by adrenocorticotropin-releasing hormone (ACTH) release in the hypophysis. C peaks one hour after awakening and increases in stressful situations, exerting a stimulating effect characterized by increases in alertness and muscular tone (Sapolsky, Romero, & Munck, 2000). T, on the other hand, is synthesized from cholesterol after conversion to several precursors in the testicles or ovaries and has a major role in the maturation of the reproductive system (McHenry, Carrier, Hull, & Kabbaj, 2014). T and C have a mutually opposing action. For instance, glucocorticoids can inhibit T action by down-regulating transcription of androgen receptors (K. L. Burnstein, Maiorino, Dai, & Cameron, 1995). Conversely, T can suppress pharmacologically-induced C secretion (Rubinow et al., 2005), and ACTH output is partly conditional on plasma T levels (Viau, 2002).

T by C (henceforth TxC) interactions can also be observed at the behavioral level, with T generally favoring approach and dominance, and C favoring fear and avoidance (Montoya, Terburg, Bos, & van Honk, 2011). In competitive contexts, participants with high T levels before a competition show higher C levels after losing, but lower C when winning (Mehta, Jones, & Josephs, 2008). Comparably, high-C men show decreases in T after winning a competitive task by a narrow margin (Wu, Eisenegger, Zilioli, Watson, & Clark, 2016), and individuals with low C and high T overbid more in competitive auctions (van den Bos, Golka, Effelsberg, & McClure, 2013). Similarly, men with greater increases in T after winning perform better in a subsequent competitive task (Zilioli & Watson, 2014).

TxC interactions have also been observed specifically in the context of aggression. In men, a pattern of high T and low C is generally related to increased aggressive behavior

(Carré, et al., 2011; Carré & Mehta, 2011; Montoya, et al., 2011). Fitting these results, individuals with high T and low C have lower self-reported empathy (Zilioli, Ponzi, Henry, & Maestripieri, 2014) and increased psychopathic traits (Glenn, Raine, Schug, Gao, & Granger, 2011). Competition-induced changes in T also correlate with subsequent aggressive behavior in both men (Carré, Campbell, Lozoya, Goetz, & Welker, 2013; Carré, Putnam, & McCormick, 2009) and rodents (Clinard, Barnes, Adler, & Cooper, 2016; Oyegbile & Marler, 2005). In women, however, a combination of high cortisol and high testosterone has been reported in BPD patients (Rausch et al., 2015), and has been related to reactive aggression (Denson, Mehta, & Ho Tan, 2013). This is in line with a study showing that C administration increases aggression in women (Böhnke, Bertsch, Kruk, Richter, & Naumann, 2010), and that higher C concentrations correlate with self-reported anger and aggressiveness in female BPD patients (Rausch, et al., 2015). These gender differences could be explained by women's relatively increased sensitivity to social threat (Else-Quest, Higgins, Allison, & Morton, 2012), whereby higher C could lead to exaggerated emotional reactions manifested in aggressive behavior.

The findings above commented seem to confirm the pro-aggressive role of testosterone. Multiple sources of evidence from the animal realm are also in agreement with this postulate, as aggressive and dominant male chimpanzees have increased T concentrations (Muller & Wrangham, 2004; Sapolsky, 1982), especially in unstable hierarchies (Sapolsky, 1991). Concordantly, testicle removal decreases aggression in some mammal species (Dixson, 1980; Lofgren et al., 2012; Sapolsky, 1982), and there is also evidence for a positive relationship between T and aggression in female non-human primates (Beehner, Phillips-Conroy, & Whitten, 2005). Given that men are more physically aggressive than females (Bettencourt & Miller, 1996) and have higher average testosterone concentrations (Keevil, MacDonald, Macdowall, Lee, & Wu, 2013), the T-aggression link has also been taken to be valid in

humans. Meta-analytic evidence does support a significant correlation between T levels and aggression, but it is much weaker than what could be expected from animal studies (Archer, Graham-Kevan, & Davies, 2005). Recent evidence has further showed that the relationship between T and aggression in humans is more nuanced than previously thought, such that T can promote prosocial behavior if that is beneficial for one's status (Eisenegger, Haushofer, & Fehr, 2011; Eisenegger, Naef, Snozzi, Heinrichs, & Fehr, 2010). In humans, status can be achieved by other means other than brute force. Since societies often punish antisocial behavior and reward prosocial behavior, it is expectable that some individuals endorse the latter over the former in order to improve their status. Indeed, the relationship between T and aggression appears to be *negative* in interdependent individuals, i.e. those that define themselves in relation to others and hence place more value on social bonds (Welker et al., 2016). Dovetailing this finding, an inverse relationship between self-reported aggression and T has been reported in women (Gladue, 1991), who generally score higher in measures of interdependence (Cross & Madson, 1997).

Most studies that have questioned the simplicity of the T-aggression relationship in humans have adopted a pharmacological approach. This line of research has shown that T lowers the perceived trustworthiness of faces (Bos, Terburg, & van Honk, 2010) but increases reciprocity in an economic exchange paradigm (Boksem et al., 2013). In a simplified poker game, T administration minimized cold bluffing, but increased calling (van Honk et al., 2016). In a study employing the UG, women who received T made fairer offers than those who received placebo (Eisenegger, et al., 2010). Even so, those who believed they had received T made more unfair offers, exemplifying the persistence of folk beliefs regarding the role of T. In another UG study in which subjects were the responders, participants who had been given T reported reduced anxiety and aggression, and accepted a tangentially higher proportion of unfair offers (Kopsida, Berrebi, Petrovic, & Ingvar, 2016). In a modified

version of the UG in which subjects could administer additional rewards and punishments to the proposers, T increased punishment after unfair offers, but also rewards after generous ones (Dreher et al., 2016). Summing up, novel evidence suggests that T might make participants initially more cautious and subsequently more reciprocal in social interactions, ultimately potentiating status-enhancing behavior. In women, who are generally more empathic (Christov-Moore et al., 2014) and less physically aggressive (Nivette, et al., 2014), T might have a stronger net prosocial effect than in men.

1.5.14. Hormonal influences on aggression-related neural structures

Both C and T act upon the amygdala, a structure rich in both glucocorticoid (Wang et al., 2013) and androgen receptors (Cunningham, Lumia, & McGinnis, 2012). Amygdala reactivity to threat has been linked to acute increases in cortisol secretion (Henckens et al., 2016; Mujica-Parodi, Carlson, Cha, & Rubin, 2014), linking with the pro-aggressive effects of state C earlier reviewed. On the other hand, individuals with higher basal cortisol levels tend to have a less reactive amygdala, as this structure presumably becomes desensitized after sustained stress exposure (Cunningham-Bussel et al., 2009; Henckens, et al., 2016). In line with this, prolonged glucocorticoid intake leads to reductions in amygdala volume (Brown, Woolston, & Frol, 2008). Hence, evidence suggests that cortisol reactivity to threat is associated with relatively increased aggression (especially in women), whereas basal cortisol is related to fear and avoidance. Indeed, persons with anxiety disorders and other pathologies characterized by high levels of fear and avoidance usually display exaggerated basal HPA activity (Elnazer & Baldwin, 2014). Importantly, however, the effect of cortisol in the brain seems to vary depending on threat escapability, such that cortisol decreases midbrain-dependent fear reactions in inescapable threat situations, whereas it increases activity in the

saliency network (ACC and AI) when danger can be avoided (Montoya, van Honk, Bos, & Terburg, 2015).

Regarding the central effects of testosterone, there is consistency across studies showing that T administration heightens amygdala reactivity to angry faces in both men and women (Bos, Panksepp, Bluthé, & Honk, 2012). Single-dose pharmacological studies permit to make causal claims on hormonal effects, but interactions between the substance and individuals' baseline hormonal levels cannot be ruled out. In an fMRI study in men, researchers circumvented this limitation by suppressing baseline differences in circulating T levels and then administering exogenous T, achieving an increase in T levels within the normal range (Goetz et al., 2014). T increased reactivity to angry faces in amygdala, hypothalamus, and PAG. The relationship between endogenous T and amygdala reactivity is however more obscure, as some studies found positive correlations in men (Derntl et al., 2009; Manuck et al., 2010), one study found an inverse correlation in men (Stanton, Wirth, Waugh, & Schultheiss, 2009) and others found none in women (Ackermann et al., 2012; Stanton, et al., 2009). A recent fMRI offers a possible explanation for these discrepancies, such that T administration increased amygdala reactivity to angry faces when they had to be approached, but *decreased* it when they had to be avoided (Radke et al., 2015). This indicates that the effect of T on the amygdala might vary upon the organism's motivational state, flexibly potentiating approach- or avoidance-like behaviors.

1.6. Outline of the thesis

In the present thesis, we draw on the GAM to test the interactive role of several neurobiological and situational variables involved in the development of aggression episodes. We used different variants of the Taylor Aggression Paradigm in order to provoke participants and trigger aggression. In each study we focused on a specific stage of aggression episodes,

introducing situational modifications in the task that allowed us to answer specific research questions.

In chapter 2, we investigated whether trait and state estimates of testosterone and cortisol impact brain reactivity to threat signals, and how this influences reactive aggression in healthy young women. That is, we investigated the interplay between presumably trait-like biological factors and transient changes in internal state in relation to aggression.

In chapter 3, we developed and validated the Fight-or-Escape paradigm (FOE), a novel task based on the TAP in which participants are given the chance to avoid confrontation. More specifically, we examined whether approach and avoidance tendencies, as measured by separate experimental tasks, predicted avoidant behavior in the FOE.

In chapter 4, we translated the FOE to the scanner to explore the neural correlates of aggressive and avoidant responses to social provocation. In the parlance of the GAM, in this and the previous chapter we investigated appraisal and decision processes preceding the aggressive encounter proper.

In chapter 5, we explored how the outcome of an aggression episode impacts ventral striatum connectivity, and how this relates to aggressive behavior. We specifically compared two positive results: winning to punish the opponent versus winning to avoid punishment only.

In chapter 6 we summarize and critically discuss our findings in relation to the theoretical and empirical literature reviewed in this Introduction.

Chapter 2: Endogenous testosterone is associated with lower amygdala reactivity to angry faces and reduced aggressive behavior in healthy young women[1]

2.1. Introduction

Reactive aggression is a phylogenetically ancient behavior by which organisms respond to threat or provocation with an overt intent to harm the attacker (Nelson & Trainor, 2007). Reactive-aggressive impulses are thought to arise from subcortical structures such as the amygdala and the periaqueductal gray, which are in turn regulated by the orbitofrontal cortex (OFC) and other prefrontal regions (Coccaro, et al., 2011). The steroid hormones testosterone (T) and cortisol (C) have been suggested as important factors for the regulation of reactive aggression, and have been shown to act on the aforementioned brain areas by binding to androgen and glucocorticoid receptors, respectively (Carré, et al., 2011; Carré & Mehta, 2011; Carré & Olmstead, 2015). Both hormones are considered to have a mutually opposing action (Rivier & Rivest, 1991; Rubinow, et al., 2005; Viau, 2002), and this divide is also apparent in human social behavior: T generally favors approach and aggression, while C leads to fear and avoidance (Montoya, et al., 2011; van Honk, Harmon-Jones, Morgan, & Schutter, 2010). This notion is supported by numerous findings in correlational (Dabbs & Hargrove, 1997; Platje et al., 2013; Popma et al., 2007; Poustka et al., 2010) and experimental settings (Carré, et al., 2009; Gordis, Granger, Susman, & Trickett, 2006; Josephs, Sellers, Newman, & Mehta, 2006), though not all evidence agrees. For instance, some studies in women have linked high, rather than low, C concentrations with aggression (Denson, Mehta, et al., 2013; Rausch, et al., 2015), and high testosterone with prosocial behavior (Boksem, et al., 2013; Eisenegger, et al.,

[1] This chapter corresponds largely to: Buades-Rotger, M., Engelke, C., Beyer, F., Keevil, B. G., Brabant, G., & Krämer, U. M. (2016). Endogenous testosterone is associated with lower amygdala reactivity to angry faces and reduced aggressive behavior in healthy young women. *Scientific Reports, 6*, 38538. doi: 10.1038/srep38538. The study was conceived by FB and UMK. I designed the study, performed the experiments, analyzed the data, and wrote the manuscript.

2010; van Honk, et al., 2016). Moreover, some controversy remains regarding whether state T and C predict aggression better than baseline values (Geniole, Carré, & McCormick, 2011). Hence, the dynamics of the relationship between these hormones and aggression are still unclear, especially in women.

Several functional magnetic resonance imaging (fMRI) studies have investigated how C and T affect the activity of brain regions involved in aggression. It has been suggested that amygdala activation in response to threat signals relates to *lower* basal C (Cunningham-Bussel, et al., 2009; Henckens, et al., 2016), but *higher* C reactivity (Henckens, et al., 2016; Mujica-Parodi, et al., 2014). On the other hand, endogenous T in men has been positively associated with amygdala reactivity to threat (Derntl, et al., 2009; Manuck, et al., 2010; but see Stanton, et al., 2009), as well as with lower OFC engagement during emotion-contingent motor control (Volman, Toni, Verhagen, & Roelofs, 2011). In women, exogenous T seems to robustly increase amygdala reactivity to angry faces (Bos, et al., 2012; Hermans, Ramsey, & van Honk, 2008; van Wingen et al., 2008), but the effects of endogenous T are less clear. For instance, van Wingen et al. (2008) found a positive relationship between T concentrations and amygdala reactivity to angry faces in women, while others have found none (Ackermann, et al., 2012; Stanton, et al., 2009). In a study using a mixed sample, endogenous T was associated to higher aggression and lower OFC activity in response to provocation (Mehta & Beer, 2009). Few fMRI studies have tested the joint role of both hormones in the context of aggression. Denson and colleagues (2013) found that T to C ratio (T:C) was linked to higher amygdala coupling with medial and dorsolateral prefrontal regions after an interpersonal provocation in males, whereas Hermans et al. (2008) reported positive relationships between T:C and hypothalamus, brainstem, and amygdala reactivity to angry relative to happy faces in women. Overall, the evidence suggests that a basal pattern of high T and low C should lead to

subcortical hyper-reactivity along with decreased ventral-prefrontal recruitment in the presence of threat or provocation.

Angry facial expressions are a common means to communicate hostility (Reed, DeScioli, & Pinker, 2014) and are thus often used as an experimental proxy of social threat. Nevertheless, angry faces as employed in most of the commented studies rarely signal a real menace, because they are usually presented as static, context-devoid stimuli. Hence, they likely elicit briefer, weaker, or even qualitatively different neural responses than those occurring in real-world aggressive interactions. It is therefore desirable to use more realistic tasks in which angry facial expressions can be interpreted as communicative signals and neural responses to these signals can be directly related to aggressive behavior. In a previous study from our group, we developed a competitive reaction time task based on the Taylor Aggression Paradigm (TAP), which we here named Social Threat Aggression Paradigm (STAP). In the STAP, participants see a short video of their rival displaying either an angry or a neutral facial expression right before they choose the volume of an aversive sound blast to be directed at their opponent. In a sample of healthy young men, we observed that OFC reactivity to angry relative to neutral faces was negatively related to aggressive behavior (i.e. less intense sound blasts; Beyer, et al., 2015). This finding fits studies showing that intermittent explosive disorder (IED) patients (Coccaro, et al., 2007) and individuals scoring high in approach motivation (Beaver, Lawrence, Passamonti, & Calder, 2008) have reduced OFC reactivity to static angry faces. Importantly, however, we also found widespread brain activation in the last phase of angry compared to neutral trials, when no facial expression was present. This suggests that contextualized angry faces frame the processing of subsequently presented information (Reed, et al., 2014) and have a longer-lasting effect than in other paradigms.

The present work had a double aim. Firstly, we investigated whether reactivity to angry faces and aggression in a direct social interaction are modulated by basal and/or acute levels of endogenous T and C. Secondly, we tested whether the findings of our previous study could replicate in a female sample. We hypothesized that morning C would be negatively related to reactive aggression, as has been shown across previous correlational studies (van Goozen, Fairchild, Snoek, & Harold, 2007), as well as in an experimental aggression study with the TAP (Böhnke, Bertsch, Kruk, & Naumann, 2010). As to T, we predicted that it would be positively associated with aggression generally, and that T levels in the afternoon would be more strongly related to aggression than in the morning (Archer, et al., 2005). Furthermore, we expected to find a competition-induced increase in both T and C (Bateup, Booth, Shirtcliff, & Granger, 2002; Kivlighan, Granger, & Booth, 2005). Regarding the effects of these hormones on neural structures, following the reviewed evidence we hypothesized that T would be linked to heightened amygdala reactivity to angry faces, reduced OFC activation, and a weaker coupling between these regions. We also expected to find a positive association between basal C and amygdala reactivity, whereas the pre-post task change in C should follow the opposite pattern.

We recruited 39 healthy female participants who provided saliva samples at home during a regular weekday, plus before and after the fMRI measurement. In addition, we obtained a pre-scan blood sample for serum T estimation. We controlled for momentary and monthly fluctuations in hormonal concentrations by having participants come during the early follicular phase of the menstrual cycle and in the afternoon (12:00-18:00), and performed the anatomical measurement on a previous day to limit scanner-related arousal. Before the functional measurement participants were introduced to their purported female opponent, a 20-year old confederate. Once in the scanner, participants played three 20-trial runs of the STAP, which was actually preprogrammed (see Methods and Fig. 2.1A). We probed whether

participants had been provoked (i.e. if they selected louder sound blasts over time and in angry trials), and inspected for correlations between mean aggression and hormonal concentrations in the morning and at scan-time. We then tested for differences in neural reactivity to angry vs neutral faces in the decision phase (i.e. when participants decided the intensity of the punishment for their opponent) and the outcome phase (i.e. when participants were informed of whether they won or lost). Further, we examined whether aggression-related hormones and within- and between-participant variability in aggressive behavior could modulate brain reactivity to angry facial expressions. Finally, we investigated the interplay of neural and endocrine factors on aggression with mediation analyses.

2.2. Methods

2.2.1. Participants

Female participants were recruited from the university via e-mails and flyers. Exclusion criteria were present or past endocrine, psychiatric, or neurologic disorder, and use of hormonal contraceptives, which reduce circulating T levels (Zimmerman, Eijkemans, Coelingh Bennink, Blankenstein, & Fauser, 2013) and alter C responsivity (Kirschbaum, Kudielka, Gaab, Schommer, & Hellhammer, 1999). Out of an initial sample of 43 participants, four participants were excluded from the analyses, three of them because they guessed that the paradigm was preprogrammed, and another because she was diagnosed of hyperthyroidism after the functional measurement had been performed. The final sample thus comprised 39 young female college students (mean age = 23.22, SD = 3.2), all right handed and fluent German speakers. This study was approved by the University of Lübeck ethics committee and performed according to the Declaration of Helsinki. Participants and confederate provided informed consent and received economic compensation.

2.2.2. Procedure

We scheduled two appointments per participant. On the first one, only the anatomical scan was recorded. We did this to familiarize participants with the scanner environment, so that C levels in the second appointment were not influenced by situational features (scanner noise, novelty, etc. Weldon et al., 2015). We also gave participants plastic tubes to provide saliva samples during a regular weekday (see "Saliva and blood collection").

The functional measurement was performed on the second appointment, which was always scheduled in the afternoon (i.e. between 12 and 18) and in the first seven days after the start of the menses. We thereby aimed to mitigate circadian (Keevil, et al., 2013) and monthly (Bui et al., 2013) fluctuations in hormonal levels and in the processing of emotional facial expressions (Derntl et al., 2008). Upon arrival, we extracted an antecubital venous blood sample (see "Saliva and blood collection"). We then introduced participants to their purported opponent, a 20 year-old female confederate who arrived 5-10 minutes later than the participant. Participant and confederate read the instructions for the STAP, and they were told that the participant could see the confederate at the beginning of each trial via webcam. The confederate behaved neutrally (i.e. avoided being friendly towards the participant) and made a casual question ("Will I play with headphones or speakers?") to make the setting more believable. The confederate was allegedly accompanied to another room to prepare for the task (she actually left) while participants provided another saliva sample. In the scanner, participants underwent a 6-minute resting-state measurement, after which the STAP started (≈30 minutes). Another 6-minute resting-state measurement was performed after the task. Once out of the scanner, participants provided an additional saliva sample and fulfilled an ad-hoc questionnaire to assess whether they had been successfully deceived. They also filled out the German versions of the Behavioral Inhibition/Approach System (BIS-BAS) scale (Strobel, Beauducel, Debener, & Brocke, 2001), and of the Aggression Questionnaire (AQ; Herzberg,

2003) to control for personality variables putatively associated with aggression and/or brain reactivity to anger (Beaver, et al., 2008). Finally, participants were debriefed and compensated.

2.2.3. Saliva and blood collection

Saliva was collected in plastic tubes (4 mL Cryovials from Salimetrics ®) using the passive drool technique (Granger et al., 2007). We requested participants to fill 4 vials on a normal weekday in which no unusual or stressful events (e.g. exams) were scheduled. Samples were collected at wake time (between 6 and 8 AM), 30 minutes later, 1 hour later, and in the evening (between 18 and 20 PM). Instructions specified that participants should fill up at least 3 mL of saliva per sample, and should refrain from: a) brushing their teeth or eating anything within 1 hour prior to collection, b) using salivary stimulants (e.g. chewing gum, lemon drops), c) consuming alcohol 12 hours prior to collection, and d) going to the dentist 48 h prior to collection (Keevil, et al., 2013). Participants were instructed to store the tubes in the refrigerator after completing the saliva collection. This delays biodegradation of the saliva samples and thus reduces within-participant variability in hormonal concentrations (Matsui et al., 2009). Participants gave the tubes back on the second measurement day, which was scheduled on the same week of the ambulatory saliva collection. Two additional samples were taken before and after scanning. Saliva samples were stored at -20°C until shipment.

Although salivary T is generally preferable as a measure of free T (Keevil, et al., 2013), we extracted a blood sample (9 ml) before the functional measurement for serum T estimation. This allows to validate the salivary T values obtained, which should correlate at around r=.39 with serum T concentrations taken at the same time of the day (Keevil, et al., 2013). The blood sample was centrifuged for 5 minutes at room temperature. Three serum aliquots of 1 ml each were extracted and stored at -80°C until shipment. Saliva and blood

samples were placed in Styrofoam boxes filled with dry ice, sealed, and shipped to author BGK's laboratory in Manchester (UK) for analysis (see "Hormone assays").

2.2.4. Social Threat Aggression Paradigm (STAP)

We employed a modified version of the competitive reaction time task which we called the Social Threat Aggression Paradigm (STAP). In the decision phase (8 seconds), participants were shown a 2-second video of the confederate displaying either an angry or a neutral face. After the video, participants selected the loudness of an aversive sound blast to be directed at their opponent on a 1 to 8 scale. In the reaction time task (4 seconds), participants had to press any button as fast as possible in response to a jittered target stimulus cued by an exclamation mark. In the outcome phase (4 seconds), participants were shown whether they had won or lost as well as the opponent's punishment selection. In case they lost, they also received the corresponding aversive tone. An example neutral trial is depicted in Fig. 2.1A. The task was programmed so that participants won approximately two thirds of the trials. Angry trials were more likely to follow trials won by the participant, and the opponent's punishment selection was on average higher in these than in neutral trials (range 5-8 vs. 3-6, means 6.4 vs 4, respectively). Opponent's punishment selections also increased gradually over the three runs (means: 4.2, 4.8, and 5.4). Before the task, we adjusted the loudness of the maximum punishment to the participants' tolerance. The button distribution in the decision phase was set up as depicted in Fig. 2.1A to balance motor activity between both hemispheres. Participants played 60 trials (20 in each of the 3 runs), plus 4 practice trials at the beginning.

The 60 2-second videos had been pre-recorded in a separate session with the confederate. In all videos (40 neutral, 20 angry), she displaced her gaze from the keyboard to the camera keeping her expression fixed. In 10 neutral videos, the confederate kept a neutral facial expression but did not look directly into the camera. These "distracted" videos were

included to make her behavior seem less artificial, and were distributed in decreasing frequency across runs (5, 3, and 2). Angry videos showed the confederate frowning and staring intensely (Reed, et al., 2014) and were distributed in increasing frequency over runs (3, 7, and 10) to make the social interaction more believable. All videos were converted to grayscale to eliminate color-related visual artifacts. All 20 angry videos and 10 normal neutral ones were rated by five female students for validation. They rated how angry, sad, concentrated and scary was the person in the video on a -9 to 9 Likert scale, and then we compared angry and neutral videos on each dimension with paired t-tests. Angry videos were perceived as angrier (t_4=6.82, p=.002), more concentrated (t_4=3.36, p=.028) and scarier (t_4=17.94, p<.001), but not sadder (t_4=1.05, p=.352) than neutral ones. Differences were in all cases normally distributed according to Shapiro-Wilk tests (all p>.226).

2.2.5. Neuroimaging data acquisition

All scans were acquired using a 32-channel head coil mounted on a Philips Ingenia 3.0T scanner. We acquired anatomical images with a standard T1-weighted gradient echo-planar sequence (180 sagittal slices, TR=7.7, TE=3.5, FOV=240, matrix=240 x 240 mm, flip angle=8°, voxel size=1 mm isotropic). Functional images were obtained with a T2*-weighted gradient echo-planar sequence for blood-oxygen level dependent (BOLD) imaging (47 axial slices per volume, TR=2.5 s; TE=25 ms; FOV=200 mm, matrix=80 x 80 mm; flip angle=90°; voxel size=2.5 mm isotropic). We recorded 3 consecutive runs of 200 volumes each. 5 dummy scans were performed at the beginning of each run to allow steady-state tissue magnetization.

2.2.6. Behavioral data analysis

We extracted the mean punishment selection per run (1 to 3) and condition (angry and neutral) for each participant. In order to probe whether aggression increased over time, and

whether participants behaved more aggressively in angry compared to neutral trials, we ran a repeated measures analysis of variance (rm-ANOVA) with factors run and condition. We proceeded identically for mean reaction times in the decision phase, and mean reaction times in the reaction time task. Additionally, we inspected whether outcome and opponent's punishment selection in preceding trials affected subsequent aggression by means of a rm-ANOVA with factors outcome (won vs lost) and punishment (high [>4] vs low [≤4]). Sphericity corrections were applied when appropriate. Significant effects ($p<.05$) were post-hoc inspected with paired t-tests. We also performed an exploratory two-sample t-test comparing mean punishment selections in the current female sample with the male sample measured in the previous study. As in our previous study (Beyer, et al., 2015), we took the mean punishment selection per participant across runs as our main aggression measure to inspect for correlations with hormones, questionnaire data, and neural activity. All analyses described in this section were performed with the *ez* package version 4.2-2 implemented in R version 3.1.3.

2.2.7. Questionnaire data analysis

We first assessed the internal consistency of the BIS, BAS, and AQ with Cronbach's alpha, and correlated scores in each of these scales with mean aggression. BAS and AQ subscales were not separately analyzed to limit the number of tests. Internal consistency was low for the BIS ($\alpha=.604$) and the BAS ($\alpha=.643$), but high for the AQ ($\alpha=.874$). None of these scales were related to aggression (all $p>.148$), and thus were not used in other analyses.

2.2.8. Hormone assays

For both saliva and serum samples, T and C concentrations were estimated with liquid chromatography tandem mass spectrometry (LC-MS/MS) as previously described (Keevil, et al., 2013). The protocol was optimized to capture free T (Macdonald et al., 2011) and C

(Perogamvros et al., 2009) and to detect the typically low T concentrations observed in women (Keevil, et al., 2013). Previously reported mean intra-assay coefficients of variation (CV) were 5.3% for T and 8.7% for C, whereas mean inter-assay CV were 9% for T and 7.8% for C (Keevil, et al., 2013; Perogamvros, et al., 2009). The lower limit of quantification (LLOQ) in saliva samples was 5 pmol/L for T, 0.8 nmol/L for C, and 0.3 nmol/L for serum T. Serum T levels have been shown to have a mean intra-assay CV of 4.0% and mean inter-assay CV of 5.6% with this technique (Keevil, et al., 2013).

2.2.9. Hormone data analysis

We computed the area under the curve with respect to the ground (AUC) for C and T across the three morning measurements (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). This measure reflects total morning output of the hormones. The evening sample was compared to the last morning sample to check for the typical circadian decay in concentrations of both hormones, but was not further analyzed. We correlated mean T concentrations across the three serum aliquots with scan-time salivary values as a validity measure, but we used salivary T values for further analyses because they better reflect bioavailable concentrations of the hormone and are not distorted by individual variability in circulating albumin and sex hormone binding globulin (SHBG; Keevil, et al., 2013). We also compared C and T concentrations before and after the TAP with paired t-tests.

We inspected whether aggression was related to T, C, or T:C in the morning and at scan time with Pearson correlation coefficients, which were deemed significant at $p<.05$. Since T:C estimates might miss out on some additive effects (i.e. high-T, high-C subjects will have similar scores as low-T, low-C ones), we performed linear regression analyses to more thoroughly investigate TxC interactions. We computed an interaction term by multiplying T and C concentrations in the morning and at scan time, and regressed mean aggression scores

against the interaction term. In the presence of a significant interaction, we performed simple slope analyses estimating the effect of T on aggression at low (-1 SD), mean, and high (+1 SD) levels of C, as in previous work (Denson, Mehta, et al., 2013). This set of analyses was performed with the *lavaan* package version 0.5-18 implemented in R version 3.1.3.

2.2.10. Neuroimaging data analysis

Neuroimaging data were analyzed with Statistical Parametric Mapping 12 (SPM 12, Wellcome Department of Imaging Neuroscience, University College London, London, UK). We manually centered all volumes on the anterior-posterior commissure to ensure cross-subject alignment. Preprocessing involved slice timing correction to the middle slice, realignment to the first functional volume, coregistration of mean functional and anatomical images, segmentation of the anatomical image, normalization to the native voxel size, and smoothing with an 8 mm full width at half maximum (FWHM) Gaussian kernel. All participants showed motion parameters below 5 mm.

At the first level, we fitted event-related models with three regressors for the decision phase (angry, neutral, and distracted, 6 seconds) and four for the outcome phase (angry won, neutral won, angry lost, neutral lost, 4 seconds). Distracted and neutral trials were modelled together in the outcome phase. The reaction time task and the sound of the opponents' punishment in the outcome phase were also modelled. We applied a 1/128 Hz high-pass filter and SPM's built-in autocorrelation function.

At the second level, we performed a one-sample t-test contrasting angry vs neutral trials in the decision phase, and a factorial analysis in the outcome phase with the factors angry vs neutral and won vs lost, as was done in our previous study (Beyer, et al., 2015). We probed whether brain activity was modulated by within-participant variability in aggressive behavior by means of a parametric modulation, as previously described (Beyer, et al., 2015). We also

performed a series of whole-brain regression analyses, testing whether brain reactivity to anger could be modulated by between-participant variability in aggression and hormonal concentrations. To this end, we regressed the angry > neutral contrast against individual aggression scores, and against each aggression-related endocrine parameter. Unless otherwise noted, the statistical threshold for whole-brain analyses was $p<0.001$ uncorrected at the voxel level, with a cluster-level family-wise error (FWE) correction of $p<0.05$.

We also conducted analyses in our *a priori* regions of interest (ROI), namely the amygdala and the OFC. We followed a similar procedure as in Beyer et al. (2015). We contrasted angry and neutral trials against baseline in the decision phase, limiting the analyses to the bilateral amygdala and left and right medial OFC ("MNI_Frontal_Med_Orb" and "MNI_Rectus") separately. Masks were created with Wake Forest university Pickatlas (Maldjian, Laurienti, Kraft, & Burdette, 2003). These analyses were thresholded at $p<0.01$ uncorrected, minimum cluster size $k>10$. This way we isolated only voxels active in the decision phase. Then, we extracted mean parameter estimates for the angry > neutral contrast in the amygdala and OFC clusters obtained. We tested for relationships between contrast values at these ROIs and aggression, T, C, and T:C by means of Pearson correlation coefficients. Correlations were considered significant at $p<0.05$, uncorrected. As the amygdala is a heterogeneous structure comprised of functionally diverse nuclei (Terburg & van Honk, 2013), we more precisely localized significant effects within this brain region. To do so, we regressed behavioral/hormonal parameters on the angry > neutral contrast limiting the analysis to the functional amygdala mask at $p<.05$ (uncorrected), $k>20$, and estimated the peaks' location using the probabilistic maps included in the Anatomy toolbox (Eickhoff et al., 2007).

Additionally, we inspected for connectivity patterns between the amygdala and the OFC, as threat-related coupling between these two areas has been related to T (Spielberg et

al., 2015), T:C (Denson, Ronay, et al., 2013) and aggression (Coccaro, et al., 2007). We performed a time-series correlation time-locked to the decision phase of angry and neutral trials using the CONN toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012). Preprocessing steps were identical as in other work (see chapters 3 and 5). We extracted ROI-to-ROI connectivity estimates (i.e. Fisher-transformed r values) separately for angry and neutral trials and compared them with a paired t-test. We then correlated connectivity strength with aggression, T, and T:C across participants. Connectivity peaks in the amygdala were also localized with the Anatomy toolbox.

2.2.11. Mediation analyses

Finally, we explored whether the hypothesized links between hormones and aggression were mediated by amygdala and/or OFC reactivity to angry faces. Additionally, we tested whether relationships between amygdala/OFC reactivity and aggression were mediated by changes in state hormones or by activity in other brain regions. Only variables related to aggression (our main outcome variable) and to the main predictor/s were included in the models, following established recommendations (R. M. Baron & Kenny, 1986). Estimation was performed with robust maximum likelihood estimation and bias-corrected accelerated (BCa) bootstrapped confidence intervals (DiCiccio & Efron, 1996). Paths were considered significant at $p<.05$ (unc.) and if the 95% BCa confidence interval did not include zero. These analyses were also performed with the *lavaan* package.

2.3. Results

2.3.1. Behavioral results

We observed significant main effects of time ($F_{2, 76}=21.93$, $p<.001$) and condition ($F_{1, 38}=6.34$, $p=.016$) on aggressive behavior, but there was no interaction between the two. As shown in Fig. 2.1B, aggression increased over runs and was slightly higher in angry (M=4.13, SE=.16;

M: mean, SE: standard error) relative to neutral trials (M=3.94, SE=.14). The opponent's punishment selection significantly affected participants' subsequent behavior ($F_{1, 38}$=20.317, p<.001), eliciting more aggression after high compared to low punishments (M=4.27, SE=0.10 vs M=3.85, SE=0.09). The outcome of the preceding trial (i.e. won vs lost) had an effect at trend level on aggression ($F_{1, 38}$=3.16, p=.083), but did not interact with the opponent's punishment selection (p=.340). Participants selected marginally higher punishments after lost (M=4.13, SE=0.10) than after won trials (M=3.99, SE=0.15). Aggressive behavior was comparable to the male sample from our previous study (t_{69}=1.15, p=.252).

For reaction times in the decision phase, there was a main effect of time ($F_{1.69, 64.54}$=4.04, p=.030), such that participants chose the punishments increasingly quickly over runs (run 1: M=1.33, SE=0.09, run 2: M=1.20, SE=0.11, run 3: M=1.10, SE=0.10, all values in seconds). Pairwise comparisons revealed that only the difference between runs 1 and 3 was significant (t_{38}=2.43, p=.020). The main effect of condition approached significance ($F_{1, 38}$=3.55, p=.067), but we observed no interactions. Participants were narrowly faster to select the punishment in angry (M=1.14, SE=.08 seconds) than in neutral trials (M=1.28, SE=.11 seconds).

Regarding the reaction time task, there was a trend towards a main effect of time ($F_{1.65, 73.25}$=3.02, p=.065) with reaction times being lowest in the second run, but we found no other effects.

2.3.2. Hormonal results

Both ambulatory C and T levels followed a typical daily pattern (Fig. 2.1C and 2.1D, respectively). While T peaked at wake time and decreased steadily throughout the day, C was highest 30 minutes after awakening and decreased steeply in the afternoon. The difference

between the last morning sample and the evening sample was significant for both C (t_{38}=5.54, p<.001) and T (t_{38}=4.42, p<.001). Means ± SE for both hormones at each time point are presented in Table 2.1.

Table 2.1. Hormonal concentrations

	Mean	SEM
T wake up	40.355	5.048
T +30 min.	32.479	3.873
T +60 min	24.145	2.425
T Evening	14.381	1.588
T pre-scan	18.122	1.419
T post-scan	18.709	1.438
Serum T	.773	.050
C wake up	10.072	1.294
C +30 min.	11.961	.920
C +60 min.	8.837	1.160
C pre-scan	4.941	.657
C post-scan	2.951	.295
C evening	1.889	.453

Values are pmol/L for testosterone and nmol/L for cortisol (C) and serum testosterone (T)

Regarding state changes in hormonal concentrations, C decreased after the fMRI measurement compared to pre-scan levels (t_{38}=4.25, p<.001). Only the pre-scan sample was thus taken to compute TxC interactions at scan-time, since interactive effects could be driven by the post-scan decrease in cortisol. There were no significant state changes in T, and, given that the pre- and post-scan samples were highly and significantly correlated (r=.77, p<.001),

the mean of the two was obtained and used as an aggregate measure of scan-time T for correlation analyses. We also computed the pre-post percent change in T as a proxy of individual endocrine reactivity. Serum and salivary T were positively correlated (r=.40, p=.011) to a degree that matched previous findings (Keevil, et al., 2013). T (t_{38}=2.11, p=.041) and C (t_{38}=4.11, p<0.001) were higher at scan-time than in the ambulatory evening sample. Morning and scan-time concentrations were moderately correlated (C: r=.40, p=.010, T: r=.47, p=.002).

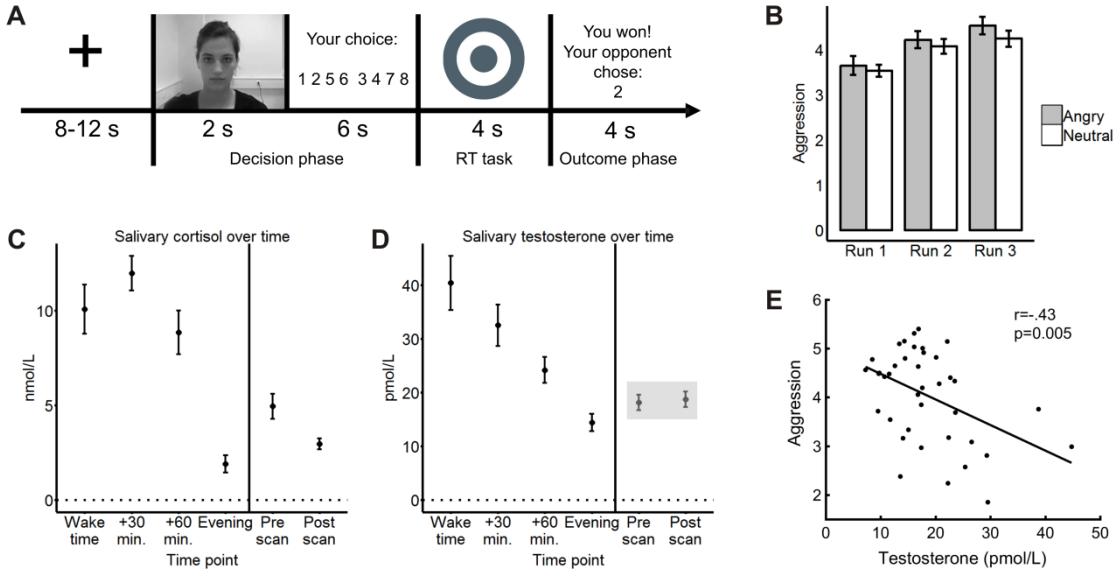

Figure 2.1 **A**: Example neutral trial of the Social Threat Aggression Paradigm (STAP). **B**: Aggressive behavior per run (left), and condition (right). *p<.05; **p<.001. **C**: Cortisol concentrations over time **D**: Testosterone concentrations over time; shadowed area depicts the time points at which testosterone concentrations were related to aggression. **E**: Scatterplot of the negative correlation between mean scan-time testosterone and mean aggressive behavior.

2.3.3. Hormone-behavior relationships

Aggression was not related to morning T, C or T:C, nor with the pre-post change in T (all p>.358). T and C did not interact in predicting aggression in the morning or before scanning (all p>.478). Mean T at scan-time was associated with aggression, but, against our predictions, the relationship was negative (r=-.43, p=0.005; Fig. 2.1E). We only took mean scan-time T for further correlations with ROI data.

2.3.4. Neuroimaging results: task effects

The main comparison of angry vs neutral trials in the decision phase revealed clusters of activation in the medial frontal gyrus (MFG), bilateral inferior frontal gyri (IFG), the orbitofrontal cortex (OFC), the bilateral posterior cerebellum, and two separate clusters in superior (STG) and middle (MTG) left temporal gyri (Table 2.2a). This pattern strongly resembles that of our previous study (Beyer, et al., 2015). Given that a cluster-level correction can be too strict to detect meaningful activations in small structures, we separately applied a false discovery rate (FDR) corrected threshold of pFDR<0.05 at the voxel level with a cluster extent threshold of k>10. With such a threshold, we also observed activity in the left caudate (Table 2.2a). We defined an 8-mm radius sphere around the group OFC cluster peak to perform ROI-to-ROI connectivity analyses with the bilateral amygdala. OFC-amygdalae coupling was lower in angry than in neutral trials (Fig. 2.2b), but the connectivity strength was unrelated to hormones or aggression (all p>.491). Amygdala connectivity peaked at x=23, y=6, z=-18 and x=-20, y=-2, z=-25 (Montreal Neurological Institute [MNI] coordinates in mm), corresponding to the basolateral amygdala (BLA).

Again paralleling our previous study, we observed strong effects of won vs lost and angry vs neutral in the outcome phase, although there was no significant interaction between the two factors. As the won vs lost effects have been extensively investigated in previous work (Beyer, et al., 2013), we report only the angry vs neutral results (Fig. 2.2C). In the outcome phase of angry relative to neutral trials, regardless of whether participants won or lost, there was widespread activation in IFG, MFG, and STG, as well as a cluster in the posterior cingulate cortex (PCC). At pFDR<0.05, k>10, activation in the left amygdala was also present (Table 2.2b). No clusters survived in the neutral > angry contrast.

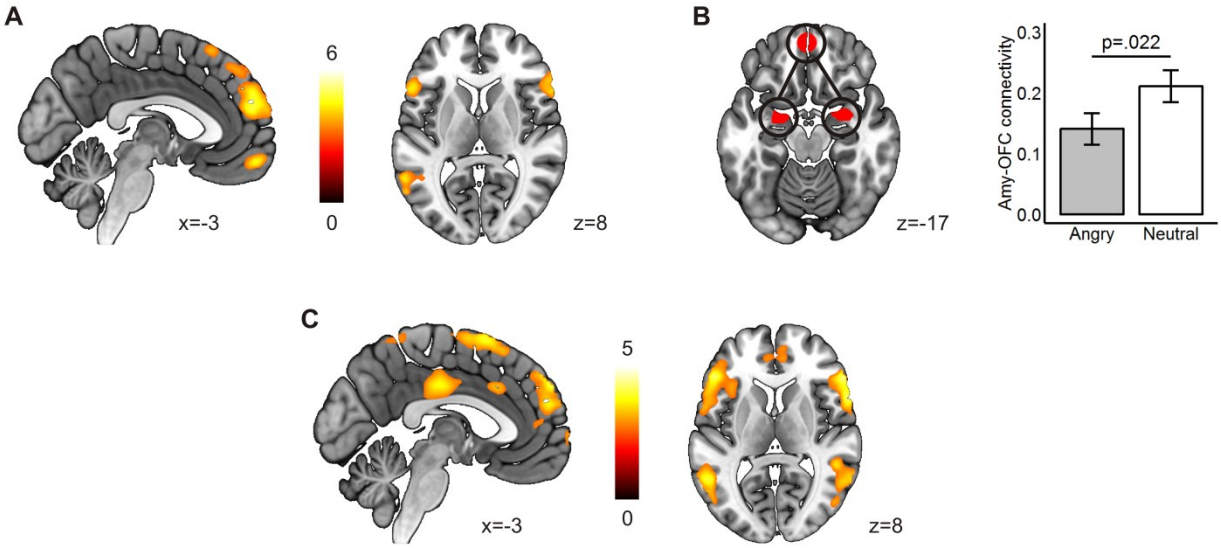

Figure 2.2 **A**: Angry > Neutral effects in the decision phase (p<.001, pFWE<.05 cluster-level corrected). **B**: Amygdala-OFC connectivity in the decision phase. **C**: Angry > Neutral effects in the outcome phase (pFDR<.05 voxel-level corrected, k>10, for visualization)

Table 2.2. Whole-brain fMRI results

Region/*Contrast*	Hem	x	y	z	k	Peak T
a) Angry > Neutral decision phase						
Inferior frontal gyrus	L	-44.5	33	0	737	6.22
Inferior temporal gyrus	R	48	8	-35	117	5.30
Cerebellum	R	23	-82	-35	152	5.26
Cerebellum	L	-22	-82	-37.5	170	5.16
Medial frontal gyrus	R/L	-4.5	55.5	20	601	5.07
Orbitofrontal cortex	R/L	-2	55.5	-15	101	4.74
Superior temporal gyrus	L	-62	-52	7.5	201	4.48
Inferior frontal gyrus	R	50.5	23	2.5	289	4.37
Middle temporal gyrus	L	-49.5	-2	-15	148	4.27
Caudate*	L	-7	8	3	24	3.66
b) Angry > Neutral outcome phase						
Inferior frontal gyrus	R	60.5	15.5	10	329	4.94
Middle temporal gyrus	R	53	-27	-5	452	4.74
Medial frontal gyrus	R/L	-9.5	63	25	262	4.65
Inferior frontal gyrus	L	-49.5	30.5	10	376	4.38
Supplementary motor area	R/L	3	10.5	67.5	344	4.38

Superior temporal gyrus	L	-57	-54.5	10	511	4.37
Posterior cingulate gyrus	R/L	5.5	-17	35	219	4.12
Amygdala*	L	-14.5	-2	-17.5	44	3.65
c) Aggression x Angry > Neutral (+)						
Lateral occipital gyrus	L	-37	-82	5	510	5.36
Superior temporal gyrus	L	-54.5	-7	2.5	216	4.61
Superior temporal gyrus	R	43	-32	10	170	4.37
d) T:C x Angry > Neutral (-)						
Middle frontal gyrus (dlPFC)	L	-20	43	43	556	6.14
Medial frontal gyrus (SMA)	R/L	-2	13	48	450	4.78

Results reported at p<.001 (uncorrected), pFWE<.05 cluster-wise corrected, except *pFDR<.05, k>10. Coordinates are mm in MNI space. Clusters ordered by peak T values. Voxel size=2.5 mm isotropic. Hem: hemisphere; k: cluster size. Only each cluster's peak is listed for clarity.

2.3.5. Neuroimaging results: brain-behavior relationships

The parametric modulation analysis did not yield significant differences between angry and neutral trials, implying that no area was significantly modulated by trial-to-trial variability in aggression. On the other hand, whole-brain regression analysis revealed three clusters of activation positively associated to aggression in the left inferior occipital gyrus, and bilateral STG (Table 2.2c, Fig. 2.3A). That is, participants with higher reactivity to anger in these areas selected on average higher punishments. We found no negative associations between aggression and brain activity.

Unlike our previous study, we found no relationships between OFC reactivity and aggression across participants, neither with the functionally defined masks nor with the cluster obtained in the decision phase (all p>.111). There was, however, a significant positive correlation between bilateral amygdala reactivity to angry faces and aggressive behavior (r=.42, p=.006; Fig. 2.3B right). This effect was maximal at coordinates x=-17, y=0, z=-25 and x=25, y=-2, z=-20, both in the BLA (Fig. 2.3B left).

2.3.6. Neuroimaging results: brain-hormone relationships

A ROI analysis revealed that amygdala reactivity to anger was negatively related to scan-time T (r=-.49, p=.001; Fig. 2.3C). We found again two clusters peaking at MNI coordinates x=-20, y=1, z=-20 and x=21, y=3, z=-18, also located in the rostral BLA. These local maxima were less than 5 mm away in any direction from the aggression-related peaks, and no more than 7 mm apart from the amygdala-OFC connectivity peaks. No other area was associated with mean scan-time T in whole-brain analyses.

Although C or T:C were not related to aggression, we ran exploratory whole-brain analyses regressing each of these parameters on the Angry > Neutral contrast in the decision phase. While we found no positive or negative effects of C, pre-scan T:C was related to lower reactivity to angry faces in the left dorsolateral prefrontal cortex (DLPFC) and in the bilateral supplementary motor area (SMA), as shown in Table 2.2d and Fig. 2.3D.

2.3.7. Mediation results

We then selected the variables to be included in mediation analyses. Scan-time T was associated with aggression and was thus defined as predictor. As amygdala reactivity correlated with both aggression and T, we included it as mediator.

Reactivity to angry faces in the bilateral STG was also associated with aggression. This region is, among other functions, central for action understanding in the context of threat detection (Parasuraman & Galster, 2013), and is connected to the amygdala through the inferior longitudinal fasciculus (Catani, Jones, Donato, & Ffytche, 2003). We hence reasoned that the threat signals stemming from the amygdala might modulate STG activity. To test this, we performed a mediation analysis with amygdala reactivity as main predictor and STG reactivity as a mediator. We extracted contrast values in the bilateral STG from the previous

Aggression x Angry > Neutral regression analysis, and ran the model. The final mediation models are depicted in Fig. 2.4.

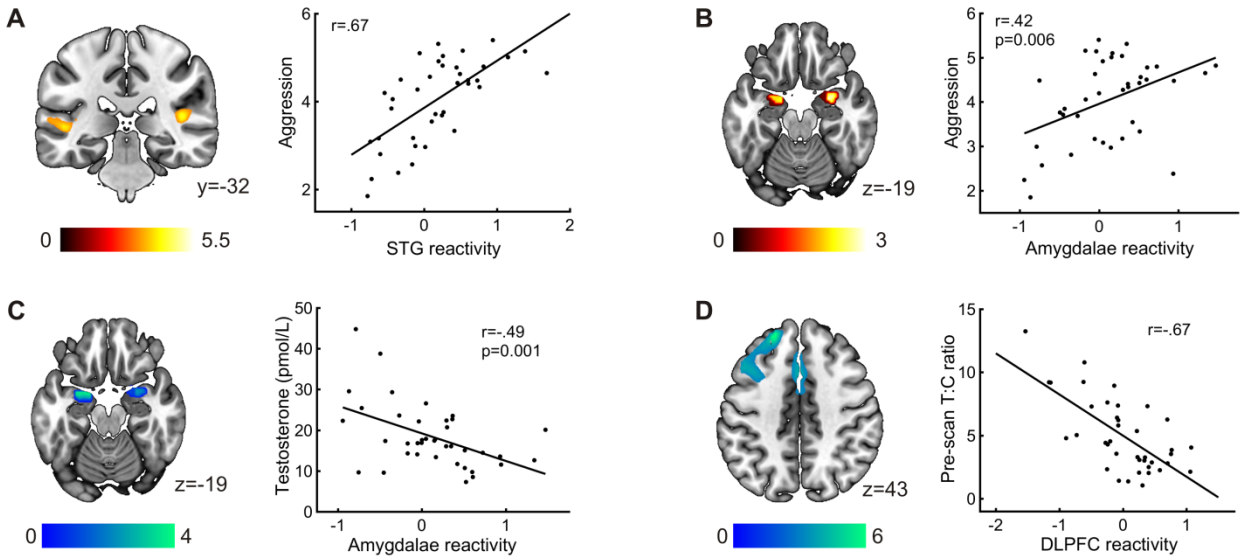

Figure 2.3 A: Bilateral superior temporal gyrus (STG) cluster positively associated with aggression in the whole-brain Angry > Neutral contrast in the decision phase (left), and scatterplot of the correlation between STG reactivity and aggression (right; p-value not computed to avoid a circular analysis). **B:** Voxels within the bilateral functional amygdala mask correlated with aggression in the Angry > Neutral contrast in the decision phase (left; $p<.05$, $k>20$) and scatterplot of the correlation between amygdala reactivity and aggression (right). **C:** Voxels within the bilateral functional amygdala mask correlated with testosterone in the Angry > Neutral contrast in the decision phase (left; $p<.05$, $k>20$), and scatterplot of the correlation between amygdala reactivity and testosterone (right). **D:** Left dorsolateral prefrontal cortex (dlPFC) and supplementary motor area (SMA) clusters negatively associated with pre-scan T:C in the whole-brain Angry > Neutral contrast in the decision phase (left), and scatterplot of the correlation between dlPFC and aggression (right; p-value not computed to avoid a circular analysis).

In model A, the indirect effect of T on aggression was not significant, and neither was the amygdala-aggression link. Therefore, amygdala reactivity did *not* mediate the relationship between scan-time T and aggression. In model B, bilateral STG activation fully and significantly mediated the relationship between amygdala reactivity and aggressive behavior.

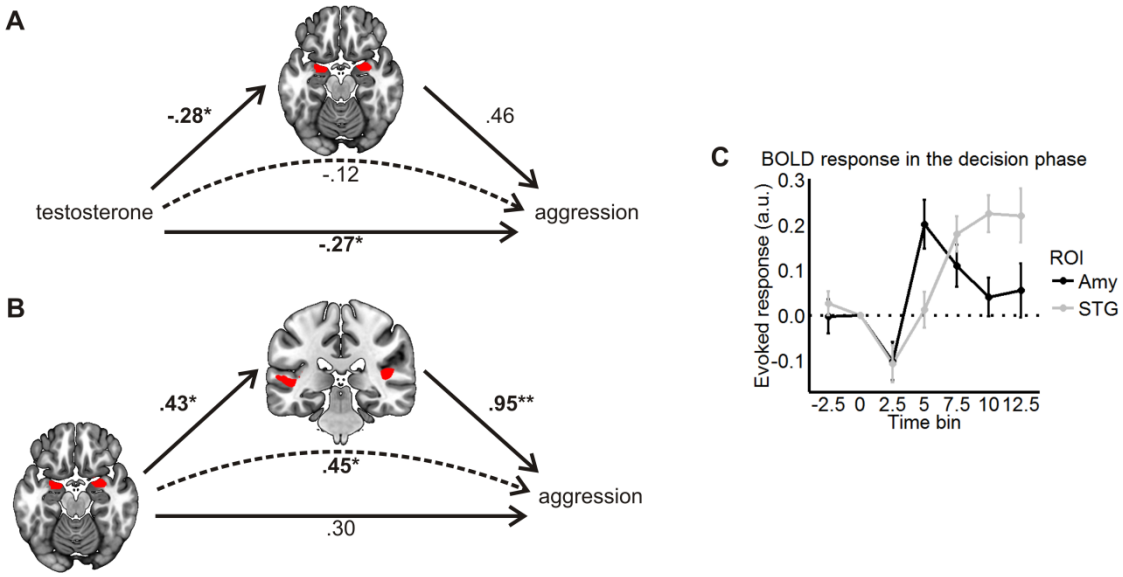

Figure 2.4 A: Amygdala reactivity did not mediate the relationship between scan-time testosterone and aggression. **B:** Bilateral superior temporal gyrus (STG) reactivity to angry faces mediated the effect of amygdala reactivity on aggression. Numbers are standardized regression coefficients. Solid lines: direct effects. Dashed lines: indirect effects. Numbers in bold depict paths significant at p<.01. The 95% BCa CI of significant paths did not include zero. *p<.01; **p<.001. **C:** Time course of the hemodynamic response in the amygdala and the STG. Values were extracted from a 5 mm sphere limited to voxels significantly active in the decision phase. Evoked responses were adjusted for all other regressors and rescaled to 0 at onset. A.u.: arbitrary units (parameter estimates).

According to this model, amygdala reactivity led to increased STG activation, which in turn resulted in heightened reactive aggression. For this model to be realistic, the amygdala should respond faster than the STG. In order to verify this, we extracted the BOLD time course of both regions in the decision phase using *rfxplot* (Gläscher, 2009). Indeed, the peak of amygdala activity in the decision phase came approximately 5 seconds earlier than that of the STG (Fig. 2.4C), strengthening the plausibility of the model.

2.4. Discussion

In the present fMRI study we probed whether endogenous T and C influence brain reactivity to angry faces and aggressive behavior in the context of a competitive interaction in healthy young women. T was negatively associated with amygdala reactivity and aggression, C was unrelated to behavioral responses, and T:C ratio was associated with lower reactivity to angry faces in dlPFC and SMA. Independently of hormonal influences, we found a positive

relationship between amygdala reactivity to anger and aggression that was mediated by STG activation. Finally, amygdala-OFC connectivity was reduced in angry trials. These results show that limbic and higher-order perceptual regions interact in the processing and response to social threat, and indicate that T can reduce amygdala reactivity while fostering strategic prosocial behavior in aggressive interactions.

2.4.1. Aggressive behavior and hormones

We found a sizeable increase in aggressive behavior over runs, mimicking other studies with the TAP that found a highly similar pattern in provoked participants, whereas non-provoked subjects showed no change in aggression over time (Böhnke, Bertsch, Kruk, & Naumann, 2010; Böhnke, Bertsch, Kruk, Richter, et al., 2010). Moreover, participants selected moderately higher punishments in angry compared to neutral trials, as well as after high selections by the opponent, and, at trend level, after losing. Participants also became quicker in selecting the punishment over time. These results closely track those of our previous study (Beyer, et al., 2015) and suggest that the paradigm successfully elicited reactive aggression.

T followed the expected decrease across the day (Wirth & Schultheiss, 2007), but we did not find a competition-boosted increase in T, nor was the change in T related to aggression. Considering that these effects do not replicate well in women (Carré & Olmstead, 2015), and, if present, occur more often after winning (Denson, Mehta, et al., 2013; Oliveira, Gouveia, & Oliveira, 2009), the fact that participants lost one third of the trials might have prevented them. Crucially and contrary to our expectations, scan-time T was negatively related to aggression. Pending replication, one might speculate that high-T participants were less intimidated by the opponent and adopted a befriending strategy over an impulsive-aggressive response. This account is supported by studies in women showing that T can reduce threat responsivity (Eisenegger, et al., 2011; Hermans et al., 2007) and increase

prosocial behavior (Eisenegger, et al., 2010; van Honk, Montoya, Bos, van Vugt, & Terburg, 2012), as well as by the lower amygdala reactivity observed in high-T participants (see "Hormonal effects on neural reactivity to anger expressions"). Importantly, T does not seem to increase pro- or antisocial behavior *per se*, but rather potentiates strategic social behavior in order to maintain or improve status or reputation (Boksem, et al., 2013; van Honk, et al., 2016). Specifically, it has been suggested that T "increases the motivation to prevent a social affront" (Eisenegger, et al., 2010). Thus, by selecting lower punishments, high-T subjects might have attempted to mitigate the opponent's aggressiveness, whereas as low-T participants might have reflexively engaged in a tit-for-tat response. However, our data do not permit to ascertain participants' underlying motives.

C levels also followed the anticipated daily trajectory, but, unlike previous work (Böhnke, Bertsch, Kruk, & Naumann, 2010), we failed to find a relationship between basal C and aggression. Other studies have also found no such link in women (Poustka, et al., 2010), and, overall, the effect seems to be more consistent in children than in older subjects (Alink et al., 2008). The pre-post scan decrease in C levels could have been partly driven by scanner-induced drowsiness or by natural circadian variation. Complementarily, it might also be that pre-scan C was heightened due to the blood extraction procedure (Dušková et al., 2015), but probably not to scanner-related stress (Weldon, et al., 2015), which had been controlled for. Unlike previous studies18, C did not moderate the effect of T on aggression. This might be due to our sample size, which was modest in comparison to the 100-150 participants recommended to detect TxC interactions (Carré & Mehta, 2011).

2.4.2. Neural reactivity to anger expressions and aggressive behavior

Task effects on brain activity were very similar to our previous study, which speaks for the reliability of the STAP and suggests that the neural correlates of anger processing in an

aggressive interaction are comparable in young men and women. In both cases, angry relative to neutral faces led to activation in inferior frontal, medial frontal, and middle/superior temporal gyri, regions often recruited when inferring others' emotions and goals (Schurz, et al., 2014). In the outcome phase, anger-related medial frontal activation was accompanied by larger temporal and inferior frontal clusters as well as by PCC and left amygdala activity. This reinforces the idea that angry faces in the context of an aggressive interaction provide relevant cues which influence the appraisal of later-occurring events (Beyer, et al., 2015). Remarkably, we found significant anger-related OFC activation across participants, whereas, in the previous study in men, OFC reactivity was only observed in non-aggressive subjects. One should in any case be cautious when directly comparing this and our previous study, as our sample size was slightly higher (7 participants more), and thus we might have reduced the likelihood of type II errors.

Basolateral amygdala (BLA) reactivity to angry faces was greater as a function of aggressive behavior. This explains the absence of amygdala effects for the angry > neutral contrast in whole brain analyses, as reactivity in this structure was heightened only among participants who responded aggressively. Although in our foregoing study in men there was no such relationship, other research has found associations between amygdala reactivity to threat/provocation and aggression in mixed-gender samples (Coccaro, et al., 2007; Gospic, et al., 2011). In another study from our lab (see Chapter 4), we found that the BLA is specifically recruited when participants actively avoid a highly threatening opponent, relative to a less threatening one. Taken together, these results suggest that the BLA signals threat relevance, flexibly facilitating either aggression or avoidance in a context-dependent manner (LeDoux, 2003). Importantly, amygdala-OFC connectivity was reduced in the decision phase of angry trials. This decoupling was likely driven by the higher OFC reactivity to anger, possibly reflecting a regulatory or social-evaluative response to interpersonal threat (Beyer, et

al., 2015). Low connectivity between these two areas has been reported in IED patients relative to healthy controls (Coccaro, et al., 2007), and related to increased negative affect during the presentation of aversive pictures (Banks, et al., 2007). Hence, the reduction in amygdala-OFC coupling observed in angry trials might have indirectly hampered the suppression of aggressive impulses (Coccaro, et al., 2011).

We did not find any difference between conditions in trial-to-trial variability in aggression. This contrasts with our previous study, in which anterior cingulate cortex (ACC) activity was upregulated as a function of aggressive behavior in angry trials (Beyer, et al., 2015). We did find associations however between bilateral STG reactivity to angry faces and aggression on a between-participant basis. STG activity has previously been shown to increase when focusing one's attention on the aggressor relative to the victim of a violent conflict (Van den Stock, Hortensius, Sinke, Goebel, & de Gelder, 2015), and when appraising unfair proposals as negative (Grecucci, Giorgetta, Bonini, & Sanfey, 2013). These findings, in addition to its general involvement in mentalizing (Schurz, et al., 2014) and threat detection (Parasuraman & Galster, 2013), suggest that persons with higher reactivity to anger in STG might be more likely to interpret the opponent's intentions as hostile and hence respond aggressively. Mediation analyses also support this view, as the link between amygdala reactivity and aggression was dependent on STG activation. Remarkably, the BOLD signal in the amygdala peaked around 5 seconds earlier than in the STG during the decision phase. Hence, whereas the amygdala might rapidly code for emotional salience (Liddell et al., 2005; Pessoa, Kastner, & Ungerleider, 2002), STG activity could more slowly subserve conscious threat perception. Supporting this account, it has been recently shown that threatening faces increase both STG and amygdala activation, but only STG reactivity correlates with emotion recognition accuracy (Sussman, Weinberg, Szekely, Hajcak, & Mohanty, 2016).

2.4.3. Hormonal effects on neural reactivity to anger expressions

BLA reactivity was negatively related to T, but did not mediate the T-aggression link. It might be that high-T individuals perceived angry faces as less threatening (Stanton, et al., 2009), as suggested by studies showing that T administration mitigates fear responses (Eisenegger, et al., 2011; Hermans, Putman, Baas, Koppeschaar, & van Honk, 2006). Alternatively, T might generally impair emotion recognition (van Honk & Schutter, 2007). In any case, this result indicates that high-T participants might have been better able to override defensive retaliatory responses, although the effect of T on aggression might be better explained by other factors such as testosterone-to-estradiol aromatization (Unger et al., 2015) than by amygdala reactivity alone. It has been proposed that T buffers acute fear through its action on the BLA, which, together with the OFC, inhibits central-medial amygdala (CMA) output to the brainstem (Terburg & van Honk, 2013). From this perspective, if high-T participants were indeed less sensitive to threat, they might have required the assumed regulatory input of the BLA to a lesser extent. Our data nonetheless suggest that the BLA does not accomplish an inhibitory or excitatory role in itself. Rather, as reasoned above, BLA activation tracks threat salience, and promotes defensive behavior adaptively.

Pre-scan T:C was related to lower dlPFC and SMA activity in the decision phase of angry vs neutral trials. These regions have been classically implicated in executive control (Miller & Cohen, 2001), and are sensitive to contextual changes during goal-directed behavior (Bahlmann, Aarts, & D'Esposito, 2015; Rudorf & Hare, 2014). Hence, participants with higher T relative to C might have perceived angry faces as less salient or challenging. A previous study also found decreased middle frontal reactivity to emotional faces as a function of endogenous testosterone in young and middle aged women (van Wingen, et al., 2008). Our finding, however, partially contradicts a report of increased dlPFC recruitment during anger control in young Asian men with high T and low C (Denson, Ronay, et al., 2013). This

divergence might be due to the use of a different paradigm (i.e. in that study subjects were asked to *actively* control the provocation-induced anger), and/or to sample characteristics. It should also be noted that we did not find associations between t:c and subcortical reactivity, in contrast to previous work (Hermans, et al., 2008).

It is worth noting that only hormonal concentrations at scan-time were related to brain reactivity and aggression. Other studies have also found that morning and afternoon T have separable effects on the processing of threat signals, such that participants with higher morning (but not afternoon) T seem to be more attentive to angry faces (van Honk et al., 1999; Wirth & Schultheiss, 2007). Also, as mentioned in the introduction, the relationship between aggression and T appears to be stronger in the afternoon and evening than in the morning (Archer, et al., 2005), and varies across the menstrual cycle (D. M. Dougherty, Bjork, Moeller, & Swann, 1997). It is also unclear whether our results are entirely comparable to those of T administration studies, in which effects are likely driven by a phasic augmentation of circulating hormone levels. In the present work, scan-time T remained unchanged after the STAP and was correlated with morning T, hence probably reflecting basal endocrine function. More research is needed on how different parameters of the infradian, circadian, and ultradian hormonal fluctuation affect reactive aggression and its neural basis. More research is needed on how different parameters of the infradian, circadian, and ultradian hormonal fluctuation affect reactive aggression and its neural basis.

2.4.4. Limitations

An important limitation of our study is that we only took ambulatory saliva samples on one regular weekday, which might not be sufficient to establish a reliable baseline (Hellhammer et al., 2007). Moreover, we did not measure anthropometric variables such as the body mass index, which could have also influenced our results (Champaneri et al., 2013). Nevertheless,

we did control for menstrual status and oral contraceptive use, and, given the prototypical temporal pattern observed in diurnal C and T levels, we deem the obtained hormonal concentrations trustable. The relatively small sample size might have anyway curtailed the sensitivity of our analyses, especially regarding C effects and TxC interactions.

It should also be taken into account that neural responses in the angry > neutral contrasts might reflect general, rather than anger-specific, emotional reactivity. Further studies could use different facial expressions (e.g. sadness) and measure participants' subjective and/or physiological reactions in order to disentangle the valence specificity of the observed effects. It would also be meaningful to employ opponents from both genders in larger, mixed samples.

2.5. Conclusion

We probed the influence of T and C on reactive aggression and its neural substrates in healthy young women. Our data support the established notion that amygdala reactivity to threat is a key factor in reactive aggression, and that this effect is accompanied by reduced amygdala-OFC coupling. However, our results further suggest a complementary mechanism by which amygdala-dependent aggressive impulses would be exacerbated by activation in the STG. Scan-time T was negatively related to aggression and amygdala reactivity, suggesting that high-T participants might have felt less threatened by the opponent. All in all, our results expand existing neural models of reactive aggression, and highlight the complex nature of T-aggression relationships in humans.

Chapter 3: Introducing the Fight-or-Escape (FOE) paradigm[2]

3.1. Introduction

Aggressive behavior is a great challenge to individuals and to society as a whole. It is thus not surprising that research on aggression has been conducted for decades and continues to be an important problem addressed by scientific experiments today (C. A. Anderson & Bushman, 1997). Laboratory experiments on human aggression usually employ one of several well-established paradigms, which are derived from different theories of human aggression. One central aspect is interpersonal provocation: two of the most widely used aggression paradigms, the point-subtraction-aggression-paradigm (PSAP; Cherek, 1981) and the Taylor Aggression Paradigm (TAP; Taylor, 1967) confront the participant with an ostensible opponent, who in some way inflicts harm upon the participant. The participant then has the option to retaliate, which is directly measured within the paradigms. This approach is based on the well-established theory that reactive aggression is usually elicited as a response to some form of provocation or frustration (Lawrence, 2006). However, one shortcoming of these paradigms is that they largely limit the participant's behavioral options to showing lower or higher levels of aggression. In real-life hostile situations, one usually has the option to avoid the aggressive interaction altogether and withdraw from the situation. Limiting the range of behavioral options possibly limits the applicability of laboratory findings to real-life aggression. Thus, implementing an escape option in laboratory aggression paradigms is an important step towards improving research on aggression (Tedeschi & Quigley, 1996).

[2] Largely corresponds to Beyer, F., Buades-Rotger, M., Claes, M., & Krämer, U. M. (submitted). Hit or run: exploring aggressive and avoidant reactions to interpersonal provocation using a novel Fight-or-Escape paradigm (FOE). The first two authors share first authorship. The study was conceived by FB and UMK. I collaborated in the design of the study, performed the experiments, analyzed the data, and wrote the manuscript.

In the Point Subtraction Aggression Paradigm (PSAP), participants are led to believe that they will interact with another player in another room while trying to earn points, which can later be exchanged for money (Cherek, 1981). The participant can earn points by pressing a button as quickly as possible. With another button, the participant can subtract points from his co-player. Provocation is implemented as the co-player's subtraction of the participant's points, whereas aggressive behavior is measured as the number of times the participant uses the point subtraction button to inflict cost upon his co-player. In some versions of the PSAP, a protective button is implemented, which protects the participant from point subtraction for a certain time (Cherek, Moeller, Schnapp, & Dougherty, 1997).

In the TAP, participants are also led to believe that they are competing against another player. Subjects are told they will engage in a reaction time (RT) competition with the opponent. They are required to respond quickly to a stimulus, and are led to believe that the faster player in each run wins. The loser gets punished with an aversive stimulus (e.g. a mild electric shock or a sound blast), which can be adjusted in intensity. In each trial, the winner determines the intensity of punishment for the loser. Provocation in this paradigm is manipulated as the punishment level assigned to the participant, whereas aggression is measured as the punishment level selected by the participant for the opponent.

These paradigms have been widely used in behavioral research on aggression and, more recently, also in research on the neural basis of aggressive behavior (Beyer, et al., 2013; Lotze, et al., 2007; Veit et al., 2010). In both paradigms, aggressive behavior is non-instrumental insofar as the main outcome of the task (i.e. winning the RT task; earning points) is not improved by aggressive behavior but may actually (in case of the PSAP) be hindered by it. One conceptual advantage of the TAP is that while in the PSAP aggression is costly (the participant cannot simultaneously subtract and earn points), in the TAP aggressive behavior can be measured independently of cost-benefit considerations. Similarly, while avoidance

behavior is assumed to be related to fear (Carver, 2004), in the PSAP the protective button may be used based on considerations of monetary tradeoff (weighing the points missed while pressing the protective button against the points saved by avoiding subtraction). As such, it may constitute an imperfect model of real-life avoidance behavior.

This latter point, however, also poses an important limitation for the extrinsic validity of the TAP. We have previously argued (Beyer, et al., 2013) that based on theories about the role of emotional reactivity in aggressive behavior (Carver & Harmon-Jones, 2009), one would expect a negative relationship between fear reactivity and aggressive behavior in the TAP. Specifically, anger, as an approach-related affect should be most reliably elicited by provocation in participants low in fear reactivity and high in approach motivation. On the other hand, participants high in fear reactivity should react to a provocative confrontation with increased avoidance tendencies, rather than aggressive approach tendencies. In an fMRI-study using the TAP, in which we measured threat reactivity as fear potentiation (FP) of the startle response, we found support for this hypothesis on a neural level (Beyer, et al., 2013). The startle response in humans can be measured as the eye-blink amplitude in response to a sudden burst of white noise. FP is defined as the amplification of this amplitude when the participant is watching threatening rather than neutral pictures, and has been shown to be a good measure of emotional reactivity to threat (Vaidyanathan, Patrick, & Bernat, 2009). In participants low in FP, we observed increased activity in areas of the so-called mentalizing network when they were confronted with a provocative opponent in the TAP. For participants high in FP, we observed the opposite effect, a reduction of activity in the mentalizing network due to provocation. The mentalizing network is a group of neural structures associated with social information processing, perspective-taking and empathy (M. D. Lieberman, 2007), and we interpreted this effect as cognitive avoidance of the aggressive interaction in highly fearful participants. However, we observed no effects on a behavioral level. One potential reason for

this is the lack of an avoidance option in the TAP. Participants high in fear reactivity had no option of actually escaping the confrontation with the aggressive opponent and consequently may have adopted a tit-for-tat-like strategy. In many everyday incidents of provocation, however, avoidance is a realistic and valid behavioral option. Thus, we expect that the ecological validity of aggression paradigms should be increased by including a true avoidance option, producing the proposed relationship between personality traits (namely fear reactivity) and aggressive behavior.

In this study, we present a novel interactive aggression paradigm with an avoidance option: the "Fight-Or-Escape" (FOE) paradigm. In a first experiment, we implemented the task in a female student sample, also measuring FP of the startle response in a similar setup as we previously used for the TAP. This experiment is designed to test our previous interpretation of the non-existing relationship between FP and aggression in the TAP. To further validate our new paradigm, in a second experiment, we combined the task with two other well-established tests of social avoidance tendencies, the approach-avoidance-task (AAT; Roelofs, Elzinga, & Rotteveel, 2005) and a dot-probe-task (DPT; MacLeod, Mathews, & Tata, 1986) using emotional facial stimuli. We expected that fear reactivity and social avoidance should be associated with avoidant behavior and less aggression towards a provocative opponent.

Experiment 1

3.2. Methods

3.2.1. Participants and procedure

Forty-three healthy female volunteers (M_{age}=22±2 years) participated in this study. Participants were invited to the lab in groups of three. They were informed that two different

experiments would be carried out: one EMG-measurement (the startle measure) and one group-task. The order of the two tasks was randomized across participants.

For the aggression task, the three participants received written instructions together. Prior to this task, participants filled out questionnaires assessing approach/avoidance tendencies and empathy (see below). After the aggression task, participants filled out a questionnaire probing for suspicion concerning the task, as well as a questionnaire on trait aggression. Finally, all participants were fully debriefed and reimbursed for their participation with 8 Euro per hour. All procedures were performed according to the Declaration of Helsinki and approved by the local ethics committee.

3.2.2. Fight-or-Escape (FOE) paradigm

The FOE-paradigm was set up as a competitive RT game for three people. Participants were instructed that each would randomly be assigned one of three characters from the "Lord of the Rings" trilogy (Tolkien, 1954a). As a background story, participants were told that Sauron and Saruman were sending out orcs to obtain the Ring of Power from Frodo, and that during the game it would be decided whether Frodo succeeds in destroying the ring or whether it is taken by his opponents. The three available characters were Frodo, Sauron and Saruman. In fact, each participant was assigned the role of Frodo. Participants were instructed that Sauron and Saruman would compete together against Frodo.

In each trial, Frodo would be playing against one of his two opponents and the winner of each trial would receive one point, with the points of Sauron and Saruman being summed together. At the end of the game, the party with the highest score (Frodo or Sauron and Saruman) would win. The winner of each trial was decided in a simple RT task: an exclamation mark was presented, followed by the picture of an orc. Frodo and his opponent had to respond to the orc by button press as quickly as possible. The faster player received

one point, whereas the loser was punished with an aversive sound. At the beginning of each round, participants selected a punishment level (i.e. the noise level, ranging from 1-8) for their opponent, in case that the participant would win. Additionally, at the beginning of each trial, Frodo had the option of putting on the ring and thus becoming invisible, in order to avoid the confrontation. In that case, nobody received a point and nobody got punished; this choice constituted the "escape" option, i.e. avoiding potential punishment while foregoing the chance of earning a point. Thus the sequence for one trial was as follows (Fig. 3.1): 1) Information on which opponent the participant was playing against. 2) Choice of putting on the ring; in case Frodo put on the ring, a short message ("You escape the orc") was displayed and the trial ended. 3) If Frodo did not put on the ring: punishment selection. 4) RT task. 5) Outcome phase: information on who won and which punishment level the opponent selected.

During screen number 2 (option of putting on the ring), Frodo was presented with the remaining number of times he could use the ring. Throughout the game, he could put on the ring a maximum number of 10 times. If a participant selected the ring after the 10 permitted escape options had been used, the message "You cannot use the ring anymore" was displayed and the trial then automatically proceeded to the punishment selection.

The game was programmed such that Frodo would win about 50% of trials. Winning and losing was not related to RT, unless participants were slower than 500ms, in which case they would lose. The behavior of Sauron and Saruman was programmed such that one opponent was non-provocative, selecting low punishments (range 1-4, mean = 2.3), whereas the other was highly provocative, selecting high punishments (range 4-8, mean = 6.0). Frodo played 20 trials against each opponent in randomized order. Thus, Frodo had the option of avoiding up to 50% of the confrontations with the provocative opponent. Frodo could distribute using the ring between the two opponents in any ratio, and he was not obliged to use all 10 escape options. Since nobody received a point in escape trials, putting on the ring

did not affect the ultimate outcome of the game. The identity of the provocative opponent (Sauron or Saruman) was randomized across participants. The intensity of the punishment (A Styrofoam scratching sound) was adapted to each participant's tolerance before starting the task.

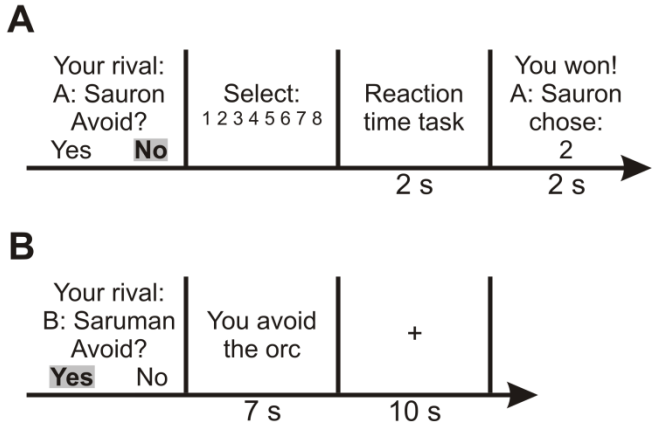

Fig. 3.1 Example trials of the Fight-or-Escape (FOE) paradigm. **A**: trial in which the participant confronted the opponent. **B**: trial in which the participant avoided the opponent. The fixation cross is only depicted for the avoidance trial, but appeared regardless of the participant's decision. See main text for details.

3.2.3. Behavioral measures

A range of behavioral measures can be derived from this task: pure aggression measures were obtained by calculating mean punishment selection for each opponent. However, this score would be identical for a participant who avoided the aggressive opponent in half the trials but otherwise retaliated with high punishment selections, and a participant who never avoided and behaved aggressively. Avoidance measures were obtained by counting the number of times the ring was used to avoid each opponent. Similarly to the problem mentioned for the mean aggression score, this avoidance measure would be identical for participants high and low in punishment selections, if both avoid the same number of trials. To address these issues, we additionally calculated a combined aggression-avoidance score by summing punishment selections for each opponent across all trials. For this measure, avoidance trials are scored as

zero. Consequently, this score is affected both by the number of times a participant chose not to play against an opponent, and by the punishment she selected when she did. Accordingly, this measure reflects the absolute amount of aggression shown towards the opponent. A medium score for the provocative opponent could be reached by a participant who frequently avoided him, but behaved aggressively in the remaining trials, or a participant who did not avoid him, but showed moderate levels of aggression.

3.2.4. Personality questionnaires

A German version (Herzberg, 2003) of the Buss and Perry Aggression Questionnaire (AQ; A. H. Buss & M. Perry, 1992) was used to assess trait aggressiveness. The AQ consists of four subscales: physical aggression, verbal aggression, hostility and anger. To assess approach and avoidance tendencies, a German version (Strobel, et al., 2001) of the behavioral inhibition and activation scales (Carver & White, 1994) was used. We also used the Interpersonal Reactivity Index (IRI; M. H. Davis, 1983) in our own translation to measure empathy and perspective taking tendencies.

3.2.5. Measurement of Fear Potentiation

To measure FP, we used a setup which was adapted from previous studies (Caseras et al., 2006; Conzelmann et al., 2009). Participants were presented with 51 pictures from the International Affective Pictures System (IAPS; Lang, Bradley, & Cuthbert, 1999). Half of these pictures were threatening (e.g. a gun pointed at the viewer, an attacking dog), the other half were neutral (e.g. a secretary on the phone; household objects). Pictures were presented in a fixed order which was set up randomly with the constraint that no more than two pictures of the same valence were presented consecutively. Each picture was presented for 6 seconds with a 12 second inter trial interval (ITI), during which a white central cross was presented on a black background. During 18 threatening and 21 neutral pictures, a short burst of white

noise (50 ms, 95 dB), was presented over speakers 1.5, 2.8 or 4.0 seconds after picture onset. For the remaining 12 pictures, the startle probe was presented during the ITI and these trials were not analyzed. To account for initial habituation of the startle response, four startle probes were presented while participants watched the fixation cross. Additionally, the first three picture trials (all neutral), were discarded.

3.2.6. EMG measurement and analysis

Two Ag-AgCl electrodes were placed below the left lower eyelid, one in line with the pupil and the other 1-2cm to the left of the first. A ground electrode was positioned centrally on the forehead. Prior to electrode placement, the skin was treated with a peeling paste and alcohol. The EMG signal was amplified and recorded at a sampling rate of 250 Hz using an EEG amplifier (32-channel Brainamp; Brain Products).

We analyzed EMG recordings with EEGLAB, a Matlab-based open-source toolbox (Delorme & Makeig, 2004). EMG signals were high-pass filtered at 10 Hz, low-pass filtered at 500 Hz and baseline-corrected using the 50 ms prior to onset of the startle probe as baseline. We then visually inspected each startle trial for artifacts. Trials with excessive noise or eyeblinks in the 50 ms baseline period were excluded. Blink magnitude was measured as the maximum absolute amplitude in an interval of 20-160 ms following the startle probe. Blink scores were z-transformed within each participant across all trials. We then subtracted the mean standardized blink amplitude for neutral pictures from the respective value for threatening trials, to get individual FP scores.

3.2.7. Correlational analyses

To investigate the relationship between FP and aggressive and avoidant behavior, we correlated FP scores with the respective behavioral measures (mean aggression, number of avoidance choices, sum of punishment selections across trials) for each opponent. We

hypothesized that FP should be negatively related to aggressive behavior and positively to avoidant behavior, resulting in a negative correlation between FP and the summed punishment score. On an exploratory level, we also correlated personality questionnaire scores with FP and behavioral measures from the task.

3.3. Results

Of the 43 participants, 16 had to be excluded due to technical problems during startle measurements and/or bad EMG data quality (13) or suspicion concerning the aggression task (3), leaving a final sample of 27 participants.

On average, participants selected higher punishment levels for the provoking (M=4.1, SE=0.3) than the unprovocative opponent (M=3.3, SE=0.3), t_{26}= 3.0, p<.01. Of the 10 avoidance options, participants used on average 5.9 (SE=0.7; range 0–10). There was no significant difference in the number of times participants avoided the provocative (M=3.3; SE=0.5) and non-provocative opponent (M=2.6; SE=0.4), p=0.33. There was also no significant difference between startle responses to neutral and threatening pictures, on average. Across participants, however, there was great variability in FP (range -.59 to .40; mean= .03; SE= .05).

We found a negative correlation between FP and the summed punishment score for the provocative opponent (r=-.45, p<.05; Fig. 3.2A). This relationship only yielded marginal significance for the non-provocative opponent (r=-.35, p=.08). A similar effect was observed for mean aggression scores, with a negative correlation between FP and aggression against the provocative opponent (r=-.48, p<.05; Fig. 3.2B), but no effect for the non-provocative opponent (r=-.23, p=.26). We found no relationship between FP and the number of times the avoidance option was chosen for either opponent or the absolute number of trials avoided (all p>.2). Mean aggression towards the provocative opponent was positively correlated with trait

anger (r=.43, p<.05). There was a negative correlation between FP and trait anger as assessed with the AQ (r=-.46, p<.05). Thus, FP was negatively related to trait anger, aggressive behavior towards the provocative opponent, as well as to a conglomerate measure of aggression and avoidance for this opponent.

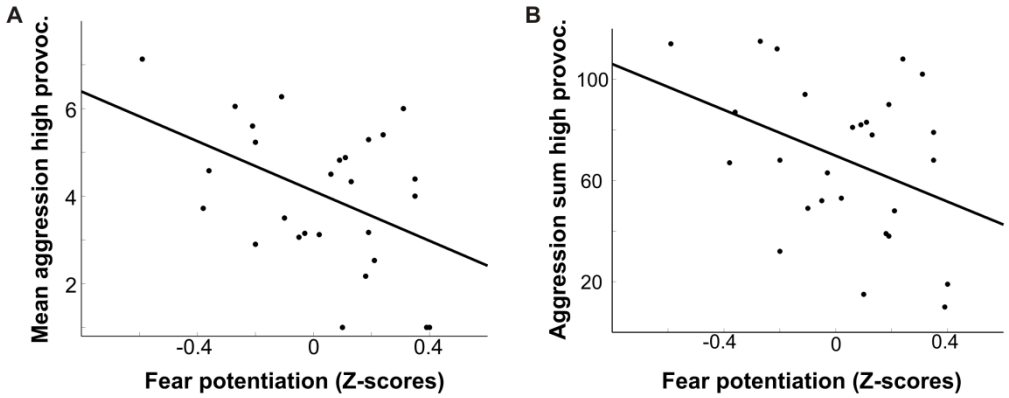

Fig. 3.2 Scatterplot depicting the negative correlations between Fear Potentiation (FP) and mean aggression (**A**), and between FP and aggression sum scores (**B**).

3.4. Discussion

In this experiment, we found a negative relationship between fear reactivity, as measured using a fear potentiated startle paradigm, and aggressive behavior towards a provocative opponent. In contrast to our previous study, where we found no such effect, participants here had the option of avoiding the aggressive interaction. Interestingly, we found no direct relationship between avoidance behavior and FP. The intrinsically low variability in avoidance (i.e. a maximum 10 times) might have curtailed the possibility to find direct relationships between avoidance and other parameters. Nonetheless, there was a negative relationship between FP and a conglomerate measure of aggression and avoidance. Thus, in situations where avoiding a confrontation is explicitly possible, participants high in fear reactivity behave less aggressively towards an aggressive opponent than participants low in fear reactivity.

The negative correlation we found between FP and trait anger supports previous findings using startle paradigms which suggest that anger is an emotion related to behavioral approach tendencies (Amodio & Harmon-Jones, 2011). FP, on the other hand, is a defensive reflex associated with behavioral avoidance, akin to freezing behavior in animals (M. Davis, Falls, Campeau, & Kim, 1993). The positive correlation we found between trait anger and aggression towards the provocative opponent further underlines the role of emotional reactivity in aggressive retaliation. Thus, people high in fear reactivity are overall less prone to feelings of anger and are less likely to retaliate against an aggressive opponent. People low in fear reactivity, on the other hand, report more angry impulses, which are related to more aggressive behavior. One has to note, however, that we conducted correlations with personality questionnaires in an exploratory approach and the observed effects would not survive multiple testing correction.

Experiment 2

3.5. Introduction

In our first experiment, we used FP as a measure of fear reactivity in order to compare our findings to our previous study using the TAP. This showed the proposed negative relationship between FP and aggression and supports our previous argument that the lack of an avoidance option in the TAP reduces interpersonal variability in aggression, as it imposes an unnatural limitation of behavioral options upon the participant. In order to further validate the paradigm and test the proposed negative relationship between avoidance and aggression, in a second experiment we combined the FOE-paradigm with two well-established tasks designed to measure implicit approach and avoidance tendencies towards social stimuli.

In the approach-avoidance-task (AAT; Roelofs, et al., 2005), participants are asked to perform approach movements (i.e. pulling a joystick towards themselves) or avoidance

movements (pushing a joystick away from themselves) in response to visual stimuli. Using stimuli that typically elicit approach or avoidance tendencies, congruency effects can be observed: participants are faster in pushing away stimuli that elicit avoidance tendencies than they are at pulling them close, whereas the opposite effect is observed for approach-related stimuli. This has been found for social stimuli as angry and happy faces (Roelofs, et al., 2005), as well as other affective stimuli, as pictures of spiders in participants high in fear of spiders (M. Rinck & Becker, 2007). The AAT thus constitutes a measure of *behavioral* avoidance.

The dot-probe-task (DPT; MacLeod, et al., 1986) assesses automatic orientation of attention towards one of two visual stimuli. In each trial, two stimuli are shortly presented at opposing sides of the screen center. After stimulus offset, a dot is presented on one side, and the participant is asked to press a corresponding response button (i.e. left or right). As individuals tend to initially allocate attention to threatening stimuli and then avoid them (Cooper & Langton, 2006; Mike Rinck & Becker, 2006), they should be slower to respond to a dot presented on the side of the threatening stimulus at long exposition times (Mogg, Bradley, Miles, & Dixon, 2004). The DPT is a measure of *attentional* avoidance.

In this experiment we used the AAT with pictures of angry and happy facial expressions and the DPT with pictures of angry and neutral facial expressions. We hypothesized that participants who showed high behavioral and attentional avoidance of angry faces would be less aggressive towards an aggressive opponent in the FOE-paradigm and would more frequently use the avoidance option for that opponent compared to participants showing less avoidance of angry faces.

3.6. Methods

3.6.1. Participants and procedure

Forty-two healthy volunteers (18 men; M_{age}= 22±3 years) participated in this study. Participants were invited to the lab in same-sex triads. They were told that they would perform three computer tasks, the first of which would be interactive. The order of the two individual tasks was randomized across participants. All procedures were performed according to the Declaration of Helsinki and approved by the local ethics committee.

3.6.2. FOE-paradigm

The setup of the FOE-paradigm was identical to that used in experiment 1, with one exception: instead of randomizing the trials against the two opponents, we now implemented an alternating order. Thus, participants knew that they would be playing against the two opponents alternatively. We modified this in order to increase the salience of the avoidance option. Whereas in experiment 1 subjects could only minimize the absolute number of trials they played against the aggressive opponent, a trial where they chose to avoid this opponent could still be followed by a trial against the same opponent. As such, the avoidance was realized on a global level rather than immediately. To make the avoidance more prominent, the alternating schedule implemented here ensured that if a participant avoided the aggressive opponent, in the following trial they would always be playing against the non-aggressive one.

3.6.3. Approach-Avoidance-Task

For this task, we used photographs of angry and happy facial expressions of the Radboud Faces Database (O. Langner et al., 2010). We used pictures of 30 individuals (15 female), each showing an angry facial expression in one picture, and a happy expression in the other. Pictures of 9 different individuals (4 female) were used for practice blocks. The pictures were cropped into an oval shape, removing hair, ears and neck.

Participants were given a standard joystick (Speedlink® Dark Tornado) as response device. At the beginning of each trial, a fixation cross was presented centrally on a white background. The participant started the trial by pressing the 'shoot' button on the joystick. Following this, one picture was presented centrally on the screen. Participants could reduce picture size by pushing the joystick away. By pulling the joystick towards themselves, they could increase picture size. Picture size was varied gradually in 7 steps.

In one block, participants were instructed to "push away" angry faces and "pull close" happy faces as quickly as possible; in the other block, the reverse instruction was given. Block order was randomized between subjects. Each block consisted of 30 happy and 30 angry trials, and the same pictures were used in both blocks. Each block was preceded by a practice run. During practice runs, each trial was followed by feedback (a green check-mark for correct reactions, a red cross for errors). The practice run for the first block consisted of 20 trials. The second practice run consisted of 28 trials, since participants had to reverse their response patterns from the first block.

For RT, we analyzed the interval between stimulus presentation and movement onset. Incorrect trials (including trials in which the initial movement was performed in the incorrect direction, followed by a correction), and trials with response latencies shorter than 150 ms or longer than 3 SD from the own mean were excluded from analysis. We calculated the pull minus push difference in RT (higher scores meaning higher avoidance) for angry and happy faces separately. We compared both biases with a paired t-test.

3.6.4. Dot-Probe-Task

For the dot-probe task, we used 40 pictures from a set of previously validated videos (Kircher et al., 2013) showing angry and neutral facial expressions of 20 different professional actors

(9 women). In each trial, the neutral and angry pictures of one person were presented together, to the left and right of the screen center.

Each trial began with a centrally located fixation cross being presented for a random interval between 500 and 1000ms. Then, the neutral and angry pictures of one person were presented for 1000ms. At picture offset, a dot was presented located to the left or right of the screen center, at the coordinate where the center of the respective picture had been. Participants were instructed to react as quickly as possible to the dot, by pressing a left button (A) on the computer keyboard if the dot was presented on the left, and a right button (L) if it was presented at the right side.

The task consisted of two blocks. In each block, 80 trials were presented, with each picture pair presented four times, once in each of the four conditions: angry picture on the left, dot probe on the left; angry picture on the left, dot probe on the right; angry picture on the right, dot probe on the right; and angry picture on the right, dot probe on the left.

We analyzed attentional approach vs. avoidance tendencies for angry pictures by subtracting mean RT for the neutral condition (dot probe in location of neutral picture) from angry conditions (dot probe in location of angry picture). The higher this score was for a given participant, the greater was this participant's reaction-time cost for reacting to a probe in the location of an angry picture. Thus, higher scores for this task represent greater attentional avoidance of angry facial expressions.

3.6.5. Correlational analyses

To explore the relationship between avoidance of threatening social stimuli and behavior in the FOE-paradigm, we correlated behavioral scores of the aggression task (mean aggression scores for each opponent; avoidance scores for each opponent; conglomerate avoidance-aggression score for each opponent) with the reaction-time scores of the AAT and the DPT.

To explore the relationship between avoidance of threatening social stimuli and behavior in the FOE-paradigm, we correlated behavioral scores of the aggression task (mean aggression scores for each opponent; summed avoidance scores for each opponent; conglomerate avoidance-aggression score for each opponent) with the reaction-time scores of the AAT and the DPT. We also correlated self-report scores with behavioral measures from the task on an exploratory basis.

3.7. Results

Of the 42 participants, 8 had to be excluded (5 due to non-deception, 1 due to incomplete task data, 2 due to extreme bias scores of ±3 SD in AAT and DPT). Hence, analyses for this experiment were performed on 34 participants (13 men).

Regarding behavior in the FoE paradigm, participants tended to select marginally higher punishments against the provoking opponent (M=4.1, SE=0.2) relative to the non-provoking one (M=3.7, SE=0.2), t_{33}=1.9, p=.06. Crucially, they avoided the provoking opponent (M=2.6, SE=0.3) more often than the non-provoking one (M=1.8, SE=0.3), t_{33}=2.1, p< .05. Participants used the avoidance option 4.3 times on average (SE=0.5; range 0-10).

We found the expected effect in the AAT, t_{33}=2.8; p<.01, such that participants showed an avoidant bias for angry faces (M=47 ms, SE=14 ms) and an approach bias toward happy faces (M=-44 ms, SE=20 ms). We did not observe the avoidant bias in the DPT (p>.2), as RTs were similar in neutral (M=365 ms, SE=8 ms) and angry trials (M=364 ms, SE=8 ms). DPT and AAT scores were uncorrelated (all p>.1).

The DPT or AAT biases for angry faces were not related to avoidance or aggression in the FOE (all p>.1). There was, however, a significant correlation between AAT scores for happy faces and avoidance against the provoking opponent (r=.43, p<.05). Namely, participants who were quicker to pull happy faces towards them, but were slower to push

them away, avoided the provoking opponent less often (Figure 3.3C). This bias was unrelated to avoidance of the non-provoking opponent (p=.9). There were no correlations between self-report data and avoidance or aggression (all p>.1).

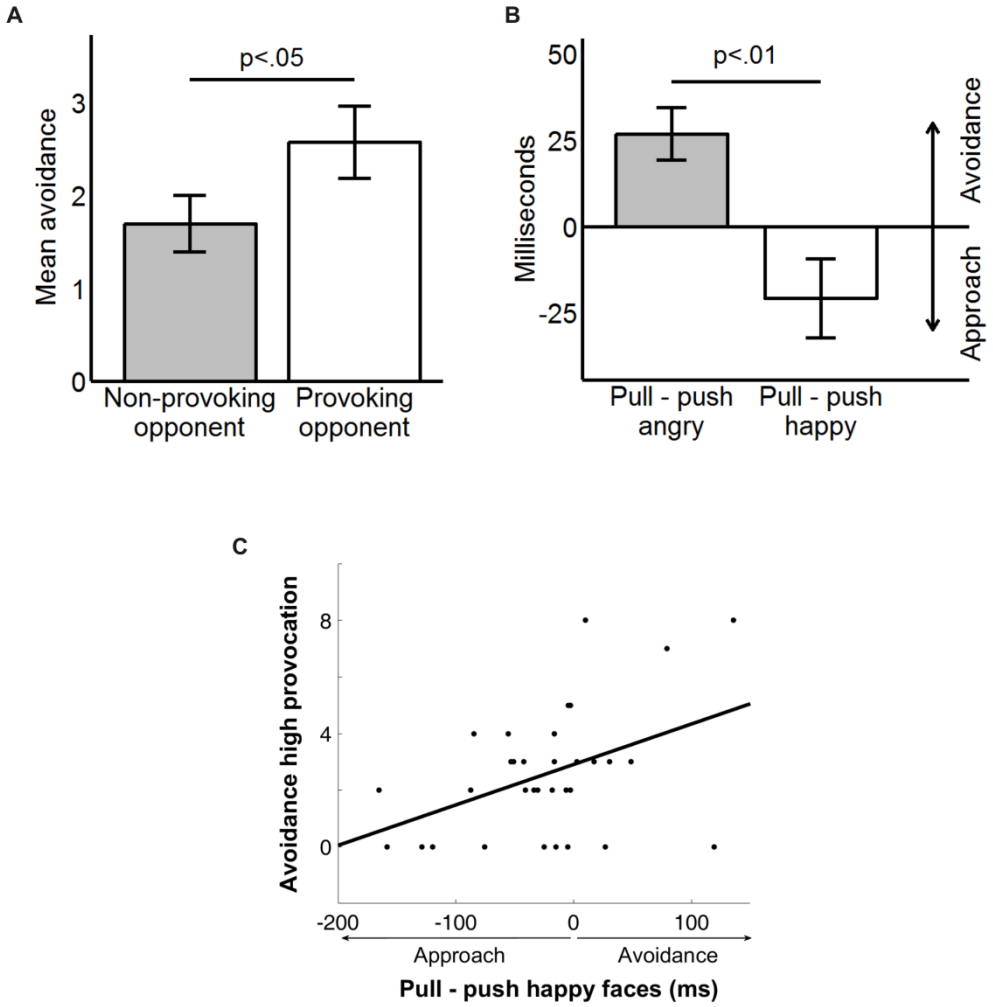

Fig. 3.3 **A**: Avoidant behavior against each opponent in the FOE. **B**: Results of the Approach-Avoidance Task (AAT). **C**: Correlation between the approach bias in the AAT (pull minus push for happy faces) and avoidant behavior against the provoking opponent in the FOE.

3.8. Discussion

In this second experiment, we modified the task to make the avoidance option more meaningful, so that participants faced each opponent alternatingly, i.e., they could not face the same opponent in two consecutive trials. As intended, this caused participants to avoid the provoking opponent more than the non-provoking one. Participants selected slightly higher

punishments against the provoking opponent than against the non-provoking one, but less so than in Experiment 1, meaning that subjects retaliated more evenly against both rivals. Given that, in Experiment 2, avoidance was a more attractive alternative strategy, and the task was more predictable, participants might have experienced an increased sense of safety and confidence. This should favor the activation of appetitive, rather than defensive, motivational systems (Lang & Bradley, 2013). Hence, perhaps aggression in Experiment 2 reflected an appetitive drive, and not so much a defensive reflex. Results of correlational analyses, which are subsequently commented, support this account.

Instead of the expected relationship between avoidant biases for angry faces and avoidance in the FOE, we found that participants with high approach scores toward happy faces engaged in more aggressive encounters in high provocation trials. Since happy faces constitute a reward signal (A. Lin, Adolphs, & Rangel, 2011; Rademacher et al., 2010), our results indicate that individuals with higher approach tendencies toward positive stimuli will tend to be more aggressive and less avoidant when provoked. This is consistent with the notion of aggression as an approach-related behavior (Berkowitz, 2012; Carver & Harmon-Jones, 2009). The fact that avoidance was related to AAT, but not to DPT scores, suggests that the escape option as implemented in the FOE reflects general avoidant tendencies, rather than implicit attentional biases.

The finding that AAT scores for happy rather than angry faces were not related to aggression deserves however further discussion. Some authors have argued that individuals showing an approach bias towards angry faces should be more aggressive, as they should be more prone to interpret angry expressions as a challenge rather than a threat, i.e. as an appetitive stimulus (Beaver, et al., 2008; van Honk, Tuiten, de Haan, van den Hout, & Stam, 2001). However, happy faces are generally less ambiguous than angry faces (Coupland et al., 2004; Parmley & Zhang, 2015), and they more clearly convey reward and positive valence

(Averbeck & Duchaine, 2009; Furl, Gallagher, & Averbeck, 2012). Hence, happy facial expressions should more consistently elicit approach motivation than angry ones. In line with this formulation, and dovetailing our findings, a recent study using the AAT in veterans found that anxious symptomatology was related to avoidance of happy faces, but not to biases toward or away from angry expressions (Clausen et al., 2016). Similarly, another study with the AAT found that approach scores towards positive stimuli predicted reactive aggression, but no effect was found for angry faces or attack-related scenes (Lobbestael, Cousijn, Brugman, & Wiers, 2016).

3.9. General discussion

Most established laboratory measures of aggression do not allow participants to avoid confrontation. We addressed this issue by developing and validating a version of the Taylor Aggression Paradigm (TAP) that included an avoidance option: the Fight-or-Escape (FOE) paradigm. In two separate experiments, we showed that reactivity to threat as measured by Fear Potentiation (FP) relates to reduced aggression and avoidance, and that those participants with stronger approach tendencies towards positive stimuli chose more frequently to engage in an aggressive interaction than participants who tended to avoid positive stimuli.

In experiment 1, participants with stronger FP responses were less aggressive on average in response to provocation. FP was also negatively related to aggression sum scores against the provoking opponent, which can be understood as a composite measure of avoidance and aggression. In our previous fMRI study, we found no relationship between threat reactivity and aggression in inescapable encounters (Beyer, et al., 2013). Here, by giving participants the possibility to avoid confrontation, we observed the previously hypothesized negative correlation between threat reactivity and aggression. Nevertheless, we found no direct relationship between FP and avoidance in the FOE (i.e. number of avoidance

options). This might be due to the fact that the avoidance option was not salient enough, as participants could face the highly provoking opponent in the trial after avoiding her.

In experiment 2, we set the task so participants played against each opponent alternatingly instead of pseudo-randomly. In so doing, the avoidance option became more meaningful, and participants avoided the provoking opponent more often than the non-provoking opponent. Crucially, subjects showing a stronger avoidant bias for happy faces used the avoidance option against the provoking opponent more frequently. We found no relationship between the AAT avoidant bias for angry faces and avoidance in the FOE. DPT scores, which represent attentional avoidance of angry faces, were also uncorrelated with behavior in the task.

Taken together, our results suggest that aggressive behavior as implemented in the FOE is driven by approach motivation to engage in aggressive interaction. Participants who tended to react to threatening stimuli with strong defensive reflexes behaved overall less aggressively towards a provoking opponent. This supports our initial theory that participants high in fear reactivity should behave less aggressively if given the opportunity to escape. On the other hand, participants with strong approach towards positive stimuli avoided the aggressive opponent less. Likely, fight-vs-flight choices in these participants were mainly driven by the prospect of potentially being able to retaliate against the aggressor, whereas participants low in behavioral approach preferred to withdraw from the aggressive interaction. These findings are in line with prominent theories of anger and aggression as driven by appetitive approach (Berkowitz, 2012; Carver & Harmon-Jones, 2009; Harmon-Jones, 2003).

3.10. Limitations and future directions

A few limitations should be mentioned. First, the avoidance option is available only *during* the game and can be used several times. In real provocation situations one can only retreat

once and usually before proper physical confrontation ensues. Moreover, there are often alternative strategies to curb the provocateur. Future studies could provide a more flexible set of responses, such as e.g. allow communication between participant and opponent, let participants choose between a social and a non-social task, offer the opportunity to compete again after the FOE, or even include an in-game option to terminate the task altogether (Ritter & Eslea, 2005). Related to this point, our sample might have suffered from range restriction due to situation selection (Dijkman & Devries, 1987), as highly avoidant subjects are unlikely to volunteer for such a study in the first place, and our participants were all healthy young students. Our task should thus be further validated on samples preselected on the basis of extreme approach and avoidance.

Chapter 4: Avoidant responses to interpersonal provocation are associated with increased amygdala and decreased mentalizing network activity[3]

4.1. Introduction

Human aggression is a complex social behavior with a profound personal and societal impact (Waters, et al., 2005). Since many instances of aggression are triggered by perceived provocation (C. A. Anderson & Bushman, 2002), many studies have investigated the social and biological factors by which individuals retaliate when provoked (Coccaro, et al., 2011; Nelson & Trainor, 2007). However, it is just as pertinent to inquire into why individuals would *avoid* confrontation, as this could provide cues on how to prevent escalation (C. A. Anderson, et al., 2008). This point has been hitherto largely overlooked. Investigating both aggressive and avoidant responses to provocation, as well as their neurobiological underpinnings, should thus help to predict the occurrence and development of aggression episodes, and ideally inform preventive and management strategies for aggressive behavior (DeWall, Anderson, & Bushman, 2011).

In laboratory aggression studies, participants are typically exposed to interpersonal threat or provocation (e.g. insults, mild electroshocks) delivered by an ostensible opponent, and experimenters measure to which extent they retaliate. Unfortunately, a non-aggressive option is not always available (Tedeschi & Quigley, 1996). Even when there is one, it usually implies not responding at all, hence not really mimicking a retreat strategy (Ritter & Eslea, 2005). Another line of research focuses on active escape, i.e. instances in which individuals must perform a task to avoid a threat. There are many studies on active escape in both humans (Löw, Weymar, & Hamm, 2015; Mobbs et al., 2009; Mobbs et al., 2007) and rodents (Bravo-

[3] Corresponds largely to Buades-Rotger, M., Beyer, F., & Krämer, U. M. (in press). Avoidant responses to interpersonal provocation are associated with increased amygdala and decreased mentalizing network activity. *eNeuro*. The study was conceived by MBR, FB and UMK. I designed the study, performed the experiments, analyzed the data, and wrote the manuscript.

Rivera, Roman-Ortiz, Montesinos-Cartagena, & Quirk, 2015; Campese et al., 2016; Ramirez, Moscarello, LeDoux, & Sears, 2015), but in these cases threat cues are often non-social, subjects cannot retaliate, and avoidance, when possible, is exclusively performance-dependent. At best, these tasks mimic encounters with predators, but not reciprocal aggressive interactions with conspecifics (Gross & Canteras, 2012).

In rodents, a circuit formed by the medial amygdala, the ventrolateral subdivision of the ventromedial hypothalamus (vlVMH) and the dorsomedial periaqueductal gray (PAG) is thought to detect danger cues from conspecifics, thereby triggering innate defensive responses (Gross & Canteras, 2012). Indeed, conspecific aggression can be optogenetically controlled by stimulating or silencing these areas (Falkner, et al., 2016; Falkner & Lin, 2014; Miczek, Takahashi, Gobrogge, Hwa, & de Almeida, 2015; Unger, et al., 2015). This neural mechanism is roughly conserved in humans, although the medial prefrontal cortex (mPFC) plays an arguably more prominent role therein (Panksepp, 2011; R. Yu, et al., 2014). Specifically, it has been suggested that mPFC encodes social dominance in concert with other brain areas involved in social cognition such as the temporal parietal junction (TPJ) or the inferior frontal gyrus (IFG) among others (Bault, Joffily, Rustichini, & Coricelli, 2011; Mason, Magee, & Fiske, 2014; Zink et al., 2008). Crucially, Ligneul et al. (2016) have recently shown that victories against a better-performing opponent recruit the mPFC, and that electrically upregulating this region potentiates dominance-based decisions. Taken together, these findings suggest that avoidant and aggressive responses to provocation might partly rely on social-cognitive processes.

In a previous fMRI study, researchers measured fear potentiation (FP) of the startle response as a measure of emotional reactivity to threat, and then set participants to play the Taylor Aggression Paradigm (TAP), an extensively employed competitive reaction time task, against two purported opponents (Beyer, Münte, Erdmann, & Krämer, 2014). In line with the

results commented in the previous paragraph, participants with higher FP had lower activity in brain areas involved in understanding others such as mPFC, TPJ, precuneus, or IFG (Schurz, et al., 2014) when confronting a highly provoking opponent. This pattern of activity in the so-called *mentalizing network* suggested that individuals high in threat reactivity cognitively disengage from the situation when provoked. This effect was nevertheless unrelated to behavior, presumably because participants did not have the chance to escape and thus overtly manifest their avoidant tendencies.

Here, we developed a version of the TAP that incorporates an avoidance option: the Fight-or-Escape (FOE) paradigm. Using the FOE, we investigated the neural correlates of aggressive and avoidant decisions against a highly and a lowly provoking opponent. Behaviorally, we expected that participants would be more aggressive against the HP than the LP, but would avoid the former more often. Drawing on the studies commented, we had two main hypotheses concerning brain activity. On a *within-participant* basis, we expected a disengagement of mentalizing regions when avoiding the HP compared to the LP. On a *between-participant* basis, we hypothesized that lower reactivity to provocation in mentalizing regions -reflecting cognitive disengagement from the aggressive interaction- would be related to escape behavior, and that this effect would be stronger for participants high in trait avoidance.

4.2. Methods

4.2.1. Participants

We recruited only female participants in order to circumvent possible gender differences in competitiveness and approach-avoidance motivation (Kivikangas, Kätsyri, Järvelä, & Ravaja, 2014). We gathered participants through flyers and emails from the local student population. One participant was excluded due to excessive head movements (>3 mm in any direction) and

three because they guessed that the paradigm was preprogrammed. Hence, the sample was comprised of 36 participants (mean age=22, SD=4) who reported to be free of psychiatric or neurological disorders. The study was approved by the university Ethics Committee and performed in accordance with the Declaration of Helsinki. Subjects provided informed consent, and were compensated for participation.

4.2.2. Procedure

On each measurement, we met participants as well as two female confederates. We told them they would play an interactive game with each other, which one of them would play from inside the scanner, and the other two with laptops connected to the scanner. Participant and confederates read the instructions together, and then the participant was taken to the scanner. We placed an MR-compatible pulse oximeter in the thumb of the left hand (see "Heart rate data acquisition"). After 4 practice trials, the functional measurement took place (≈30 minutes). The paradigm was presented through scanner-compatible goggles, with diopter-matched lenses if required. Participants' responses were recorded with two 4-button devices strapped to their waist. After the TAP, participants fulfilled a series of questionnaires, plus two computerized tasks (see "Computerized and self-report measures").

4.2.3. The Fight-or-Escape (FOE) Paradigm

The task was developed as a variant of the Taylor Aggression Paradigm or TAP (Taylor, 1967). The TAP is a widely used competitive reaction time task that elicits aggression through provocation. Here, we operationalized provocation as an aversive sound blast (i.e. a Styrofoam scratching noise). In this version of the TAP, which we called Fight-or-Escape (FOE) paradigm, participants faced each of the two purported opponents in alternating order, and had the option to avoid a limited number of trials.

In the escape phase (6 s), participants were informed of which of the two opponents they were playing with, and had to choose whether to avoid the encounter. If they did, they waited until the next trial (7 s + inter-trial interval). They could do so a maximum of 5 out of the 20 trials in each of the 3 runs. They played equally often against each opponent. If they did not avoid, they had to select the loudness of a sound blast (scale 1-8) to be later directed at their purported opponent (selection phase, 3 s). This was followed by the reaction time task (2 s), in which they had to press a button quicker than the rival when a jittered target (0.5 s) appeared. In the outcome phase (2 s), they were informed of whether they had won or lost and of the opponent's punishment selection. If they lost, they also received the corresponding sound blast. The inter-trial interval (10-12 s) had a randomized variable length (see Fig. 4.1).

The task was preprogrammed such that participants lost two thirds of the trials, equally distributed against both opponents. One opponent selected on average higher punishments (high punisher, HP; range 4-8) than the other (low punisher, LP; range 1-5). The cover story was set in the Lord of the Rings (LOTR) universe (Tolkien, 1954b) to engage participants and to make the task easier to understand. Participants were told that they would play as Frodo (protagonist of the novel), and that the opponents would play as Sauron and Saruman (antagonists). The avoidance option was phrased as "putting the Ring on" because said ring confers invisibility to its wearer in the LOTR mythology (Tolkien, 1954b). The targets in the reaction time task were 3 orc pictures from the LOTR movies, presented in random order. For half of the participants the HP was Sauron, and for the other half it was Saruman. The maximal loudness of the sound blast was adapted to each participant's tolerance. The paradigm was programmed and implemented in Presentation® (Version 16.5, www.neurobs.com).

4.2.4. Computerized and self-report measures

Approach-Avoidance Task (AAT)

This task was used to measure approach-avoidance motor responses to threat and reward signals, which were operationalized as happy and angry faces, respectively (Roelofs, Minelli, Mars, van Peer, & Toni, 2009; Volman, et al., 2011). Participants played with a joystick (Speedlink® Dark Tornado), which they either had to pull or push when seeing a happy or an angry face, depending on the condition. The size of the faces was gradually increased in size when pushing and decreased when pulling, giving the impression that the images approached or receded. Participants should thus be slower to pull angry faces "towards them", and slower to push happy faces "away", allowing to calculate an avoidant bias score from reaction times (Roelofs, et al., 2005).

Pictures were extracted from the Radboud Faces Database (Oliver Langner et al., 2010). We used photographs of 30 persons (15 female), each showing an angry facial expression in one picture, and a happy expression in the other. Pictures of 9 different individuals (4 female) were used for practice blocks. The pictures were cropped into an oval shape, removing hair, ears and neck.

In a first block, participants had to pull happy faces towards them, and push angry faces away. In a second block, the rule was reversed. The size of the faces was increased or decreased in 7 gradual steps. Images had an initial size of 6 cm (≈7.6° of the visual angle), and could be shrunk to 2 cm (≈1.2°) or enlarged to 22 cm (≈13.9°). Each trial started by pressing the joystick's trigger button, and a central fixation cross of was presented between trials. Each block consisted of 30 happy and 30 angry trials, and we used the same pictures in both blocks. Each block was preceded by a practice run, in which feedback was provided: a green check-mark for correct reactions, and a red cross for errors. The practice run for the first

block consisted of 20 trials, whereas the second consisted of 28 trials because participants had to learn the reverse rule. No feedback was provided during the task proper.

The main outcome measure was the response latency until the first movement. We rejected trials in which the first movement was in the wrong direction, as well as trials with responses shorter than 150ms or longer than twice the participant's own standard deviation. We calculated an implicit avoidant bias as the pull-push reaction time difference in angry trials, and an implicit approach bias as the pull-push reaction time difference in happy trials (Roelofs, et al., 2009). Reaction times per condition were: push angry: 573±8 ms [M±SE], pull angry: 644±16 ms, push happy: 630±15 ms, pull happy: 577±10 ms.

Dot-Probe Paradigm (DPP)

The DPP is a well-established measure of attentional avoidance (MacLeod, et al., 1986). In the version employed here, participants were presented with an angry and a neutral face, which were followed by a target (a dot) appearing either on the right or the left side of the screen. In half of the trials, the target was presented in the former position of the angry face (congruent condition) and in the other half in the previous location of the neutral face (incongruent condition). As individuals tend to initially allocate attention to threatening stimuli and then look away (Cooper & Langton, 2006), we programmed the task with a long exposition time (1s) to measure this general avoidant bias.

Face stimuli were 40 pictures extracted from a set of previously validated videos (Kircher *et al.*, 2013). The pictures were stills of 20 professional actors (9 women) displaying angry and neutral facial expressions. In each trial, two pictures of the same person with a neutral and an angry expression were presented together, to the left and right of the screen center.

Each trial began with a fixation cross being presented for a jittered interval between 500 and 1000ms. Then, two pictures (4.5 cm/≈3.6° each) of one person with a neutral and angry expression were presented for 1000ms. At picture offset, a dot appeared to the left or right, at the position where the center of the corresponding picture had been. Participants were instructed to react as quickly as possible to the dot, by pressing the A key on the computer keyboard if the dot was presented on the left, and L if it was presented on the right side.

Participants completed 80 trials in each of the two blocks. Each picture pair was presented four times with the following configurations: angry face and target on the left (congruent); angry face and target on the right (congruent); angry face on the left, target on the right (incongruent); angry face on the right, target on the left (incongruent). Scores were calculated as the difference in reaction time between congruent and incongruent correct trials, such that higher values would indicate attentional avoidance of the threatening stimulus. Reaction times per condition were: congruent: 391±10 ms, and incongruent: 390±11 ms.

Both computerized tasks were run in Presentation® (Version 17.2, www.neurobs.com) on a Dell Latitude E6400. The monitor had a resolution of 1440 x 900 pixels and a 60 Hz refresh rate. Participants' head was at a distance of ≈90 cm from the screen during the AAT and ≈70 cm during the DPP (the joystick was removed and the computer was brought closer for the DPP).

Questionnaires

Participants fulfilled the Harm Avoidance (HA) scale from the revised Temperament and Character Inventory (TCI-R) in German (Brandstrom, Richter, & Nylander, 2003), a dichotomous 35-item measure with four subscales: anticipatory worry, fear of uncertainty, shyness, and fatigability. We also used the German version of Carver and White's BIS scale (Strobel, et al., 2001), which is based on Gray's biopsychological theory of personality and is

thought to measure punishment sensitivity and general avoidant tendencies. The scale employed here has 7 items, scored on a 1 to 4 ("Completely disagree" to "Completely agree") Likert scale. In addition, participants fulfilled the German version of the Liebowitz Social Anxiety Scale (LSAS; Stangier & Heidenreich, 2005), which uses a 4-point scale to measure fear ("none" to "severe") and avoidance ("never" to "usually") of different social situations.

We used a questionnaire to check whether the experimental manipulation had succeeded. Participants rated the unpleasantness of the highest and lowest tones (scale 1-8), and the fairness of their two opponents (1-8). We assessed whether they had been successfully deceived with three qualitative questions ("Have you noticed anything special in the opponents' behavior?", "Have you followed any specific strategy during the game?", and "What do you think this study investigated?") and in the debriefing. In addition, we administered the 27-item German version of the Aggression Questionnaire (AQ; Herzberg, 2003), an ad-hoc translated German version of the Revised Competitiveness Index (RCI; Harris & Houston, 2010) with 14 items, and one extra question inquiring on weekly hours of videogame use. Scores for all questionnaires were: BIS: 19.92±0.60, LSAS-Fear: 15.44±1.65, LSAS-Avoidance: 14±1.72, HA: 12.56±1.01, RCI-Competitiveness: 28.03±1.17, RCI-Contentiousness: 15.25±0.63, and AQ: 2.00±0.06. The internal consistency of all scales was satisfactory (α= [.780, .961]).

4.2.5. Heart rate data acquisition

Photoplethysmography was performed with an Invivo Precess Model 3160 pulse oximeter attached to the thumb of the left hand. The device had the following technical specifications: saturation precision=70 to 100% ± 3%, pulse range=30 to 240 bpm ± 3 bpm, pulse accuracy=±3, LED1 wavelength= 663 nm, LED2 wavelength= 948 nm, LED1 output power= 66.9 uW, LED2 output power= 39.1 uW, pulse duration=$6.06*10^{-4}$ s.

4.2.6. Neuroimaging data acquisition

We acquired all scans with a 32-channel head coil mounted on a Philips Ingenia 3.0T scanner supporting gradient echo-planar imaging (EPI). We obtained anatomical images with a T1-weighted EPI sequence (180 sagittal slices, TR = 7.7, TE = 3.5, FOV = 240, matrix = 240 x 240 mm, flip angle = 8°, voxel size = 1 mm isotropic). For functional scans, we used a T2*-weighted EPI scanning protocol sensitive to changes in the blood-oxygen-level dependent (BOLD) signal (47 axial slices per volume, TR = 2.5 s; TE = 25 ms; FOV = 200 mm, matrix = 80 x 80 mm; flip angle = 90°; voxel size = 2.5 mm isotropic). The beginning of each trial of the task (see "Fight-or-Escape [FOE] Paradigm" section) was timed to coincide with the start of a volume in order to reduce sampling variability in the escape phase, our main epoch of interest. We recorded 3 consecutive runs of 216 volumes each (i.e. 9 minutes per run), with 5 dummy scans at the start of each run to permit steady-state tissue magnetization.

4.2.7. Behavioral data analysis

We first calculated the cumulative proportion of subjects who used up all avoidance options in each trial, which provides information on participants' overall strategy (i.e. whether they exhausted all avoidance options early in the run or saved them for later trials). Subsequent inferential statistics were computed with linear mixed-effect models (LMMs, also known as hierarchical or multilevel models), which are appropriate for unbalanced data and permit to model trial-wise behavior (Aarts, Dolan, Verhage, & van der Sluis, 2015; Baayen, Davidson, & Bates, 2008). These analyses were performed with the *lmerTest* package version 2.0-33 (Kuznetsova, Christensen, Bavay, & Brockhoff, 2015) implemented in R version 3.1.3. First, we tested the effects of provocation (coded as 0=low, 1=high) and run (coded as -1, 0 and 1 for runs one to three) on aggression and avoidance. Provocation and run were defined as fixed-effect factors, whereas subject was defined as a random factor. Additionally, we ran a

LMM on trial-wise response latencies in the escape phase with factors decision (coded as 0=fight, 1=avoid), run, and provocation, aiming to test whether participants took more time to choose the avoidance or fight options against one or the other opponent. In this model, decision was defined as a participant random-effect variable. We also modelled reaction times in the punishment selection phase and in the reaction time task with factors provocation and run. For avoidance, we fitted a generalized LMM (function *glmer*) with a logit link function for binomial outcomes, whereas all other variables were modelled with standard LMMs (function *lmer*) for continuous outcomes. We report parameter estimates (b) and their associated statistics (t/Z and p values) for all LMMs. Where appropriate, we performed pairwise post-hoc t/Z-tests (Tukey-adjusted for multiple comparisons) with the function *lsmeans* (Lenth, 2016). The means/proportions and standard errors extracted with *lsmeans* were used to plot the observed effects.

Regarding self-report data, we first analyzed the manipulation check. We inspected whether participants rated the lowest and the highest tone differently, and whether they perceived the high punisher as less fair than the low punisher by means of paired t-tests. Finally, we tested whether the mean difference in avoidance or aggression between the high and low provoker (i.e. the provocation effect) was related to any of the avoidance or control measures with Pearson correlation coefficients. We only used questionnaires' total scores, and not subscales, to avoid inflating the number of tests. Only measures related to avoidance or aggression were used to explore further brain-behavior relationships. Significance was set at $p<.05$ (uncorrected). These analyses were performed with built-in R functions.

4.2.8. Heart rate data analysis

We visually inspected the heart rate (HR) data for artifacts, which left 30 participants with complete usable data. We removed linear drifts with the *detrend* function from MATLAB

R2015b and computed HR change as the pulse count difference between the 1 s pre-stimulus baseline and each second of the escape and selection phases (Bradley, Codispoti, Cuthbert, & Lang, 2001). Since we expected an initial freezing response to provocation (Lang & Bradley, 2013), we extracted the maximum deceleration (i.e. the minimum value) across the duration of the escape phase for each trial. We then ran a series of LMMs on these scores with within-subject factors decision, run, and provocation as we did for behavioral data. In the selection phase, we expected to observe a preparatory HR acceleration (van Honk *et al.*, 2001), and so we extracted the maximum acceleration in HR relative to baseline. In this case, we fitted a run by provocation LMM.

4.2.9. Neuroimaging data analysis

We used Statistical Parametric Mapping 12 (SPM 12; Wellcome Department of Imaging Neuroscience, University College London, London, UK) implemented in MATLAB R2015b for the analysis of neuroimaging data. We first realigned all scans manually according to the anterior-posterior commissure. We applied standard preprocessing steps, namely slice-timing correction to the middle slice, realignment to the first functional volume, coregistration of anatomical and mean functional images, segmentation of the anatomical image with the standard SPM12 "Segment" function (known as "New Segment" in SPM8), normalization to the native voxel size in Montreal Neurological Insitute (MNI) space, and smoothing with an 8 mm full width at half maximum (FWHM) Gaussian kernel.

We then fitted two first level models to answer our different research questions. In the first one, we defined two 6 seconds in the decision phase for high and low provocation. We also specified two regressors in the selection phase for high and low provocation trials (3 seconds), and four 3-second regressors in the outcome phase (won and lost against the high and low provoker). We modelled as well the reaction time task (onset of the target), the sound

of the punishment in lost trials, and movement parameters derived from realignment as regressors of no interest. The onset of motor responses (i.e. button presses) rather than reaction times were also included as nuisance regressors, since modelling the latter can remove genuine decision-related activity (Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008). Regressors were convolved with the canonical hemodynamic response function, and we applied a 128s high-pass filter to remove signal drifts, as well as SPM's autoregressive function.

In order to test for differential brain reactivity to the high relative to low provocation, we performed one sample t-tests contrasting high vs low trials in the escape phase. In addition, we performed multiple regression analyses to probe whether avoidance could predict brain reactivity to provocation across participants. Hence, we regressed brain activity in the high > low contrast in the escape phase against the provocation effect for avoidance (i.e. difference in number of avoided trials between HP and LP). Although our focus was on the escape phase, we also analyzed the selection and outcome phases. In the selection phase, we compared high vs low provocation, and regressed the behavioral provocation effect for aggression (i.e. difference in mean punishment selections between HP and LP) on this contrast. In the outcome phase, we conducted a flexible factorial analysis with factors won vs lost and high vs low provocation, and tested both main effects (won > lost, lost > won, high > low, low > high) and the interaction (won high > won low: lost low > lost high and won low > won high: lost high > lost low) with t-tests.

In a second set of analyses, we investigated brain activity associated with fight versus avoidance decisions, and whether these could be modulated by provocation. In these analyses, we only included participants who avoided each opponent at least once in the same run (n=27). We did so for two reasons. First, including participants with too few trials can reduce statistical power when conducting analysis of variance (ANOVA) on unbalanced data (Tibon

& Levy, 2015). By including only participants who avoided both opponents in the same run we ensured that a minimum amount of avoidance trials per run (i.e. 2 out of 20 or 10%) was modelled and thereby achieved higher power to detect avoidance-related effects. Such an approach is common practice in fMRI studies conducting performance-dependent contrasts (Madipakkam, Rothkirch, Guggenmos, Heinz, & Sterzer, 2015; Marchewka et al., 2016; Rodehacke et al., 2014). Second, this procedure permits to create a balanced second level design, which ensures orthogonality between the different effects (McFarquhar, 2016). First-level models were fitted identically as before, but included four regressors in the escape phase defined by participants' decision: avoid high, avoid low, fight high, and fight low. Pauses (7 seconds) when using the avoidance option were also modelled separately for the HP and the LP. If participants used up all avoidance options in a given run, the escape phase was modelled as a pause, since there was no decision to be taken. If they never decided to avoid or avoided only one opponent in a given run, that run was excluded from the analysis. At the second level we performed a flexible factorial ANOVA with factors decision (avoid vs fight) and provocation (high vs low). Given that there were considerably less avoidance than fight trials, we assumed unequal variance between the levels of this factor. We then tested the main effects of decision (avoid > fight and fight > avoid) and its interaction with provocation (avoid high > avoid low: fight low > fight high and avoid low > avoid high: fight high > fight low). In all analyses, we applied a whole-brain Family-Wise Error (FWE) corrected threshold of $p<.05$ at the cluster level with a cluster-forming threshold of $p<.001$ (uncorrected). Anatomical regions were labelled according to the atlases implemented in xjView (http://www.alivelearn.net/xjview8/). We extracted mean parameter estimates for each condition within a 7.5 mm radius sphere (i.e. 3 voxels) around the peak of significant clusters for data visualization using MarsBar (http://marsbar.sourceforge.net/).

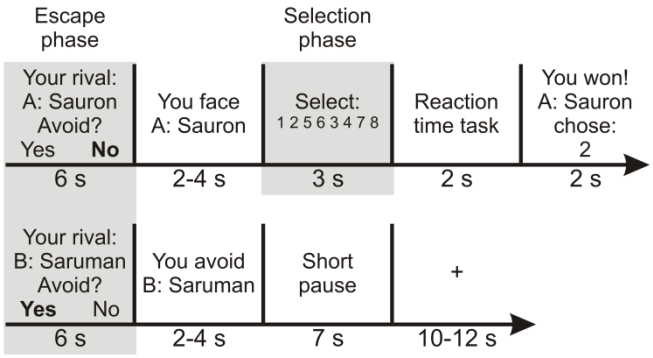

Fig. 4.1 Example trials of the Fight-or-Escape (FOE) paradigm. **A**: trial in which the participant confronted the opponent. **B**: trial in which the participant avoided the opponent. The fixation cross is only depicted for the avoidance trial, but appeared regardless of the participant's decision. See main text for details.

4.3. Results

4.3.1. Behavioral and self-report results

Participants rated the loudest tone as more distressing than the lowest one (t_{35}=18.03, p<.001), and perceived the HP to be less fair than the LP (t_{35}=9.20, p<.001). No more than 25% of participants (i.e. 9 out of 36) used up all avoidance options in any run. Furthermore, participants tended to exhaust avoidance options towards the end of each run (see Fig. 4.2A).

As predicted, participants were more likely to avoid the HP than the LP (b=0.50, Z=4.00, p<.001; Fig. 4.2B). In the punishment selection phase, they selected higher punishments against the HP compared to the LP (b=0.69, t_{1806}=8.86, p<.001; Fig. 4.3C). There were no main effects of run (both p>.097). The run by provocation interaction was significant for punishment selections (b=0.32, $t_{1804.4}$=3.40, p<.001) and near-significant for avoidance (b=0.29, Z=1.91, p=.055). Hence, we compared avoidance and aggression scores for high vs low provocation in each run to clarify whether participants learned the difference between both opponents over time. For avoidance, the difference between the HP and the LP was significant in the second (Z=2.12, p=.033) and third runs (Z=3.65, p<.001), but not in the first (p=.282). For aggression, it was significant in all runs (Run 1: $t_{1802.60}$=2.40, p=.016, Run

2: $t_{1802.88}$=5.67, p<.001, Run 3: $t_{1802.65}$=7.22, p<.001). The increasing effect sizes indicate that the provocation effect became stronger over time for both avoidance and aggression.

Participants were generally quicker to decide in the escape phase when facing the HP than the LP (b=-0.09, $t_{2091.9}$=-3.34, p<.001) and over time (b=-0.08, $t_{2085.9}$=-3.55, p<.001). Importantly, we found a significant interaction (b=0.21, $t_{2067.20}$=2.70, p=.006) between provocation and decision (Fig. 4.2D), such that participants were faster to choose the fight option against the HP than the LP ($t_{2082.95}$=-3.23, p=.001), but took a comparable amount of time to avoid each opponent (p=.112).

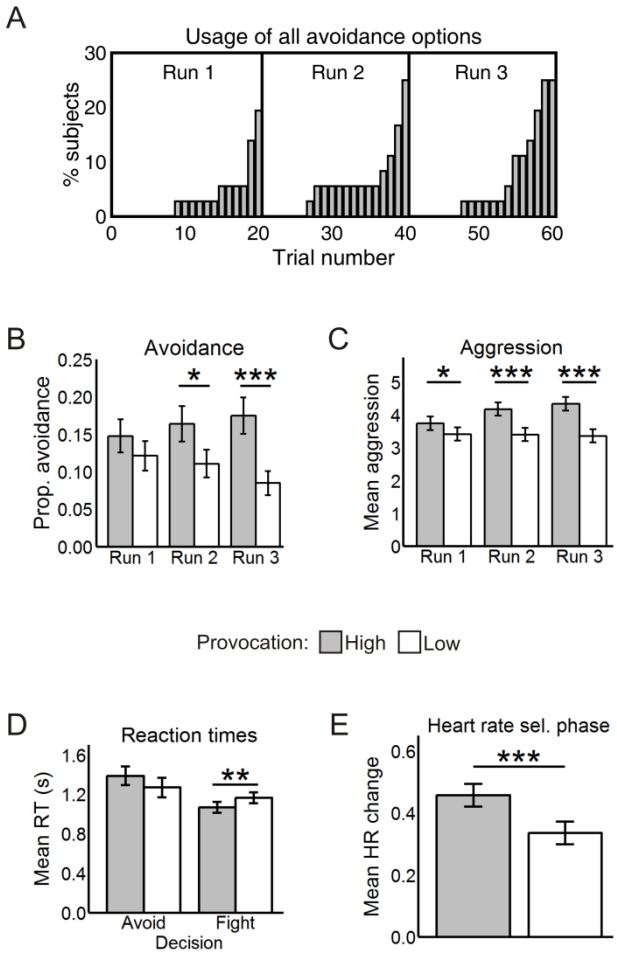

Fig. 4.2 A: Cumulative proportion of participants who used all avoidance options in each trial. **B:** Mean avoidance by run and provocation. **C:** Mean aggression by run and provocation. **D:** Reaction times in the escape phase by decision and provocation. **E:** Heart rate results in the selection phase by provocation. All values in this figure are mean ± standard error. *p<.05; **p<.01, ***p<.001.

There were no other main effects or interactions (all p>.265). Reaction times in the punishment selection phase were unaffected by run or provocation (p>.123 for main effect and interaction). In the reaction time task participants became faster over runs (b=-0.02, t_{1804}=-2.58, p=.009), but there were no main or interactive effects of provocation (both p>.634).

DPT scores, avoidance scores in the AAT, personality and control measures were not associated with the provocation effect for either avoidance or aggression (all p>.178). However, participants who rated the HP as more unfair than the LP showed stronger provocation effects for both avoidance (r=.40, p=.015) and aggression (r=.52, p=.001). Furthermore, participants with a higher approach bias in the AAT (i.e. those who were faster to pull relative to push happy faces) were, at trend level, less avoidant (r=-.32, p=.054) and more aggressive (r=.32, p=.052) when facing the HP relative to the LP.

4.3.2. Heart rate results

Regarding the escape phase, we found no differences between opponents in HR reactivity (p=.362), nor a run by provocation interaction (p=.275). HR did not differ either between fight and avoid decisions (p=.605), nor was there a decision by provocation (p=.284) or three-way interaction (p=.225). In the selection phase, we found a main effect of provocation (b=0.12, t_{1507}=4.05, p<.001; Fig. 4.2E) such that participants had a higher HR increase relative to baseline in HP (0.46±0.03) than in LP trials (0.34±0.03). Time had no main or interactive effects (both p>.341).

4.3.3. Neuroimaging results

Neural reactivity to provocation in the escape phase

No region was differentially active in the high > low or low > high comparisons in the escape phase. In regression analyses, we did not find any association between brain activity in this contrast and the provocation effect for avoidance across participants.

Neural activation in fight vs avoid decisions

When participants decided to avoid, we observed widespread bilateral activation across the superior temporal sulcus (STS), medial prefrontal cortex (mPFC), inferior frontal gyrus (IFG), and posterior cingulate cortex (PCC) extending to the ventral precuneus among other regions (Table 4.1a and Fig. 4.3A). When participants decided to fight, we found increased activation in motor and somatosensory cortex, left orbitofrontal cortex (OFC), dorsal precuneus, ventral thalamus, and middle occipital lobe (Table 4.1b; Fig. 4.3B).

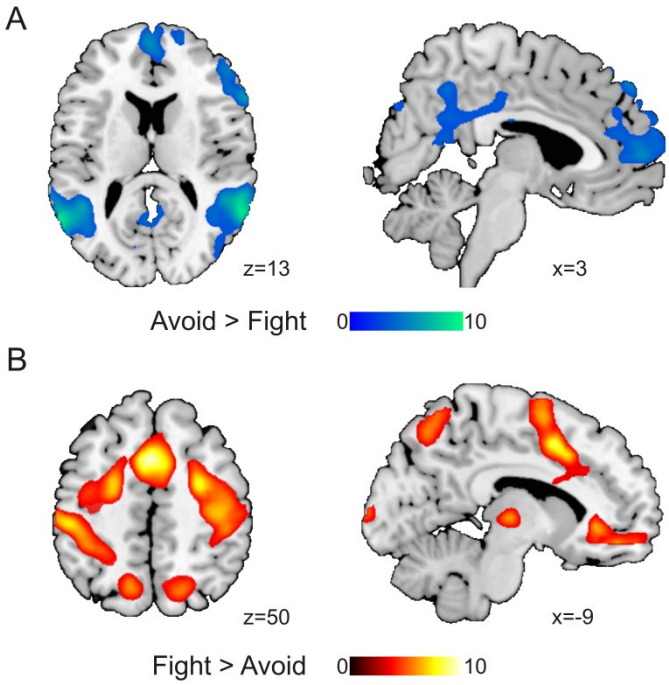

Fig. 4.3 Avoid > Fight contrast. Avoid decisions were linked with activation in regions such as the medial frontal gyrus (mPFC), the temporo-parietal junction (TPJ), the posterior cingulate cortex (PCC) extending into the ventral precuneus, and the right inferior frontal gyrus (IFG). **B:** Fight > Avoid contrast. Fight decisions were associated with increased activation in bilateral somatomotor cortex, orbitofrontal cortex (OFC), ventral thalamus, and dorsal precuneus. Statistical parametric maps are thresholded and presented at p<.001, pFWE<.05 cluster-level corrected.

Interaction between provocation and decision

The right amygdala, the cuneus extending to the middle precuneus, and bilateral posterior cerebellum showed increased activation when participants chose the avoidance option in the high compared to low provocation condition (Table 4.1c, Fig. 4.4). The intraparietal sulcus extending to the supramarginal gyrus (SMG), middle frontal gyrus, mPFC, IFG, inferior parietal lobule (IPL) covering the temporo-parietal junction (TPJ), and sgACC showed lower activation when participants avoided the HP relative to the the LP (Table 4.1d; Fig. 4.5).

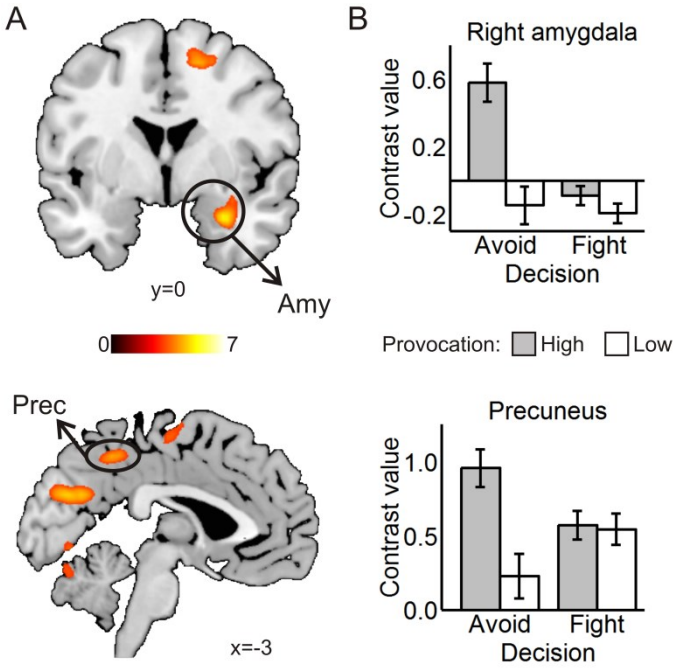

Fig. 4.4 Avoid high > Avoid low : Fight low > Fight high contrast. **A:** amygdala (Amy), precuneus (Prec), cuneus, and posterior cerebellar clusters. Statistical parametric maps are thresholded and presented at p<.001, pFWE<.05 cluster-level corrected. **B:** Contrast values by decision and provocation in basolateral amygdala (above) and precuneus (below). Values are mean ± SE within a 7.5 mm sphere around the local peak.

Neural reactivity to provocation in the selection phase

In the punishment selection phase, we found increased reactivity to provocation (high > low) in inferior frontal gyrus (IFG), inferior parietal lobule (IPL) covering the temporo-parietal junction (TPJ), posterior cingulate cortex (PCC), precuneus, and the brainstem peaking in the

periaqueductal gray (PAG) but including the red nucleus and the ventral thalamus (Table 4.2a; Fig. 4.6A and 4.6B). No clusters survived in the opposite contrast (low > high), and no region was associated with the provocation effect for aggression across participants.

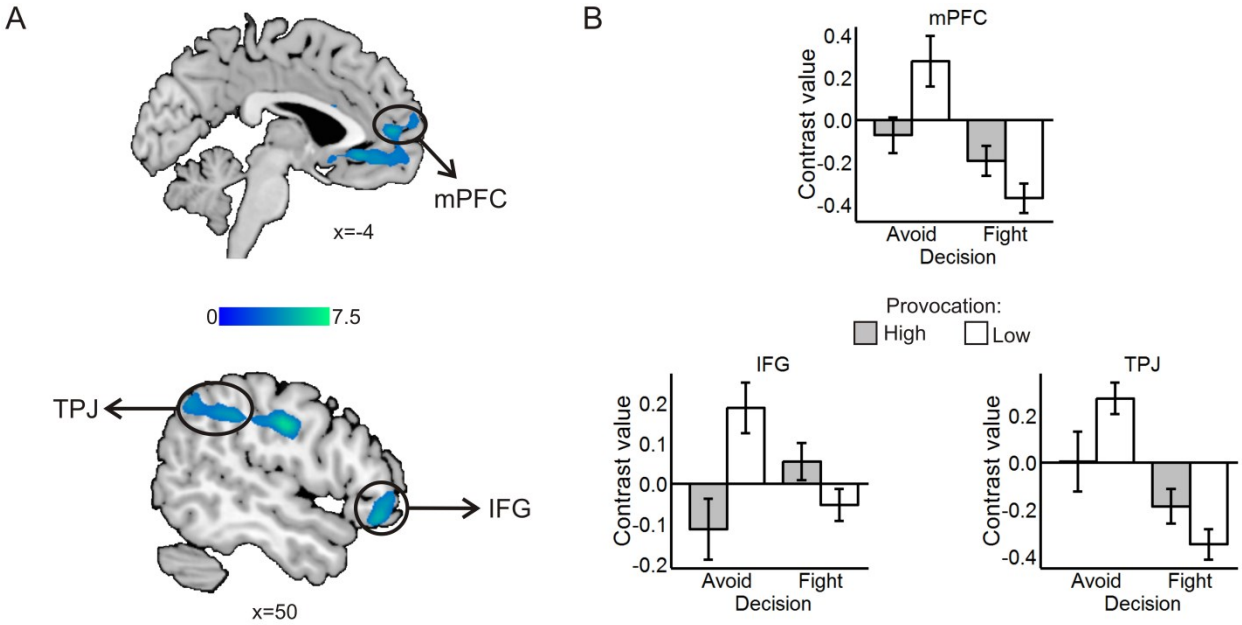

Fig. 4.5 Avoid low > Avoid high : Fight high > Fight low contrast. **A:** clusters in subgenual anterior cingulate cortex (sgACC) extending into the rostral medial prefrontal cortex (mPFC), in temporo-parietal junction (TPJ), and in inferior frontal gyrus (IFG). Statistical parametric maps are thresholded and presented at p<.001, pFWE<.05 cluster-level corrected. **B:** Contrast values by decision and provocation in mPFC (above), IFG (below left) and TPJ (below right). Values are mean ± SE within a 7.5 mm sphere around the local peak.

Brain activity in the outcome phase

In the outcome phase, won relative to lost trials elicited brain activity in the ventral striatum (VS; Table 4.2b; Fig. 4.6C) as well as in a number of cortical areas, chiefly in the dorsolateral prefrontal and posterior parietal cortex. Lost relative to won trials were associated with large clusters across the superior temporal gyrus, as well as with activation in the lingual gyrus and the anterior insula (AI; Table 4.2c; Fig. 4.6D). There were no main or interactive effects of provocation in the outcome phase.

Table 4.1. brain activity in the escape phase (n=27).

Region/*Contrast*	k	Peak T	x	y	z
a) Avoid > Fight					
Superior temporal gyrus	10182	9.63	65.5	-49.5	20
		9.49	50.5	-42	20
		9.23	58	-52	10
Inferior frontal gyrus	322	6.43	63	23	12.5
		5.42	55.5	25.5	5
		4.84	50.5	40.5	12.5
Middle frontal gyrus	1067	6.29	25.5	25.5	37.5
Superior frontal gyrus		5.88	18	53	40
Medial frontal gyrus		5.85	0.5	58	7.5
b) Fight > Avoid					
Superior frontal gyrus	9856	10.09	-22	-2	65
		9.69	28	-7	57.5
Supplementary motor area		9.60	-7	10.5	45
Anterior cerebellum	193	7.81	30.5	-52	-32.5
		5.55	20.5	-54.5	-22.5
Precuneus	276	7.43	18	-67	55
Ventral thalamus	156	6.06	-9.5	-17	2.5
		4.22	0.5	-29.5	2.5
		3.47	3	-19.5	-5
Superior occipital gyrus	164	5.88	18	-97	15
		5.18	13	-99.5	2.5
c) Avoid high > Avoid low : Fight low > Fight high					
Anterior cerebellum	648	7.33	40.5	-52	-30
Posterior cerebellum		4.55	8	-72	-20
		4.26	18	-72	-22.5
Cuneus	2747	6.67	-17	-57	25
Middle occipital gyrus		6.66	-32	-74.5	25
Cuneus		6.18	13	-79.5	27.5
Posterior cerebellum	287	5.53	-27	-69.5	-27.5
		4.97	-34.5	-52	-30

Lingual gyrus	308	5.45	13	-47	2.5
		4.58	23	-64.5	-5
		3.82	23	-49.5	7.5
Amygdala	182	5.34	30.5	3	-22.5
Temporal pole		5.13	43	10.5	-17.5
Inferior frontal cortex		4.25	30.5	15.5	-22.5
Supplementary motor area	259	4.78	28	10.5	62.5
		4.57	13	-4.5	60
		4.33	13	-14.5	60
d) Avoid low > Avoid high : Fight high > Fight low					
Middle frontal gyrus	4812	7.48	28	53	0
Middle frontal gyrus		6.99	-24.5	40.5	25
Postcentral gyrus		6.30	-32	-22	35
Postcentral gyrus	238	6.72	53	-9.5	35
		5.39	63	-4.5	30
		4.36	55.5	-9.5	22.5
Inferior parietal lobe	232	5.66	55.5	-57	42.5
		5.00	53	-37	40
		4.68	55.5	-64.5	30

Results reported at p<.001 (uncorrected), pFWE<.05 cluster-wise corrected. Coordinates are in MNI space. Clusters ordered by peak T values. Voxel size=2.5 mm isotropic. K: cluster size.

4.4. Discussion

We investigated the neural correlates of fight and avoidance decisions in response to provocation. Participants avoided the HP more often, but selected louder sound blasts when confronting her. Nevertheless, these behavioral provocation effects were not related to computerized and self-report measures of avoidance on a between-participant basis. Fight decisions yielded increased activation in OFC, dorsal precuneus, and the sensorimotor cortex among other regions. Avoidance decisions were accompanied by increased activity in the mentalizing network, but this effect was less pronounced in high provocation trials. On the other hand, amygdala, precuneus, and posterior cerebellum were more active when avoiding

the HP compared to the LP. Our study therefore identifies direct neural correlates of fight-or-flight decisions, and helps to delineate the contribution of OFC, amygdala, and mentalizing regions in aggressive and avoidant responses to provocation.

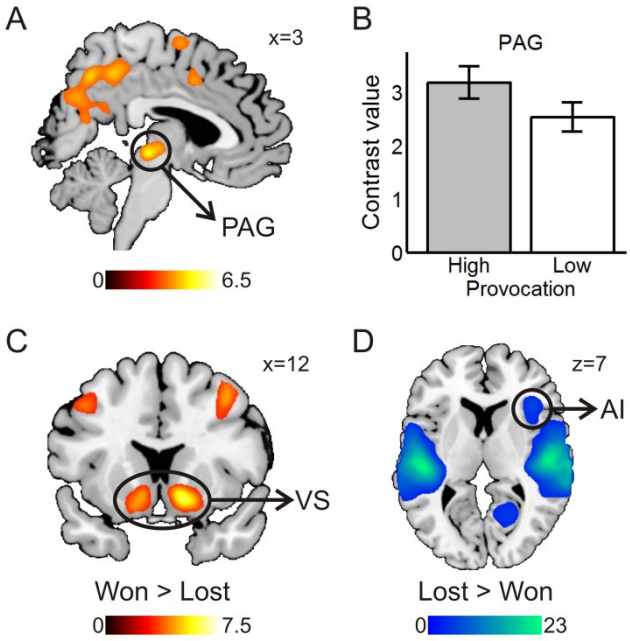

Fig. 4.6 A: High > Low provocation contrast in the punishment selection phase. High provocation elicited activation in periaqueductal gray (PAG) extending to the ventral thalamus, precuneus, supplementary motor area (SMA), temporo-parietal junction (TPJ), and inferior frontal gyrus (IFG) among others. **B:** Contrast values by provocation in the PAG. Values are mean ± SE within a 7.5 mm sphere around the local peak. **C**: Won > Lost contrast in the outcome phase. Winning was associated with activity in ventral striatum (VS), middle frontal and inferior occipital areas among others. **D**: Lost > Won contrast in the outcome phase. Losing was related to activity in anterior insula (AI), superior temporal gyrus and lingual gyrus among other regions. Statistical parametric maps are thresholded and presented at p<.001, pFWE<.05 cluster-level corrected.

Table 4.2. Brain activity in the selection and outcome phases (n=36)

Region/*Contrast*	k	Peak T	x	y	z
a) High > Low selection phase					
Inferior frontal gyrus	790	6.50	50.5	20.5	5
		5.48	43	30.5	-2.5
		5.00	28	-4.5	2.5
	429	5.53	43	-52	45
		4.22	55.5	-47	37.5
		4.11	38	-67	55
Posterior cingulate cortex	1691	5.47	13	-44.5	35

Calcarine sulcus		5.24	-14.5	-62	17.5
Precentral gyrus		4.79	-39.5	-24.5	57.5
Periaqueductal gray	218	5.39	15.5	-24.5	-15
		5.16	3	-24.5	-5
Putamen	129	4.94	25.5	8	-10
		3.68	18	0.5	-15
Supplementary motor area	241	4.90	10.5	-2	67.5
		4.21	8	10.5	62.5
		3.86	-2	-4.5	65
Orbitofrontal cortex	99	4.88	-32	18	-17.5
		4.21	-32	33	-10
Middle occipital gyrus	93	4.41	-47	-84.5	7.5
Middle temporal gyrus		4.34	-59.5	-69.5	5
		3.91	-52	-74.5	10
Midcingulate cortex	82	4.19	8	5.5	40
		3.73	-9.5	8	40
b) Won > Lost outcome phase					
Inferior occipital gyrus	805	7.48	-19.5	-97	-5
		5.60	-27	-94.5	12.5
		5.11	-17	-99.5	12.5
Fusiform gyrus	1568	6.82	35.5	-44.5	-22.5
Inferior occipital gyrus		6.23	28	-92	-5
		5.73	20.5	-97	0
Ventral striatum	246	6.49	13	13	-10
Supramarginal gyrus	724	6.06	-47	-67	35
		4.96	-54.5	-49.5	50
		4.63	-44.5	-57	42.5
Superior temporal gyrus	529	5.67	55.5	-59.5	32.5
		5.06	50.5	-62	45
		3.95	55.5	-47	50
Inferior frontal gyrus	259	5.29	50.5	38	-10
		4.49	40.5	55.5	-7.5
Middle frontal gyrus	1763	5.04	-29.5	23	45

Medial prefrontal cortex		4.68	-4.5	60.5	15
Middle frontal gyrus		4.54	-14.5	43	42.5
Ventral striatum	192	4.90	-12	3	-15
		4.68	-12	13	-12.5
		3.33	-9.5	18	2.5
Middle frontal gyrus	537	4.67	38	10.5	47.5
		4.57	23	23	57.5
		4.08	13	35.5	52.5
Inferior frontal gyrus	463	4.60	-37	50.5	-7.5
		4.60	-47	43	-10
		4.30	-54.5	30.5	2.5
c) Lost > Won outcome phase					
Superior temporal gyrus	3055	22.95	-47	-24.5	7.5
		18.72	-54.5	-32	12.5
		17.28	-39.5	-32	12.5
Superior temporal gyrus	3331	22.71	50.5	-19.5	7.5
		19.21	65.5	-27	12.5
Anterior insula		6.58	38	25.5	5
Lingual gyrus	293	5.19	20.5	-59.5	5
		3.55	15.5	-72	25

Results reported at p<.001 (uncorrected), pFWE<.05 cluster-wise corrected. Coordinates are in MNI space. Clusters ordered by peak T values. Voxel size=2.5 mm isotropic. K: cluster size.

4.4.1. Behavioral and cardiac responses to provocation

Participants avoided the HP more frequently and selected higher punishments against her than against the LP. They were also generally quicker to decide when facing the HP, and more so when choosing to fight her. In the selection phase and reaction time task, subjects became quicker over time independently of the opponent or the decision, implying that they learned the task dynamics. Importantly, participants did not generally use all avoidance options. When they did, it was in the last trials, when the end of the run neared. This suggests that participants might have also pondered whether to use up avoidance options or save them for

later, which, as pointed out by an anonymous reviewer, adds another layer of cognitive complexity to the decision process. We did not find the expected deceleration in HR in the escape phase, likely because having the opportunity to escape conferred participants a sense of safety (Lang & Bradley, 2013). In the selection phase, however, heart rate increased when facing the HP, reflecting a typical *circa-strike* physiological response to an incoming threat (Haller et al., 2014; Lang & Bradley, 2013). The magnitude of this response was small due to the short duration of this epoch (i.e. 3 seconds). Overall, results suggest that the FOE paradigm was able to successfully provoke participants and motivated genuine fight and flight responses.

There was no direct relationship between any measure of trait avoidance and avoidant behavior in the FOE. Rather, avoidance and aggression were related to fairness ratings, indicating that participants' behavior depended more on transient appraisals of the opponent than on broad avoidant tendencies. Alternatively, variability in avoidance might have been too low to detect correlations with personality (Bates, Zhang, Dufek, & Chen, 1996). It can also be that the computerized avoidance tasks inadequately captured the construct of interest, given that the DPT has shown poor reliability in non-clinical samples (Schmukle, 2005). Notably, the AAT approach bias toward happy faces was marginally associated with decreased avoidance and increased aggression against the HP relative to the LP. This indicates that participants with high approach motivation were more sensitive to provocation, and is in line with theories suggesting than anger and aggression are approach-driven behaviors (Berkowitz, 2012; Carver & Harmon-Jones, 2009). The finding that participants were faster to select the fight –but not the avoidance- option against the HP also fits this notion. Further developments of the FOE could include e.g. a shorter time limit to decide and/or additional tradeoffs for fight and avoid decisions. Such constraints would better recreate real world aggression episodes, in which responses to provocation are likely adopted

after one-shot, impulsive decisions (Lowe & May, 2011; Simons, Wills, Emery, & Spelman, 2015).

4.4.2. Neural responses to provocation

We found no general differences in neural activation between the HP and the LP in the escape phase despite the stark contrast in behavior towards each opponent. In previous studies, differential brain reactivity to provocation was only observed when deciding the punishment intensity (Beyer, Münte, Erdmann, et al., 2014; Krämer, et al., 2007), which presumably entails qualitatively different cognitive processes than choosing whether to avoid or fight. The latter decision might have been complex enough to constrict differences in neural reactivity to provocation, which were only visible within avoid and fight trials. This is supported by the widespread brain activity observed in the avoid vs fight contrasts, and the greater reaction times in avoid relative to fight decisions.

When selecting the punishment, some areas of the mentalizing network (IFG, TPJ, precuneus) were activated in high relative to low provocation trials. This could reflect a more intense deliberation of the consequences of punishing the HP, as is often assumed in competitive or bargaining paradigms (Assaf et al., 2009; Krach et al., 2008; Polosan et al., 2011). This effect was accompanied by increased activity in the PAG extending to the ventral thalamus and other midbrain nuclei, an established defensive reaction to imminent threat (Mobbs, et al., 2009; Mobbs, et al., 2007). Remarkably, no mPFC activity was observed in this epoch. This finding is in agreement with studies reporting a shift from prefrontal to subcortical activation as a function of threat proximity (Mobbs, et al., 2009; Mobbs, et al., 2007). Across studies, the mPFC is less strongly recruited in tasks involving quick, one-shot judgements, than in tasks that require inferring stable characteristics (Schurz, et al., 2014). Hence, provocation seems to foster alertness and rapid social-cognitive processing in the

seconds preceding the aggressive encounter. It is noticeable that other studies with the TAP did not report PAG reactivity to provocation during punishment selection (Beyer, Münte, Erdmann, et al., 2014; Krämer, et al., 2007). This might be due to the fact that the punishment selection phase here was shorter (3 vs 6 seconds), so the sense of incoming threat was probably heightened.

4.4.3. Neural activation underlying fight decisions

When participants decided to fight, we observed increased activity in OFC, sensorimotor regions, the ventral thalamus, and the bilateral precuneus. Somatomotor activation in retaliatory decisions could reflect preparatory processes, which might have been exacerbated by the potential value of the task's outcomes (Iyer, Lindner, Kagan, & Andersen, 2010). Importantly, this motor activity is unlikely due to differences in reaction times between fight and avoid decisions, which were small in absolute value (around 115 ms) and had been controlled for by including button presses as a nuisance regressor. Medial OFC activation was also increased when participants decided to fight relative to avoid. This shows that the function of the OFC in aggression is not confined to impulse control (Blair, 2001; Mehta & Beer, 2009). Instead, our results indicate that this area accomplishes a more general evaluative role (Stalnaker, et al., 2015). From this perspective, the observed OFC activation could correspond to threat assessment (Beyer, et al., 2015), and/or vindictive approach motivation (Seymour, et al., 2007). Notably, the ventral thalamus was recruited in fight relative to avoid decisions. This subregion has been proposed to integrate motivational and motor proprioceptive inputs, thereby contributing to action selection (Bosch-Bouju, Hyland, & Parr-Brownlie, 2013). The above commented somatomotor activity and the fact that the PAG cluster observed in the selection phase extended into the ventral thalamus concur with this interpretation.

4.4.5. Neural activation underlying avoid decisions

In a previous fMRI study from our group, participants with high emotional reactivity to threat had less mentalizing network activity when facing a provoking opponent, but there was no relationship between this effect and aggressive behavior (Beyer, Münte, Erdmann, et al., 2014). Here, by giving participants an escape option, we expected to observe a direct link between threat-induced deactivation of mentalizing regions and active avoidance. Indeed, activity in mentalizing and "mirror neuron" regions was generally increased when deciding to avoid an aggressive encounter, but was relatively reduced when avoiding the HP. This suggests that participants engaged in mentalizing processes during avoidance decisions, but disengaged from the situation when they perceived high threat, i.e. in HP trials. This is consistent with the reduced reaction times for these trials, with the increased amygdala activation, and with studies showing that social stress disrupts social cognition both at the behavioral (Smeets, Dziobek, & Wolf, 2009) and neural level (Nolte et al., 2013). Nonetheless, our interpretation is partly based on reverse inference (Hutzler, 2014), and so alternative explanations, such as e.g. reduced cognitive effort (Halko, Hlushchuk, Hari, & Schürmann, 2009) could also account for these effects. In fact, considering that the escape phase was relatively long (i.e. 6 seconds), we might have captured not only decision-related but also post-decision cognitive processes. This might especially concern activation in the rostral prefrontal cortex and the inferior parietal lobe, which increases during post-decision evaluation and correlates with self-reported uncertainty about the chosen option (Wan, Cheng, & Tanaka, 2016). Provocation might thus impair the cognitive processes leading to avoidance decisions as well as the reevaluation of such decisions. On the other hand, mPFC and IFG activity in avoidance decisions could also correspond to the experience of safety, given that relief from pain has been associated with increased activation in these areas (Leknes, Lee, Berna, Andersson, & Tracey, 2011). If that was the case, the sense of relief and the

corresponding BOLD signal in these regions should be greater when avoiding the HP, but we observed the opposite pattern. Relief is therefore unlikely to explain the present results. All in all, we deem it reasonable to assume that activity in these areas corresponds at least in part to social-cognitive processes, although their exact nature and timing (i.e. pre- vs post-decision) cannot be ultimately clarified with the present data.

Remarkably, regions typically regarded as part the mirror neuron system such as the anterior SMG showed a similar pattern of activity in avoid decisions as mentalizing ones. Mirror neurons are thought to be involved in automatic action perception, whereas mentalizing regions contribute to the more complex understanding of others' cognitive and emotional state (D. Y. J. Yang, Rosenblau, Keifer, & Pelphrey, 2015). Although some authors argue that these processes are independent (Catmur, 2015), others construe mirror activity as necessary for higher-level mentalizing inferences (Tidoni & Candidi, 2016). Our data suggests that both systems have a convergent role in deciding how individuals respond to provocation. Supporting this formulation, mirror and mentalizing systems increase their coupling during real-time social interactions (Sperduti, Guionnet, Fossati, & Nadel, 2014), and reactivity to emotional stimuli in the precentral gyrus –a motor mirror region- has been related to aggression (Beyer, Münte, & Krämer, 2014).

Unlike the rest of the mentalizing network, we observed a spatial gradient in the precuneus. Dorsal regions were recruited in fight responses, ventral parts were involved in avoid decisions, and the middle area was specifically activated when avoiding and when about to face the HP relative to the LP. Our data agrees with the proposed functional segregation between a *sensorimotor* and a *limbic* precuneus (Margulies et al., 2009), and suggests that the differential activation of these subareas might contribute to either aggressive or avoidant responses to provocation.

We found that the right basolateral amygdala was more active when avoiding a highly relative to a lowly provoking opponent. Amygdala reactivity to threat has been linked to aggression (Gospic, et al., 2011; Lotze, et al., 2007; McCloskey, et al., 2016). Albeit this is often interpreted as an approach-driven phenomenon (Beaver, et al., 2008), the paradigms used in these studies lack an avoidance option. The present data imply that the amygdala is not involved in either approach or avoidance *per se* (Fernando, Murray, & Milton, 2013; Weymar & Schwabe, 2016), but potentiates defensive behavior adaptively. If escape is possible, amygdala activation will favor avoidance, if not, it will facilitate aggression (Lang & Bradley, 2013; LeDoux, 2003). Furthermore, our results indicate that this structure can signal threat *proactively,* thereby contributing to controlled decision processes (Pessoa, 2010).

It is also worth noting that posterior cerebellar activity closely resembled that of the amygdala, rather than that of the motor cortex. Activation in the cerebellum was highest when avoiding the HP, and its peak was located on its posterior aspect extending to the vermis. The latter subregion is often termed *limbic* cerebellum (Schmahmann, Weilburg, & Sherman, 2007), and is thought to be involved in inferential social-cognitive processes (Van Overwalle, Baetens, Mariën, & Vandekerckhove, 2014). However, since the function of the cerebellum in higher-order cognition is still far from clear (Koziol *et al.*, 2014), its precise contribution to avoidant behavior cannot be readily delimited.

In summary, participants showed increased amygdala activity, reduced mentalizing network activity, and short reaction times in trials in which they avoided the HP. The most plausible explanation of these results is that, on a trial-wise basis, perceived threat caused participants to avoid the opponent both cognitively (not thinking about her intentions) and behaviorally (choosing not to engage in the confrontation). This is consistent with the proposed role of the amygdala in coordinating cortical responses to threat (Pessoa & Adolphs, 2010).

4.4.6. Brain activity in the outcome phase

Winning relative to losing was linked to activity in the VS. This midbrain dopaminergic structure is thought to code for both general and social rewards (Sescousse, Caldú, Segura, & Dreher, 2013). Indeed, VS activation after wins is a highly robust finding in fMRI studies with the TAP (Beyer, Münte, Erdmann, et al., 2014; Brunnlieb, et al., 2013; Emmerling, et al., 2016; Krämer, et al., 2007) as well as in other competitive paradigms (Bault, et al., 2011; Delgado, Schotter, Ozbay, & Phelps, 2008; Kätsyri, Hari, Ravaja, & Nummenmaa, 2013). Losing was linked to large clusters of activity in the superior temporal gyrus, peaking in the auditory cortex and hence attributable to the sound blast. However, we also observed a defeat-related cluster on the AI, again mimicking previous findings (Beyer, Münte, Erdmann, et al., 2014; Krämer, et al., 2007) and presumably corresponding to the obnoxiousness of the punishment (Lamm, Decety, & Singer, 2011). This might also account for lingual gyrus activity, which has been related to aversive learning with auditory stimuli (Gu et al., 2016; McTeague, Gruss, & Keil, 2015; Miskovic & Keil, 2014). Results in the outcome phase thus replicated established effects, arguing for the reliability of the paradigm.

4.4.7. Limitations

Some shortcomings of the study should be kept in mind. First, the limited temporal resolution of fMRI does not permit to reliably isolate brain activity preceding the decisions, which were taken on average in little more than 1 second. This also implies, as commented earlier, that brain activity in the escape phase might partly reflect post-decision cognitive processes. Electroencephalography could better track the temporal dynamics of such a quick process (Gluth, Rieskamp, & Büchel, 2013). Note also that onsets and data acquisition in the escape phase were synchronous, which might compromise statistical power. We did so because the crucial contrasts in this study involved a reduced number of trials, so that making onsets and

acquisition asynchronous could add unwanted variability. Furthermore, at relatively short sampling rates (i.e. below 2.68 seconds), both methods yield similar estimates of the BOLD response (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000). Second, as our sample was limited to healthy young women, it should be tested whether similar results can be obtained in other populations. Third, we used heart rate to keep comparability with previous studies (Lang & Bradley, 2013; van Honk, Tuiten, Hermans, et al., 2001), but the relationship between vagal input and heart rate is not linear and might be better captured by heart period (Berntson, Cacioppo, & Quigley, 1995). Fourth, our paradigm did not tap important variables that can influence how one responds to provocation, such as presence of bystanders (Vasquez, et al., 2013). This is an interesting venue for future studies with the FOE.

4.5. Conclusions

In the present study we explored aggressive and avoidant reactions to interpersonal provocation, as well as their underlying neurophysiological signature. Cardiac and behavioral data suggest that the FOE, our newly developed competitive task, successfully provoked participants. We showed that avoidance was related to activity in mentalizing regions, whereas retaliation was associated with OFC and somatomotor activation. Activity in some areas was however modulated by provocation. Specifically, the right amygdala was upregulated when avoiding a provoking opponent, whereas certain mentalizing and mirror regions (mPFC, TPJ, IFG, SMG) showed relatively decreased activation. Taken together, our results indicate that avoidant responses to provocation might stem from anticipatory threat signaling and are associated with reduced perspective-taking. Moreover, our study suggests that threat escapability is a major situational factor that should be considered in laboratory measures of aggression.

Chapter 5: Ventral striatum connectivity during physical aggression[4]

5.1. Introduction

To understand aggression it is not only important to study its antecedents, but also its aftermath (C. A. Anderson & Bushman, 2002). How rewarding the outcome of an aggressive interchange is for an individual may induce him/her to refrain from further aggressive behavior, or, on the opposite extreme, to hound the opponent, which may ultimately lead to violence escalation (C. A. Anderson, et al., 2008). This applies not exclusively to the specific instance in which aggression occurs, but also translates to other situations through effects on motivation (e.g. anticipating the outcome of an upcoming aggressive encounter) and learning (e.g. repeatedly experiencing punishment or reward in aggressive interactions).

The ventral striatum (VS) is one of the main target sites of midbrain dopaminergic projections and is considered to be central in reward processing (Camara, Rodriguez-Fornells, Ye, & Münte, 2009; Haber & Knutson, 2010; McClure, York, & Montague, 2004). Functional magnetic resonance imaging (fMRI) studies have shown that the VS is robustly activated by both primary and secondary reinforcers (Sescousse, et al., 2013). Nevertheless, increases in VS activation have also been reported when presenting attractive (Bray & O'Doherty, 2007) or happy faces (A. Lin, et al., 2011), and even with more abstract social rewards such as positive impressions from others (Izuma, Saito, & Sadato, 2008). Whether the valuation of social and non-social rewards is subserved by similar structures or computational principles is still an open debate, but there seems to be a remarkable degree of overlap between the two (Ruff & Fehr, 2014).

[4] This chapter largely corresponds to Buades-Rotger, M., Brunnlieb, C., Münte, T. F., Heldmann, M., & Krämer, U. M. (2016). Winning is not enough: ventral striatum connectivity during physical aggression. *Brain Imaging and Behavior, 10*(1), 105-114. doi: 10.1007/s11682-015-9370-z. The study was conceived by CB, MH, and UMK, and conducted by CB. I re-analyzed the data and wrote the manuscript.

In paradigms setting a competitive context, fMRI studies demonstrated that VS activity is enhanced when participants win (Delgado, et al., 2008), but even more so when their earnings are higher than their competitors' (Bault, et al., 2011; Dvash, Gilam, Ben-Ze'ev, Hendler, & Shamay-Tsoory, 2010; Fliessbach et al., 2007). This latter effect can be accompanied by greater feelings of *Schadenfreude* (Bault, Coricelli, & Rustichini, 2008), suggesting that a rival's suffering could make a victory more pleasant. Consistent with this notion, VS activation while watching an unfair opponent receives an electric shock correlates positively with desires of revenge (Singer, et al., 2006). Likewise, the pleasure experienced from others' misfortune is associated with VS activation when reading their story (H. Takahashi et al., 2009).

Comparable findings were made in fMRI studies using the Taylor Aggression Paradigm (TAP; Taylor, 1967), a competitive reaction time game in which players can physically punish their opponent when winning a trial. Importantly, the opponent is usually a confederate of the researchers and the course of the game is preprogrammed. In fMRI studies with the TAP, VS activation after wins has been consistently observed (Brunnlieb, et al., 2013; Krämer, et al., 2007). Nonetheless, this raises an interpretative challenge: is victory reinforcing only because punishment is avoided, or also because the opponent is punished?

To help answer this question, we employed a version of the TAP that permits to isolate victory- and punishment-related effects (Brunnlieb, et al., 2013). In this study, participants play a modified TAP with two kinds of trials: in "passive" trials participants can be punished if they lose, but are not allowed to punish their opponent if they win; in "active" trials they can punish when winning, but cannot be punished when losing. Participants were given either intranasal arginine-vasopressin or placebo and played the TAP in the scanner. For both groups, whole-brain general linear model (GLM) analyses revealed VS activation after won relative to lost trials (Brunnlieb, et al., 2013). The GLM analysis with a factorial design

(factor win vs. lose and factor active vs. passive) did not show interactions in the VS on a corrected level. Here, we extend these analyses by investigating differences in functional connectivity (FC) of the VS in active relative to passive trials in the placebo group. We focused on the placebo group because we were interested in the neural correlates of aggression in general, rather than in the effect of a specific hormone.

FC analysis permits to uncover regions that share similar dynamics of activity across trials and thus allows exploring functional networks that might otherwise remain hidden in standard GLM analyses (Camara, et al., 2009). FC between different areas of the reward network has previously been shown to predict the subjective value of social stimuli (Smith, Clithero, Boltuck, & Huettel, 2014), long-term memory performance after reward learning (Hamann, Dayan, Hummel, & Cohen, 2014), and individual differences in reward sensitivity (Costumero et al., 2013).

Although winning trials should be rewarding simply because participants avoid an aversive outcome (Kim, Shimojo, & O'Doherty, 2006), we hypothesized that punishing the opponent would be more reinforcing than winning alone based on previous studies (Bault, et al., 2008; Singer, et al., 2006; H. Takahashi, et al., 2009). Hence, we expected increases in connectivity between the VS and other reward-related areas in won active as compared to won passive trials. Our main hypothesis was that FC between VS and orbitofrontal cortex (OFC) would be enhanced, given the involvement of these two regions in reward valuation (Liu, et al., 2011). Connectivity between the VS and the OFC has been shown to correlate positively with behavioral persistence (Jung et al., 2010) and associative learning rates (van den Bos, Cohen, Kahnt, & Crone, 2012), whereas it is reduced during the presentation of pleasant stimuli in anhedonia (Keller et al., 2013) and dysphoria (Sabatinelli et al., 2014). In order to test our hypothesis, we performed seed-based FC analyses in data from the previous fMRI study (Brunnlieb et al., 2013).

5.2. Materials and methods

5.2.1. Participants

Only data from the control group in Brunnlieb et al. (2013) were used. Thus, results reported here are based on 15 University students, all male (age = 19–32, mean = 25.8, SD = 3.4), right handed, and without self-reported psychiatric or neurologic illnesses. Informed consent was obtained for each participant. The study had been approved by the University of Magdeburg's ethical committee and was performed according to the Declaration of Helsinki.

5.2.2. Procedure

Subjects were led to believe they would play a reaction time task against an unacquainted opponent. They were introduced to a 26-year old male confederate of the experimenters. After eight practice trials outside the scanner, participants were placed in the scanner and played the TAP (see Task Design). They subsequently filled out self-reports measuring trait aggression (Aggression Questionnaire, AQ; Herzberg, 2003) and trait empathy (Interpersonal Reactivity Index, IRI; Davis, 1983) and were debriefed regarding the true research goals and procedure of the study. The AQ (German version) has four scales: physical aggression, verbal aggression, anger and hostility. The IRI (own translation) has also four scales: perspective taking, fantasy, empathic concern, and personal distress.

5.2.3. Task design

At the beginning of each trial, participants were informed of the type of trial: in active trials (signaled by the German word for punishment) participants could punish their opponent after winning, but could *not* be punished after losing; in passive trials (preceded by the German word for threat), participants could *not* punish their opponent after winning, but could be punished after losing. In each trial (Fig. 5.1), participants had to select the intensity of a sound blast to be directed at their opponent on a 1 to 4 scale (decision phase, 6 seconds). Then,

participants had to respond more quickly than their alleged opponent to a target stimulus, cued by an exclamation mark (2 s). After this, subjects were informed of the opponent's punishment level selection (4 s); this screen was followed by the outcome phase (6 s), in which they were told whether they won or lost and the corresponding outcome, determined by the type of trial.

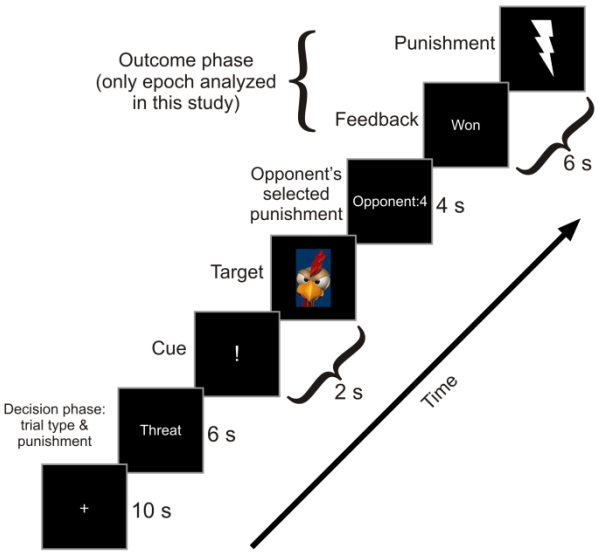

Fig. 5.1 Representative time course of a trial. After a fixation cross, the decision phase started (6 s). The trial type was indicated by the corresponding German word: "punishment" for active trials and "threat" for passive ones (see Task Design); meanwhile participants selected the loudness of the punishment to be received by their opponent. Then the reaction time task followed, in which participants responded as fast as possible to a target stimulus -a cartoon- cued by an exclamation mark (2 s). Then the opponent's selected punishment level was displayed (4 s), followed by the outcome phase (6 s), in which participants were informed of the trial's result. In the passive trial depicted, the participant would have won and avoided the punishment, but his opponent would have not been punished.

Passive and active trials were grouped in blocks of 4, and 24 trials per condition were presented in each of the two runs. The punishment was a polystyrene scratching noise, whose maximal loudness was adjusted to each participant's tolerance.

5.2.4. Data acquisition

All images were recorded with a 3-T Siemens Magnetom Trio Syngo MR 2004A Scanner. The structural image consisted of 192 sagittal slices (T1-weighted MPRage procedure, 256 x 256 matrix, FOV = 256 mm, slice thickness = 1 mm). Functional images were acquired

ascendingly and continuously with a gradient-echo-EPI sequence (TR = 2 s; TE = 30 ms; FOV = 224 mm; flip angle = 80°; matrix = 64 x 64; slice thickness = 3.5 mm), each volume with 32 transversal slices of 3.5 mm isotropic voxels. 342 volumes were recorded for each of the two runs.

5.2.5. Data analysis

Preprocessing and statistical analyses were performed with the Statistical Parametric Mapping 12b toolbox (SPM12b, Wellcome Department of Imaging Neuroscience, University College London, London, UK). Functional images were slice time corrected to the middle slice, realigned to the first functional volume, spatially normalized to 2 mm isotropic voxels in Montreal Neurological Institute (MNI) space, and smoothed with an 8-mm full width at half maximum –FWHM- Gaussian kernel. The first level design comprised the regressors "passive" and "active" (6 s) for the decision phase. For the outcome phase, won and lost trials were specified separately for passive and active conditions, resulting in 4 regressors: "won passive", "lost passive", "won active", and "lost active" (6 s each). The opponent's punishment selection (4 s) and the target (2 s) were also included as regressors, though they were not analyzed; movement parameters estimated from realignment were introduced as well. The standard SPM autoregressive model was applied, and regressors were convolved with the canonical hemodynamic response function. At the second level, a one-sample t-test was performed contrasting won > lost trials in the outcome phase.

We then tested if the VS responded differentially to whether the opponent was punished or not. Although the original ROI analysis showed no significant interaction between block (passive vs active) and outcome (won vs lost) in the VS functional peak of activation (Brunnlieb, et al., 2013), we decided to perform a post-hoc comparison between won active and won passive trials for explorative purposes. To do so, we defined a ROI around the VS

group activation peak in the won > lost contrast. We extracted the mean parameter estimates across all participants for won passive and won active trials separately, and compared them with a paired t-test (see Fig. 2).

For the connectivity analysis, we defined seed regions in the left and right VS (5 mm sphere around x=±10, y=12, z=-8). These regions were chosen based on meta-analyses of reward processing (Diekhof, Kaps, Falkai, & Gruber, 2012; Liu, et al., 2011), and covered areas of functional activation found in this and previous studies with the TAP (Brunnlieb, et al., 2013; Krämer, et al., 2007).

We used the Conn toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012) for our FC analysis. Conn has previously been used to assess task-dependent connectivity in other event-related studies (Glass et al., 2011; Porges & Decety, 2013). The preprocessing pipeline involved regressing out the raw blood-oxygen-level dependent (BOLD) signal from both individual white matter and cerebrospinal fluid masks as well as motion parameters from each voxel through least squares estimation (Whitfield-Gabrieli & Nieto-Castanon, 2012). Main condition effects were also included as confounds, as is common practice in FC analysis (Nair et al., 2014; O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012).

Our FC measure was the Fisher-transformed bivariate correlation coefficient between the average BOLD signal in the seed regions and every other voxel in the brain at time points defined by the first level design. Thus, statistical brain maps of the correlation values (r-maps) were obtained for each participant and each condition. The second level analysis was performed on the r-maps, contrasting won active against won passive conditions across participants with a one-sample t-test. An uncorrected voxel-level threshold of $p<.005$ was selected, with a false discovery rate (FDR) threshold of $q<.05$ at the cluster level.

Finally, we explored associations between connectivity values and behavioral parameters of interest (punishment selections in the decision phase of active and passive trials

–and their difference-, response latency in the reaction time task of active and passive trials – and their difference-, trait aggression and its four facets, and trait empathy and its four facets) with Pearson correlation coefficients. Given that a Bonferroni-corrected threshold might be too restrictive for this set of analysis, the robustness of the correlations was assessed through the bootstrapped 95% confidence intervals (CI) obtained with 1000 iterations. Values were considered significant if the CI did not include zero (D'Angelo et al., 2012).

5.3. Results

The parameter estimates in the VS peak of activation were significantly higher for won active than for won passive trials (t[14]=2.5, p=0.02), as can be seen in the evoked response curves from Fig. 5.2 obtained with rfxplot (Gläscher, 2009).

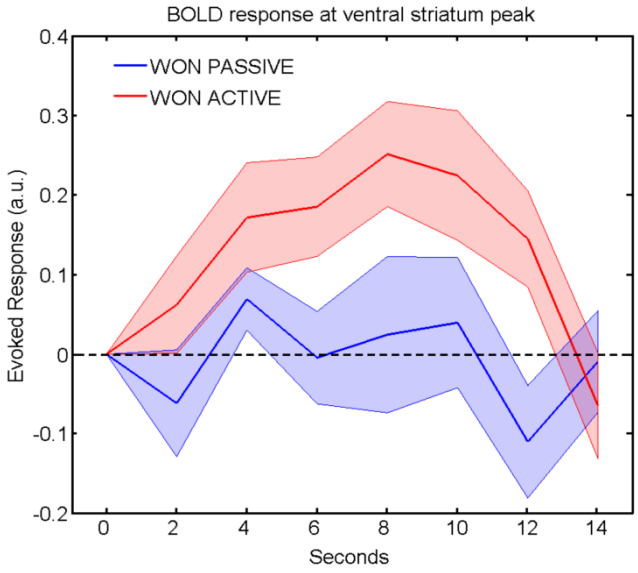

Fig. 5.2 Evoked responses in a right ventral striatum (VS) region of interest (ROI) for won passive and won active conditions. The ROI was defined as a 5 mm sphere limited to supra-threshold voxels around the group VS peak in the won > lost contrast (MNI coordinates 10, 4, -4). Lines represent the time course of the VS parameter estimates for each condition; colored areas represent the corresponding standard errors. Evoked responses were adjusted for all other regressors and rescaled to 0 at onset. A.u.: arbitrary units (parameter estimates)

The bilateral VS seed (colored green in Fig. 5.3) showed a stronger positive correlation in active compared to passive won trials with the medial orbitofrontal cortex (OFC), the bilateral supplementary motor area (SMA), a cluster including the left dorsal posterior insula

(PI) and the dorsal thalamus, and two separate bilateral clusters in the inferior occipital gyrus (IOG). Thus, when participants won a trial in which their opponent was punished, there was increased VS connectivity with OFC, SMA, IOG, left PI and left thalamus (Fig. 5.3, red; Table 5.1a).

As to brain areas negatively associated with the VS seed in the won active > won passive contrast, only one cluster in the left inferior frontal gyrus (IFG) partially covering the anterior insula (AI) survived multiple comparison correction. That is, in won trials, VS-IFG/AI connectivity was relatively lower when the opponent was punished, as compared to when he was not (Fig. 5.3, blue; Table 5.1b).

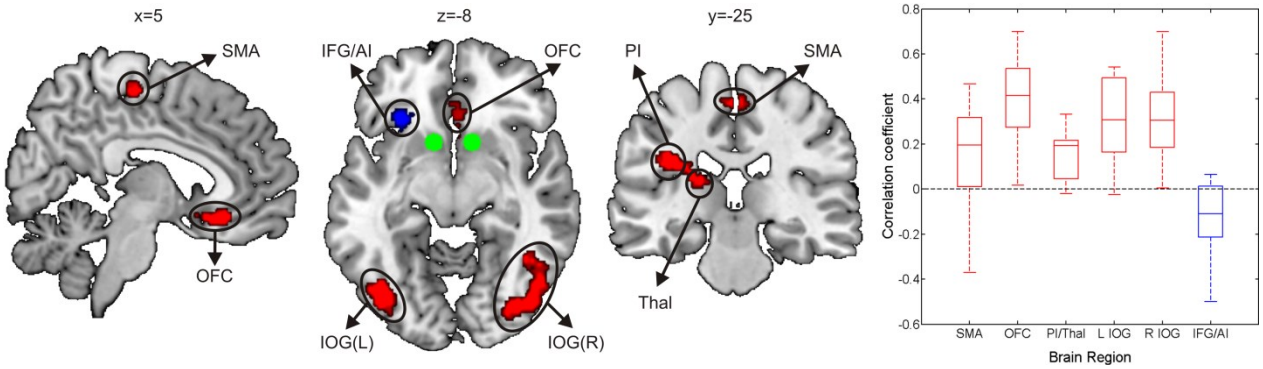

Fig. 5.3 Ventral striatum (VS) connectivity in the won active > won passive contrast (i.e. opponent punished > opponent not punished). The boxplot shows connectivity values in won active trials across participants; the central box represents the 1st, 2nd (i.e. the median) and 3rd quartiles of the distribution, whereas the whiskers represent the minimum and maximum values. Green: bilateral VS seed. Red: increased connectivity with the VS. Blue: decreased connectivity with the VS. SMA: supplementary motor area. OFC: orbitofrontal cortex. IOG(L): left inferior occipital gyrus. IOG(R): right inferior occipital gyrus. IFG: left inferior frontal gyrus. PI: posterior insula. Thal: thalamus.

To assess the behavioral relevance of the connectivity changes, we tested for correlations with behavioral parameters from the paradigm and questionnaire data. Most notably, connectivity between VS and PI/thalamus correlated positively with punishment selections in active minus passive trials (r=.67, 95% BCa CI [.21, .90], p=.006). VS-IFG/AI connectivity correlated positively with trait anger (r=.66, 95% BCa CI [.03, .93], p=0.006), whereas no correlations with trait empathy were found. Also, VS-SMA connectivity showed a

trend towards a negative association with reaction times in active minus passive trials (r=-.53, 95% BCa CI [.03, -0.77], p=.04).

Table 5.1. Functional connectivity (FC) with ventral striatum (VS) seed

a) Greater FC with VS in won active vs. won passive trials							
Hem	Region	BA	x	y	z	T	k
R	Inferior Occipital Gyrus	19	46	-72	-6	5.17	472
L	Inferior Occipital Gyrus	19	-40	-74	-8	4.49	159
L	Insula	13	-40	-22	28	6.61	144
R	Orbitofrontal gyrus	11	4	30	-12	4.63	129
R	Supplementary motor area	6	2	-22	60	4.35	114
b) Lower FC with VS in won active vs. won passive trials							
Hem	Region	BA	x	y	z	T	k
L	Inferior Frontal Gyrus	47	-32	20	-20	-4.98	174

Results are reported at p<.005 (height threshold) and p<0.05 (FDR extent threshold). Clusters are ordered by size (bigger to smaller). Hem = hemisphere, BA = Brodmann Area, xyz = MNI coordinates, T = peak T-values, k = cluster size.

5.4. Discussion

In the present study we investigated whether functional connectivity between the ventral striatum and other brain areas differs between winning to punish someone and winning to avoid being punished oneself.

Comparing the BOLD response in the ventral striatum when punishing the opponent relative to just avoiding punishment yielded increased activity in the former condition. This indicates that the effect of winning in the VS might have been partly driven by whether the rival was punished or not. However, a factorial analysis revealed no interaction between outcome (won/lost) and type of trial (active/passive) in the original article (Brunnlieb, et al., 2013). Although we might have lacked power to obtain a significant interaction, caution is warranted when interpreting this effect until future studies replicate it.

Regarding the connectivity results, the comparison of won active versus won passive trials revealed that the striatum had a greater FC with the orbitofrontal cortex (OFC), the bilateral inferior occipital gyri (IOG), a cluster including the left posterior insula (PI) and left dorsal thalamus, and the supplementary motor area (SMA). The VS signal correlated negatively with that of the left inferior frontal gyrus/anterior insula (IFG/AI) in the same comparison.

The won active > won passive contrast revealed enhanced connectivity between the VS and the medial orbitofrontal cortex (OFC). The OFC is one of the core reward areas (Diekhof, et al., 2012), and has been shown to track the subjective value of stimuli (Winecoff et al., 2013). Our OFC connectivity cluster partially coincided with the one identified by Clithero & Rangel (2014) for money, food, and other rewards - the latter category including social rewards. Furthermore, it was located medially, a subregion more sensitive to reward than to punishment, in contrast to lateral orbitofrontal regions (Kringelbach, 2005). Porges & Decety (2013) recently reported that higher FC between nucleus accumbens and subgenual cingulate cortex when watching mixed martial arts fights, as compared to capoeira dances, concurred with increasing pleasure ratings. This result suggests that FC between VS and subcallosal areas of the medial frontal lobe may partly underlie the gratifying value of seeing others in pain. This may also be true for our results, as participants were competing against a provocative opponent, which gave them a reason to enjoy their rival's misfortune. It is also notable that our medial OFC connectivity cluster was in close vicinity to the aforementioned subcallosal cluster.

We also found a significant increase in FC between the VS and bilateral inferior occipital gyri (IOG) in won active minus won passive trials. IOG activity is a typical correlate of attentional efforts or demands, both with neutral and emotional stimuli (Pessoa, et al., 2002; Vuilleumier, 2005). Therefore, the VS-occipital FC enhancement that arose when the

opponent was punished could be due to the higher emotional salience of the event. This possibility is highlighted by the fact that ventrolateral occipital clusters also emerged in the whole-brain analysis for won vs lost trials (Brunnlieb, et al., 2013). Moreover, this region was recently found to increase its FC strength with the VS in marijuana users presented with cannabis-related cues, as compared to neutral cues (Filbey & Dunlop, 2014). This further suggests that VS-occipital coupling might be relevant in signaling the salience of potentially rewarding events.

We also observed increased VS connectivity with a cluster extending from the posterior insula (PI) to the left dorsal thalamus when the opponent was punished in contrast to when he was not. At a more restrictive height threshold of $p<0.0025$ (uncorrected), the cluster was limited to the PI and adjacent white matter, indicating that the PI signal could have extended artificially to the thalamus. However, the PI receives multiple somatosensory inputs from the thalamus, which compels us not to rule out completely the thalamic connectivity observed (Singer, Critchley, & Preuschoff, 2009). The PI region is known to be involved in low-level somatosensory representations (Kurth, Zilles, Fox, Laird, & Eickhoff, 2010; Nieuwenhuys, 2012), and it is responsive to pleasant tactile stimulation (Ebisch, Ferri, & Gallese, 2014; May, Stewart, Tapert, & Paulus, 2014). Accordingly, connectivity between the VS and the PI has been previously reported after winning outcomes (van Holst, Chase, & Clark, 2014). On the other hand, the dorsal thalamus has been postulated to carry arousal-laden information to the nucleus accumbens (Sescousse, et al., 2013), and significant VS-thalamus FC has been reported both for wins and losses in previous studies (Camara, Rodriguez-Fornells, & Münte, 2008; Cho et al., 2012). Although in our study we compared two positive outcomes (avoiding *versus* delivering a punishment), participants with higher VS-PI/thalamus connectivity values tended to select stronger punishments against their opponents. Hence, the higher VS-

PI/thalamus coupling we observed when the opponent was punished as opposed to when he was not might be linked to a reward-boosted increment in arousal.

The VS correlated positively with the supplementary motor area (SMA) when the opponent was punished, but not in the absence of punishment. SMA activity normally accompanies motor response selection (Nachev, Kennard, & Husain, 2008), but no motor action was required in the outcome phase. Previous studies have found that increased FC between the VS and premotor areas when presented with positively valued stimuli correlated with trait measures of reward sensitivity (Passamonti et al., 2009; Weiland et al., 2013). The authors suggested that this functional link might serve to convert the hedonic value of a reinforcer into readiness for action, with the ultimate aim of obtaining more rewards (Passamonti, et al., 2009; Weiland, et al., 2013). From this point of view, if punishing the opponent was more gratifying than only avoiding punishment, the VS-SMA connectivity might serve to maintain or improve motor performance in the reaction time task of active trials, when the opponent could be punished. Partially supporting this idea, VS-SMA connectivity strength as a trend correlated negatively with reaction times in active minus passive trials -that is, higher connectivity values were associated to faster reactions in active compared to passive trials. Hence, the observed VS-SMA coupling might correspond to an increased urge to punish the opponent, reflected behaviorally in a better competitive performance.

A negative correlation emerged between the VS and a left inferior frontal gyrus (IFG) cluster extending to the anterior insula (AI) when contrasting won active against won passive trials. The AI seems to be key in the representation of internal states (Singer, et al., 2009), most notably self-experienced and vicarious pain (Corradi-Dell'Acqua, Hofstetter, & Vuilleumier, 2011; Singer et al., 2004). In the Porges & Decety (2013) study mentioned previously, the VS-subgenual connectivity cluster extended to the AI, whereas in our study

the VS-AI correlation was *negative*; this might be due to the fact that in Porges & Decety (2013) the participants were passively viewing violent scenes and thus were free to pity the wounded, while in our study participants were "responsible" for punishing their competitor. This situation clearly does not incite any feelings of empathy. On the other hand, the IFG is part of the human mirror neuron system, and is highly responsive to imitative movements (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). Hence not surprisingly, IFG activity is also often observed in social interaction paradigms and interpreted as contributing to mentalizing or perspective-taking processes (Polosan, et al., 2011). Our IFG connectivity cluster was on the left hemisphere around Brodmann area 47, matching a region activated by script-induced guilt (Shin et al., 2000), and an empathy-related peak identified by Liakakis et al. (2011). A recent study found that VS-IFG connectivity induced by happy faces was lower as a function of autistic traits in healthy adults (Sims, Neufeld, Johnstone, & Chakrabarti, 2014). In the present study, the correlation between the VS and the IFG/AI was weaker (i.e. more positive) for anger-prone individuals, who might be expected to be less forgiving (Berry, Worthington, O'Connor, Parrott, & Wade, 2005) and more aggressive (Parrott & Zeichner, 2002) when provoked, and might be therefore less compassionate towards their rival. Altogether, our results suggest that VS-IFG/AI decoupling when punishing a provocative opponent could reflect a reduction of the empathic response.

We cannot infer to which extent the connectivity results reported here reflect subjective reward or not, as we did not assess how satisfied participants were with the different outcomes. Nonetheless, we limited our analysis to the outcome phase of won trials, in which participants experienced victory and punishment avoidance at the same time, and, accordingly, a strong VS activity was elicited. Adding to the evidence from the literature discussed throughout this article, Knutson, Katovich, & Suri (2014) have provided initial proof that positive affect can be inferred from VS activity.

An inherent limitation of FC studies and thus also the current study is the correlational nature of the analysis. Considering that we had no clear hypothesis on the directionality of the activations, we decided to use bivariate correlations, which are also easy to interpret. They provide, however, little insight on the temporal dynamics of the observed connectivity results. Future studies assessing social reward could benefit from using dynamic causal modelling or Granger causality analysis (Cho, et al., 2012; Filbey & Dunlop, 2014) in order to uncover directional relationships in the activation pattern of the reward network.

5.5. Conclusion

Summing up, our study has shown that the human ventral striatum (VS) increases its coupling with brain areas related to reward, attention and arousal when punishing a provoking opponent, as compared to simply avoiding a punishment. This effect concurs with a decoupling between VS and regions presumably related to pain and empathy. Such a brain response would be adaptive in an aggressive interaction by disregarding the competitor's well-being. Backing this idea, it has been shown that brain responses may adapt selectively to contexts of cooperation or competition (Decety, Jackson, Sommerville, Chaminade, & Meltzoff, 2004), possibly to adjust behavior in accordance to the situation. As this is likely subject to individual differences, it should be tested in future studies with in-task subjective ratings and/or physiological measures that permit to infer subjects' enjoyment.

Chapter 6: General discussion

6.1. Summary of findings

In the present work we have expanded on existent experimental aggression paradigms, aiming to develop engaging and realistic tasks that better capture the flow of aggressive interactions despite the experimental constraints, and thus permit to capture direct relationships between brain activity and reactive aggression.

In chapter 2, we probed whether state and trait estimates of the steroid hormones testosterone (T) and cortisol (C) could modulate brain reactivity to angry faces and subsequent aggressive behavior. To that end, we collected data from 39 healthy young women who provided saliva samples during a regular weekday as well as before and after scanning. In the scanner, participants played the Social Threat Aggression Paradigm (STAP), a previously established task devised to link reactivity to anger expressions and aggression. First, we largely replicated the previously observed effects of the paradigm on brain activity, and showed a reduction in amygdala-orbitofrontal (OFC) coupling when the opponent looked angry. Furthermore, we showed that amygdala reactivity to angry faces was related to aggression, and that this relationship was mediated by superior temporal gyrus (STG) activation. This implies that the amygdala subserves quick detection of danger signals, whereas the STG might relate to slower threat perception and assessment. While C had no effect on aggression, T at scan-time modulated amygdala-dependent aggressive responses to angry facial expressions. Specifically, high-T subjects were less aggressive and had lower amygdala reactivity to angry faces, suggesting a lower emotional excitability in these participants.

In chapter 3, we developed a paradigm that allows to measure avoidant responses to provocation, namely the Fight-or-Escape paradigm. In the FOE, participants can either face or

avoid a highly (HP) and a lowly (LP) provoking opponent in a competitive reaction time task. In a first experiment, we showed that participants (n=27, all women) with higher emotional reactivity to threat were less aggressive and more avoidant. In a second experiment (n=34, 13 men), we showed that participants with stronger approach tendencies confronted the HP more often.

In chapter 4, we used the FOE paradigm in the fMRI scanner in order to explore the neural basis of aggressive and avoidant decisions. Following previous evidence, we expected that participants with stronger avoidant tendencies would avoid the provoking opponent more often, both behaviorally (i.e., selecting the escape option more frequently), and cognitively (i.e., not thinking about the opponents' intentions). This should be reflected in relatively lower activation of the mentalizing network in response to provocation, an effect that should be stronger in highly avoidant participants. Although we did not find the expected effect on a between-subject basis, some regions of the mentalizing network were less engaged when avoiding the highly provoking opponent, whereas the amygdala was upregulated. This implies that avoidant responses to provocation are subserved by heightened threat signaling and social-cognitive disengagement.

Finally, in chapter 5 we explored the functional architecture of the ventral striatum (VS), a brain region crucial for reward processing, when participants won an aggressive interaction. VS activation has been repeatedly observed for won trials in the TAP, but this effect could be due to punishment avoidance rather than to punishment delivery. To better clarify this issue, we compared won trials in which participants punished their opponent (active) against those in which they could not (passive). In active trials, the VS increased its coupling with regions related to reward valuation (OFC), attention (inferior occipital gyrus, IOG), arousal (posterior insula/thalamus, PI/thalamus), and motor planning (supplementary motor area, SMA), whereas it had reduced connectivity with a region involved in empathic

processes (inferior frontal gyrus/anterior insula, IFG/AI). Importantly, connectivity strength between VS and PI/thalamus was related to increased aggression, whereas the VS-IFG/AI link was weaker in participants high in trait anger. VS-SMA coupling was related at trend level with faster reaction times in active trials. Our data suggest that positive outcomes in a competitive context can shape ongoing neural responses, which in turn favor aggressive behavior.

6.2. Strengths and limitations

The most important limitation of the current work is the artificiality of the experimental tasks employed. Participants played on a computer or inside the MRI scanner, which precludes direct bodily confrontation as would occur in a real brawl. However, we did include elements aimed at making the tasks more realistic and engaging for participants. The fact that in all studies participants were clearly provoked, selecting higher punishments in high relative to low provocation trials and/or over time, indicates that we succeeded in this enterprise. This in turn speaks for the reliability of the TAP, which, despite being flexibly analyzed across studies (Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014), has yielded remarkably robust results during its 50 years of existence.

Neuroimaging results from chapter 2 also suggest that subjects were highly immersed in the aggressive encounter. In many experiments brain activity is measured while participants are presented with a series of faces that convey no meaning in themselves. In some cases, the task might require to press a button when a fixation cross appears, or to indicate whether two faces are identical. In these paradigms, faces provide no information beyond that required for successful task performance. In chapter 2, we circumvented this shortcoming by making angry faces informative about the opponent's intentions towards the participant, and,

consequently, elicited robust activity in mentalizing areas. We were thus able to observe direct links between brain reactivity to facial expressions and aggression.

Our work also overcomes an important limitation of previous paradigms by allowing the joint measurement of aggression and avoidance in response to provocation. Although some of the tasks outlined in section 1.3 also offer the possibility to adopt a prosocial strategy, none of them allow subjects to avoid confrontation, even when this is arguably the most common response to provocation in everyday life (Harrison, Ahn, & Adolphs, 2015). We have shown that novel insights can be gained by enriching participants' behavioral repertoire, improving the naturalness of experimental tasks and thus better recreating real-life social interactions (Ferguson & Rueda, 2009; Zaki & Ochsner, 2012). Albeit results from the TAP cannot be readily generalized to extreme instances of aggression such as murder (Ferguson & Rueda, 2009), laboratory aggression tasks have good predictive and external validity regarding its milder manifestations (C. A. Anderson & Bushman, 1997; Giancola & Parrott, 2008). Moreover, it can be argued that studying the normal range of a given behavior is necessary to better understand its extreme cases (D. J. Anderson, 2012).

Another limitation in this regard is that the samples employed were limited to healthy college students. Some authors have raised concerns regarding the validity of these samples, since it is unclear how well findings obtained from such subjects can generalize to clinical populations or to other sociocultural contexts (Henrich, Heine, & Norenzayan, 2010; Payne & Chappell, 2008). Even within the student population, those who volunteer to participate in such studies might already be predisposed to do so partly because of personality characteristics (B. Marcus & Schütz, 2005). From this perspective, our research subjects might not be fully representative of the whole population they belong to, and they might have been already familiar with the research environment. Although this can cause "good subject" bias (i.e. the tendency to respond in accordance to the perceived expectation of the

experimenter; Nichols & Maner, 2008), the fact that we only included participants who believed they were playing against somebody else suggests that we measured genuine aggressive behavior (C. A. Anderson & Bushman, 1997). Furthermore, college student samples have been shown to have only minimal differences relative to the general population regarding self-reported antisocial behavior, making them appropriate to study this construct (Wiecko, 2010). An additional advantage of these samples is that they are highly homogeneous in age, cognitive ability and education level, and thus provide increased internal validity for the effects observed – i.e., they limit confounds related to sampling variability. For all these reasons, although caution must be exercised when extrapolating our findings to pathological aggression, we contend that our results relate to everyday instances of anger and aggression as experienced by healthy individuals and apply at least to such situations.

6.3. Implications for neurobiological models of aggression

Our results help to clarify the role of specific brain regions on aggressive behavior. In the experiment presented in chapter 2, in which the aggressive encounter was inevitable, amygdala reactivity to threat was related to aggressive retaliation. This is in line with the model of Coccaro and colleagues (2011) presented in the Introduction, according to which aggressive impulses originate in the amygdala and brainstem. However, in the paradigm used in chapter 4, in which threat could be escaped, amygdala activation was related to avoidance. This indicates that amygdala reactivity to threat does not have a univocally pro-aggressive effect. Rather, as pointed out in the corresponding chapters, the amygdala –more specifically the basolateral amygdala- facilitates approach or avoidance adaptively. These results allow to reconcile the well-known involvement of the amygdala in fear conditioning (LeDoux, 2014) with its more recently acknowledged involvement in approach-related behaviors (Weymar & Schwabe, 2016). Importantly, our data also indicates that amygdala reactivity is more strongly

related to reactive than to proactive aggression (De Dreu, Scholte, van Winden, & Ridderinkhof, 2015).

In chapter 5 we also observed context-dependent brain responses to aggression outcomes, so that winning to punish the opponent evoked a more widespread pattern of ventral striatum connectivity than winning to avoid being punished oneself. This implies that, while the VS might not be initially involved in the *generation* of aggressive impulses at the start of a confrontation, it might have a relevant role in the *perpetuation* of aggression in subsequent events. Our data suggests that this vindictive aspect of aggression and its neural correlates (VS activity and connectivity) should be taken into account in future neurocognitive models of aggression, as is done in basic psychological accounts of the phenomenon (DeWall, et al., 2011).

Data from chapters 2 and 4 compels to reformulate the function of the rostral/dorsal mPFC in aggression. In chapter 2, the mPFC displayed higher reactivity to angry than to neutral faces. In chapter 4, the mPFC showed increased activity in avoidance decisions, but was less strongly recruited when avoiding a highly provoking opponent. As mentioned in chapter 4, the rostral mPFC is involved in monitoring social dominance signals (Ligneul, et al., 2016). Our data agrees with this evidence and suggests that mPFC activation might subserve the assessment or prediction of the potential threat posed by the opponent. Hence, in the context of aggression the mPFC should be better understood not as a top-down inhibitory control region, but as a general social-cognitive hub involved in threat or dominance assessment.

As we have commented in the Introduction, according to the model by Coccaro et al. (2011) the OFC can act as a buffer against aggressive behavior, but also as a pro-aggressive input. In this work, we have characterized the circumstances under which the OFC can operate as a brake or as an enabler of aggression. In chapter 2, the OFC lowered its coupling

with the amygdala during trials in which the opponent looked angry, which could have indirectly facilitated aggression. In chapter 4, the OFC was recruited when participants decided to fight. Finally, in chapter 5, the OFC increased its coupling with the VS when the rival was punished, as opposed to when he was not. Many possible functions for the OFC have been suggested, be it social signal encoding, reward valuation, or general task monitoring (Stalnaker, et al., 2015). Our data cannot speak for the general function of the OFC, but help to delimit the role of this brain region in aggression depending on which region it is coupled with, and in which context. Specifically, the OFC seems to favor the appetitive aspect of aggression in concert with the VS when punishing a provoking opponent (Seymour, et al., 2007), but it can buffer retaliation by downregulating amygdala reactivity to threat (Coccaro, et al., 2011).

Another aspect worth commenting is gender dimorphism in aggression. In chapter 2, we observed no differences in mean punishment selections between the current female sample and the male sample employed in an earlier study with an equivalent paradigm, which is in line with the meta-analytic finding that provocation mitigates gender differences in aggression (Bettencourt & Miller, 1996). However, an interesting dissociation emerged, such that amygdala, but not OFC reactivity to angry faces was related to aggression in women, whereas the opposite pattern was found in men (Beyer, et al., 2015). In women, the OFC lowered its coupling with the amygdala following the presentation of angry faces, and hence could have only modulated aggression indirectly. It is also noteworthy that we did not find any effects in the ACC. This contrasts with previous studies in men wherein increased ACC reactivity to provocation was observed when participants selected high punishments (Beyer, et al., 2015; Krämer, et al., 2007). While these findings might point to a more direct role of subcortical regions in women's aggression, it could be precipitous to draw conclusions from these results

until a study allowing direct between-gender comparison against the same opponent is conducted.

A further contribution of the present work is to stress the importance of networks and connectivity between different brain areas. Indeed, for most operations brain regions operate in conjunction with others rather than in isolation, even if the contribution of some brain regions dominates over others in a specific cognitive process (McIntosh, 1999). In chapters 2 and 5 we performed functional connectivity analyses that revealed brain networks involved in the processing of facial threat cues and victory, respectively. In chapter 2, additionally, we performed a mediation analysis which showed that, across participants, fast amygdala responses to angry faces were linked to the slower STG reactivity, which in turn led to increased aggressive behavior. Although the importance of crosstalk between different brain areas is hardly novel, we have uncovered functional interactions in the context of aggressive interactions that might prove relevant to understand the physiology of aggression beyond the information provided by the activity of single brain areas.

6.4. Conceptual implications

The previously commented amygdala effects agree with the notion that approach and avoidance motivation are not mutually exclusive, since both dispositions have a similar physiological profile and can lead to analogous behavioral strategies when the available options are limited (Lang & Bradley, 2013). In a study using the Preemptive Strike Game (PSG; see section 1.3.5), an immense majority (≈75%) of participants who chose the attack option reported that they would have withdrawn from the task had they been given the chance to do so (Simunovic, et al., 2013). These results dovetail ours in suggesting that fear might boost aggression in the absence of alternatives. In chapter 2 we have additionally shown that participants with high testosterone were less aggressive and had lower amygdala activation,

which can be arguably interpreted as reduced fear (LeDoux, 2014). Aggressive behavior in the STAP might hence constitute a measure of defensive aggression, whereas, in the FOE, it might rather reflect an appetitive retaliatory drive (Albert, et al., 1993). This is not to say that approach and avoidance motivation are equivalent. In fact, in the second experiment from chapter 3 and in chapter 4 we have found clearly distinctive effects of approach and avoidance motivation on aggression, such that mostly the former was associated with aggressive decisions. Rather, our data suggests that fearful responses might be expressed as aggression in the absence of an escape option. If, on the other hand, avoidance is possible, choosing to retaliate would reflect approach motivation. Rodent studies also provide support for this idea. Rats bred for low and high anxiety display more aggression than non-selected animals when caged with an intruder (Neumann, Veenema, & Beiderbeck, 2010), but only low-anxiety rats have increased DA release in the VS during the encounter (Beiderbeck et al., 2012). This is also in line with McNaughton and Corr's theory of defensive behavior, according to which avoidance motivation can lead to fight responses if threat cannot be avoided (McNaughton & Corr, 2004). Supporting this notion, hypothetical scenarios involving inescapable threat can elicit both freezing and aggression, which are correlated with each other in women (Blanchard, Hynd, Minke, Minemoto, & Blanchard, 2001). Interestingly, high-anxiety female rats have even been reported to be slightly more aggressive than their low anxious counterparts (Bosch, Meddle, Beiderbeck, Douglas, & Neumann, 2005; de Jong, Beiderbeck, & Neumann, 2014), indicating that the link between fear and aggression in inescapable encounters might be stronger in females.

Related to the latter point, our results contribute to an expanding literature that compels to abandon simplistic notions about the effects of T on human behavior. Specifically, we showed that T might in some instances reduce aggression in women, which might be a mechanism to improve one's status (Eisenegger, et al., 2011). One might speculate that, since

cultural pressures tend to favor prosocial over antisocial behavior in women (Nivette, et al., 2014), the former might be more readily adopted as a status-attaining strategy in high-T individuals. However, as commented in the Introduction and in chapter 2, multiple other social and biological factors are involved in the T-aggression relationship, such that hormones like oxytocin (Campbell, 2013) or estradiol (Stanton & Edelstein, 2009) as well as other potential mediating variables such as self-construal (Welker, et al., 2016) are also likely to play an important role in women's aggression.

The results from chapter 5, in which VS connectivity was modulated by the outcome of the aggressive encounter, can also offer some insight regarding the development of aggression. First, these data might help to explain why some individuals become aggressive in competition, such that they would process a rival's misfortunes as more rewarding and/or intense. In line with this formulation, it has been shown that both the own team's success and a rival team's misfortune elicit similar VS activity, and the strength of this activation predicts the willingness to aggress a fan of the opposite team (Cikara, Botvinick, & Fiske, 2011). Second, our results imply that not only negative events such as frustration but also *positive* outcomes can have a pro-aggressive effect. This links with studies in humans and rodents showing that victories increase subsequent aggression (Carré, Campbell, et al., 2013; Clinard, et al., 2016; Oyegbile & Marler, 2005). Rewarding outcomes could thus favor a sense of dominance over others and thereby contribute to the perpetuation of aggression (Blanchard & Blanchard, 2003). Such a mechanism might be at play in some cases of bullying, mobbing, and other types of status-based harassment (Garandeau, Lee, & Salmivalli, 2014).

6.5. Future directions

Throughout the present thesis we have offered possibilities to expand our findings. In this last section we will briefly suggest some more possible follow-up studies that would permit to establish, complement, and further the presented results.

Extending the study from chapter 2, one could inspect whether participants react differently to faces expressing emotions of different valence. For instance, the opponent could smile in some trials and look sad in others. This might be even more meaningful in paradigms such as the Ultimatum Game, which involve strategic play and in which such "cheeky" actions might elicit either rage or compassion depending on how they are interpreted. Regarding Chapter 3, it would be interesting to contrast trials in which participants must avoid (forced avoidance) against trials in which participants must fight (forced fight). This would also allow dissociating brain activity specifically related to decision processes, to preparatory activity, and to safety. Taking this idea further, one could allow participants to choose among the trial types from chapter 5 (i.e. those in which only the rival can be punished versus those in which only the player can) in order to measure participants' retaliatory versus altruistic tendencies. This is similar to studies in which participants can choose between an aggressive and a non-aggressive videogame (Ferguson et al., 2008).

In all studies, we have highlighted that one could back up conclusions derived from the TAP by including subjective ratings throughout the task. If this was however done on a trial-wise basis, it could induce metacognitive brain activity (Molenberghs, Trautwein, Böckler, Singer, & Kanske, 2016), and perhaps influence participants' behavior due to priming effects. That is, asking participants to rate their guilt or anger might activate cognitive schemata and behavioral scripts associated with these emotions (Berkowitz, 2012). Furthermore, including these questions could induce suspicion regarding the true goal of the research, such that one should include filler items (e.g. "How hungry are you?") to disguise the construct being

measured, and maybe avoid items directly related to aggression altogether. Given that in the TAP, punishment selection is a relatively unambiguous and direct measure of aggression, it would probably be redundant to tap anger- or aggression-related constructs, and such efforts could be better devoted to measure other, more basic affective dimensions (e.g. positive and negative affect). In any case, we propose that these measures should be administered between runs in order to not interfere with the course of the task.

From a broader perspective, a comprehensive research agenda that includes genetic, neuroimaging, and experience sampling data would allow to answer lingering questions that are yet to be addressed in the aggression literature, such as how well neuroimaging findings predict aggressive behavior in everyday life (Heller et al., 2015), or how genetic makeup impacts the development of aggression-related brain structures in interaction with micro- and macrostructural social factors (Little et al., 2015). Although such large multidimensional datasets are prone to producing false positives, data-driven techniques applied to these data can complement hypothesis-driven research and ultimately help to advance the field on both theoretical (e.g. determining which model best explains variability in aggressive behavior; Lamb, Annetta, Hoston, Shapiro, & Matthews, 2017) and applied grounds (e.g. which variables are most useful for predicting aggression among inmates; Gilbert & Daffern, 2010).

Finally, it would be highly interesting to employ some of our paradigms not only to clinical populations, but also to *a priori* mentally healthy populations displaying unusually elevated aggression, such as sports fans (Cikara, et al., 2011; van der Meij et al., 2015). As commented earlier, the extent of VS reactivity to a rival team's defeat predicts self-reported eagerness to aggress a fan of that team (Cikara, et al., 2011). A study employing a version of the TAP in Chapter 5 with fans of rival clubs would permit a direct test of this prediction, and could provide insight into the neural basis of the still prevalent phenomenon of hooligan violence (Ostrowsky, 2016). On a similar note, considering the increasing relevance of online

harassment (Fisher, Gardella, & Teurbe-Tolon, 2016; Rost, Stahel, & Frey, 2016), predicting aggression in indirect social interactions is also likely to become an important goal in our subfield. An interesting venue for further studies would thus be to use the TAP in order to identify predictors of cyber-aggression.

References

Aarts, E., Dolan, C. V., Verhage, M., & van der Sluis, S. (2015). Multilevel analysis quantifies variation in the experimental effect while optimizing power and preventing false positives. *BMC Neuroscience, 16*, 94. doi: 10.1186/s12868-015-0228-5

Achterberg, M., van Duijvenvoorde, A. C. K., Bakermans-Kranenburg, M. J., & Crone, E. A. (2016). Control your anger! The neural basis of aggression regulation in response to negative social feedback. *Social Cognitive and Affective Neuroscience, 11*(5), 712-720. doi: 10.1093/scan/nsv154

Ackermann, S., Spalek, K., Rasch, B., Gschwind, L., Coynel, D., Fastenrath, M., . . . de Quervain, D. J. F. (2012). Testosterone levels in healthy men are related to amygdala reactivity and memory performance. *Psychoneuroendocrinology, 37*(9), 1417-1424. doi: http://dx.doi.org/10.1016/j.psyneuen.2012.01.008

Adams, D., Barnett, S. A., Bechtereva, N. P., Carter, B. F., Delgado, J. M. R., Diaz, J. L., . . . Wahlstrom, R. (1990). The Seville Statement on Violence. *American Psychologist, 45*(10), 1167-1168. doi: 10.1037/0003-066x.45.10.1167

Aghajani, M., Colins, O. F., Klapwijk, E. T., Veer, I. M., Andershed, H., Popma, A., . . . Vermeiren, R. R. J. M. (2016). Dissociable relations between amygdala subregional networks and psychopathy trait dimensions in conduct-disordered juvenile offenders. *Human brain mapping, 37*(11), 4017-4033. doi: 10.1002/hbm.23292

Albert, D. J., Nanji, N., Brayley, K. N., & Madryga, F. J. (1979). Hyperreactivity as well as mouse killing is induced by electrical stimulation of the lateral hypothalamus in the rat. *Behavioral and Neural Biology, 27*(1), 59-71. doi: http://dx.doi.org/10.1016/S0163-1047(79)92756-0

Albert, D. J., Walsh, M. L., & Jonik, R. H. (1993). Aggression in humans: What is its biological foundation? *Neuroscience & Biobehavioral Reviews, 17*(4), 405-425. doi: 10.1016/S0149-7634(05)80117-4

Alia-Klein, N., Goldstein, R. Z., Kriplani, A., Logan, J., Tomasi, D., Williams, B., . . . Fowler, J. S. (2008). Brain Monoamine Oxidase A Activity Predicts Trait Aggression. *The Journal of Neuroscience, 28*(19), 5099-5104. doi: 10.1523/jneurosci.0925-08.2008

Alink, L. R. A., van Ijzendoorn, M. H., Bakermans-Kranenburg, M. J., Mesman, J., Juffer, F., & Koot, H. M. (2008). Cortisol and externalizing behavior in children and adolescents: Mixed meta-analytic evidence for the inverse relation of basal cortisol and cortisol reactivity with externalizing behavior. *Developmental Psychobiology, 50*(5), 427-450. doi: 10.1002/dev.20300

Amodio, D. M., & Harmon-Jones, E. (2011). Trait emotions and affective modulation of the startle eyeblink: on the unique relationship of trait anger. *Emotion, 11*(1), 47-51. doi: 10.1037/a0021238

Anderson, C. A., Buckley, K. E., & Carnagey, N. L. (2008). Creating Your Own Hostile Environment: A Laboratory Examination of Trait Aggressiveness and the Violence Escalation Cycle. *Personality and Social Psychology Bulletin, 34*(4), 462-473. doi: 10.1177/0146167207311282

Anderson, C. A., & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1*(1), 19-41.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual review of psychology, 53*, 27-51. doi: 10.1146/annurev.psych.53.100901.135231

Anderson, D. J. (2012). Optogenetics, Sex, and Violence in the Brain: Implications for Psychiatry. *Biological Psychiatry, 71*(12), 1081-1089. doi: http://dx.doi.org/10.1016/j.biopsych.2011.11.012

Anstrom, K. K., Miczek, K. A., & Budygin, E. A. (2009). Increased phasic dopamine signaling in the mesolimbic pathway during social defeat in rats. *Neuroscience, 161*(1), 3-12. doi: http://dx.doi.org/10.1016/j.neuroscience.2009.03.023

Archer, J., Graham-Kevan, N., & Davies, M. (2005). Testosterone and aggression: A reanalysis of Book, Starzyk, and Quinsey's (2001) study. *Aggression and Violent Behavior, 10*(2), 241-261. doi: http://dx.doi.org/10.1016/j.avb.2004.01.001

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences, 18*(4), 177-185. doi: http://dx.doi.org/10.1016/j.tics.2013.12.003

Assaf, M., Kahn, I., Pearlson, G. D., Johnson, M. R., Yeshurun, Y., Calhoun, V. D., & Hendler, T. (2009). Brain Activity Dissociates Mentalization from Motivation During an Interpersonal Competitive Game. *Brain Imaging and Behavior, 3*(1), 24-37. doi: 10.1007/s11682-008-9047-y

Averbeck, B. B., & Duchaine, B. (2009). Integration of social and utilitarian factors in decision making. *Emotion, 9*(5), 599-608. doi: 10.1037/a0016509

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi: http://dx.doi.org/10.1016/j.jml.2007.12.005

Baczkowski, B. M., van Zutphen, L., Siep, N., Jacob, G. A., Domes, G., Maier, S., . . . van de Ven, V. (2016). Deficient amygdala–prefrontal intrinsic connectivity after effortful emotion regulation in borderline personality disorder. *European Archives of Psychiatry and Clinical Neuroscience*, 1-15. doi: 10.1007/s00406-016-0760-z

Bahlmann, J., Aarts, E., & D'Esposito, M. (2015). Influence of Motivation on Control Hierarchy in the Human Frontal Cortex. *The Journal of Neuroscience, 35*(7), 3207-3217. doi: 10.1523/jneurosci.2389-14.2015

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. *The Journal of Neuroscience, 27*(31), 8161-8165. doi: 10.1523/jneurosci.1554-07.2007

Bandura, A. (1978). Social Learning Theory of Aggression. *Journal of communication, 28*(3), 12-29. doi: 10.1111/j.1460-2466.1978.tb01621.x

Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala–frontal connectivity during emotion regulation. *Social Cognitive and Affective Neuroscience, 2*(4), 303-312. doi: 10.1093/scan/nsm029

Baron, R. A., & Eggleston, R. J. (1972). Performance on the "aggression machine": Motivation to help or harm? *Psychonomic Science, 26*(6), 321-322. doi: 10.3758/bf03328631

Baron, R. A., & Richardson, D. R. (1994). *Human Aggression*. New York: Plenum.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. doi: 10.1037/0022-3514.51.6.1173

Barrett, L. F., & Barrett, D. J. (2001). An Introduction to Computerized Experience Sampling in Psychology. *Social Science Computer Review, 19*(2), 175-185. doi: 10.1177/089443930101900204

Bates, B. T., Zhang, S., Dufek, J. S., & Chen, F. C. (1996). The effects of sample size and variability on the correlation coefficient. *Medicine & Science in Sports & Exercise, 28*(3), 386-391.

Bateup, H. S., Booth, A., Shirtcliff, E. A., & Granger, D. A. (2002). Testosterone, cortisol, and women's competition. *Evolution and Human Behavior, 23*(3), 181-192. doi: http://dx.doi.org/10.1016/S1090-5138(01)00100-3

Bault, N., Coricelli, G., & Rustichini, A. (2008). Interdependent Utilities: How Social Ranking Affects Choice Behavior. *PLoS ONE, 3*(10), e3477. doi: 10.1371/journal.pone.0003477

Bault, N., Joffily, M., Rustichini, A., & Coricelli, G. (2011). Medial prefrontal cortex and striatum mediate the influence of social comparison on the decision process. *Proceedings of the National Academy of Sciences, 108*(38), 16044-16049. doi: 10.1073/pnas.1100892108

Beaver, J. D., Lawrence, A. D., Passamonti, L., & Calder, A. J. (2008). Appetitive Motivation Predicts the Neural Response to Facial Signals of Aggression. *The Journal of Neuroscience, 28*(11), 2719-2725. doi: 10.1523/jneurosci.0033-08.2008

Beehner, J. C., Phillips-Conroy, J. E., & Whitten, P. L. (2005). Female testosterone, dominance rank, and aggression in an Ethiopian population of hybrid baboons. *American Journal of Primatology, 67*(1), 101-119. doi: 10.1002/ajp.20172

Beiderbeck, D. I., Reber, S. O., Havasi, A., Bredewold, R., Veenema, A. H., & Neumann, I. D. (2012). High and abnormal forms of aggression in rats with extremes in trait anxiety – Involvement of the dopamine system in the nucleus accumbens. *Psychoneuroendocrinology, 37*(12), 1969-1980. doi: http://doi.org/10.1016/j.psyneuen.2012.04.011

Berkowitz, L. (1983). Aversively stimulated aggression: Some parallels and differences in research with animals and humans. *American Psychologist, 38*(11), 1135-1144. doi: 10.1037/0003-066x.38.11.1135

Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin, 106*(1), 59-73. doi: 10.1037/0033-2909.106.1.59

Berkowitz, L. (2012). A Different View of Anger: The Cognitive-Neoassociation Conception of the Relation of Anger to Aggression. *Aggressive Behavior, 38*(4), 322-333. doi: 10.1002/ab.21432

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1995). The metrics of cardiac chronotropism: Biometric perspectives. *Psychophysiology, 32*(2), 162-171. doi: 10.1111/j.1469-8986.1995.tb03308.x

Berry, J. W., Worthington, E. L., O'Connor, L. E., Parrott, L., & Wade, N. G. (2005). Forgivingness, Vengeful Rumination, and Affective Traits. *Journal of Personality, 73*(1), 183-226. doi: 10.1111/j.1467-6494.2004.00308.x

Bertsch, K., Grothe, M., Prehn, K., Vohs, K., Berger, C., Hauenstein, K., . . . Herpertz, S. C. (2013). Brain volumes differ between diagnostic groups of violent criminal offenders. *European Archives of Psychiatry and Clinical Neuroscience, 263*(7), 593-606. doi: 10.1007/s00406-013-0391-6

Bettencourt, B. A., & Miller, N. (1996). Gender differences in aggression as a function of provocation: A meta-analysis. *Psychological Bulletin, 119*(3), 422-447. doi: 10.1037/0033-2909.119.3.422
Bettencourt, B. A., Talley, A., Benjamin, A. J., & Valentine, J. (2006). Personality and aggressive behavior under provoking and neutral conditions: A meta-analytic review. *Psychological Bulletin, 132*(5), 751-777. doi: 10.1037/0033-2909.132.5.751
Beyer, F., Münte, T. F., Erdmann, C., & Krämer, U. M. (2013). Emotional reactivity to threat modulates activity in mentalizing network during aggression. *Social Cognitive and Affective Neuroscience, 9*(10), 1552-1560. doi: 10.1093/scan/nst146
Beyer, F., Münte, T. F., Erdmann, C., & Krämer, U. M. (2014). Emotional reactivity to threat modulates activity in mentalizing network during aggression. *Social Cognitive and Affective Neuroscience, 9*(10), 1552-1560. doi: 10.1093/scan/nst146
Beyer, F., Münte, T. F., Göttlich, M., & Krämer, U. M. (2015). Orbitofrontal Cortex Reactivity to Angry Facial Expression in a Social Interaction Correlates with Aggressive Behavior. *Cerebral Cortex, 25*(9), 3057-3063. doi: 10.1093/cercor/bhu101
Beyer, F., Münte, T. F., & Krämer, U. M. (2014). Increased neural reactivity to socio-emotional stimuli links social exclusion and aggression. *Biological Psychology, 96*, 102-110. doi: 10.1016/j.biopsycho.2013.12.008
Beyer, F., Münte, T. F., Wiechert, J., Heldmann, M., & Krämer, U. M. (2014). Trait Aggressiveness Is Not Related to Structural Connectivity between Orbitofrontal Cortex and Amygdala. *PLoS ONE, 9*(6), e101105. doi: 10.1371/journal.pone.0101105
Blair, R. J. R. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry, 71*(6), 727-731. doi: 10.1136/jnnp.71.6.727
Blair, R. J. R., & Cipolotti, L. (2000). Impaired social response reversal. A case of "acquired sociopathy". *Brain, 123*(6), 1122. doi: 10.1093/brain/123.6.1122
Blanchard, C. D., & Blanchard, R. J. (2003). What can animal aggression research tell us about human aggression? *Hormones and Behavior, 44*(3), 171-177. doi: http://dx.doi.org/10.1016/S0018-506X(03)00133-8
Blanchard, C. D., Hynd, A. L., Minke, K. A., Minemoto, T., & Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defense patterns of non-human mammals. *Neuroscience & Biobehavioral Reviews, 25*(7–8), 761-770. doi: http://dx.doi.org/10.1016/S0149-7634(01)00056-2
Böhnke, R., Bertsch, K., Kruk, M. R., & Naumann, E. (2010). The relationship between basal and acute HPA axis activity and aggressive behavior in adults. *Journal of Neural Transmission, 117*(5), 629-637. doi: 10.1007/s00702-010-0391-x
Böhnke, R., Bertsch, K., Kruk, M. R., Richter, S., & Naumann, E. (2010). Exogenous cortisol enhances aggressive behavior in females, but not in males. *Psychoneuroendocrinology, 35*(7), 1034-1044. doi: http://dx.doi.org/10.1016/j.psyneuen.2010.01.004
Boksem, M. A. S., Mehta, P. H., Van den Bergh, B., van Son, V., Trautmann, S. T., Roelofs, K., . . . Sanfey, A. G. (2013). Testosterone Inhibits Trust but Promotes Reciprocity. *Psychological Science*. doi: 10.1177/0956797613495063
Bos, P. A., Panksepp, J., Bluthé, R.-M., & Honk, J. v. (2012). Acute effects of steroid hormones and neuropeptides on human social–emotional behavior: A review of single administration studies. *Frontiers in Neuroendocrinology, 33*(1), 17-35. doi: 10.1016/j.yfrne.2011.01.002
Bos, P. A., Terburg, D., & van Honk, J. (2010). Testosterone decreases trust in socially naïve humans. *Proceedings of the National Academy of Sciences, 107*(22), 9991-9995. doi: 10.1073/pnas.0911700107
Bosch-Bouju, C., Hyland, B., & Parr-Brownlie, L. (2013). Motor thalamus integration of cortical, cerebellar and basal ganglia information: implications for normal and parkinsonian conditions. [Review]. *Frontiers in Computational Neuroscience, 7*(163). doi: 10.3389/fncom.2013.00163
Bosch, O. J., Meddle, S. L., Beiderbeck, D. I., Douglas, A. J., & Neumann, I. D. (2005). Brain Oxytocin Correlates with Maternal Aggression: Link to Anxiety. *The Journal of Neuroscience, 25*(29), 6807-6815. doi: 10.1523/jneurosci.1342-05.2005
Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*(3), 276-298. doi: 10.1037/1528-3542.1.3.276
Brandstrom, S., Richter, J., & Nylander, P.-O. (2003). Further development of the temperament and character inventory. *Psychological Reports, 93*(3f), 995-1002. doi: 10.2466/pr0.2003.93.3f.995
Bravo-Rivera, C., Roman-Ortiz, C., Montesinos-Cartagena, M., & Quirk, G. J. (2015). Persistent active avoidance correlates with activity in prelimbic cortex and ventral striatum. *Frontiers in Behavioral Neuroscience, 9*. doi: 10.3389/fnbeh.2015.00184

Bray, S., & O'Doherty, J. (2007). Neural Coding of Reward-Prediction Error Signals During Classical Conditioning With Attractive Faces. *Journal of Neurophysiology, 97*(4), 3036-3045.
Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. *Neuron, 68*(5), 815-834. doi: http://dx.doi.org/10.1016/j.neuron.2010.11.022
Brown, E. S., Woolston, D. J., & Frol, A. B. (2008). Amygdala Volume in Patients Receiving Chronic Corticosteroid Therapy. *Biological Psychiatry, 63*(7), 705-709. doi: 10.1016/j.biopsych.2007.09.014
Brunnlieb, C., Münte, T. F., Krämer, U. M., Tempelmann, C., & Heldmann, M. (2013). Vasopressin modulates neural responses during human reactive aggression. *Social Neuroscience, 8*(2), 148-164. doi: 10.1080/17470919.2013.763654
Bryant, F. B., & Smith, B. D. (2001). Refining the Architecture of Aggression: A Measurement Model for the Buss–Perry Aggression Questionnaire. *Journal of Research in Personality, 35*(2), 138-167. doi: http://dx.doi.org/10.1006/jrpe.2000.2302
Buades-Rotger, M., & Gallardo-Pujol, D. (2014). The role of the monoamine oxidase A gene in moderating the response to adversity and associated antisocial behavior: a review. *Psychology Research and Behavior Management, 7*, 185-200. doi: 10.2147/prbm.s40458
Bui, H. N., Sluss, P. M., Blincko, S., Knol, D. L., Blankenstein, M. A., & Heijboer, A. C. (2013). Dynamics of serum testosterone during the menstrual cycle evaluated by daily measurements with an ID-LC–MS/MS method and a 2nd generation automated immunoassay. *Steroids, 78*(1), 96-101. doi: 10.1016/j.steroids.2012.10.010
Burnstein, E., & Worchel, P. (1962). Arbitrariness of frustration and its consequences for aggression in a social situation1. *Journal of Personality, 30*(4), 528-540. doi: 10.1111/j.1467-6494.1962.tb01687.x
Burnstein, K. L., Maiorino, C. A., Dai, J. L., & Cameron, D. J. (1995). Androgen and glucocorticoid regulation of androgen receptor cDNA expression. *Molecular and Cellular Endocrinology, 115*(2), 177-186. doi: http://dx.doi.org/10.1016/0303-7207(95)03688-1
Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on hostile versus instrumental aggression dichotomy? *Psychological Review, 108*(1), 273-279. doi: 10.1037/0033-295x.108.1.273
Bushman, B. J., DeWall, C. N., Pond, R. S., & Hanus, M. D. (2014). Low glucose relates to greater aggression in married couples. *Proceedings of the National Academy of Sciences, 111*(17), 6254-6257. doi: 10.1073/pnas.1400619111
Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*(3), 452.
Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *J Pers Soc Psychol, 63*(3), 452-459.
Buss, D. M., & Shackelford, T. K. (1997). Human aggression in evolutionary psychological perspective. *Clinical Psychology Review, 17*(6), 605-619. doi: 10.1016/S0272-7358(97)00037-8
Butter, C. M., Mishkin, M., & Mirsky, A. F. (1968). Emotional responses toward humans in monkeys with selective frontal lesions. *Physiology & Behavior, 3*(2), 213-215. doi: 10.1016/0031-9384(68)90087-5
Byrd, A. L., & Manuck, S. B. (2014). MAOA, childhood maltreatment and antisocial behavior: Meta-analysis of a gene-environment interaction. *Biological Psychiatry, 75*(1), 10.1016/j.biopsych.2013.1005.1004. doi: 10.1016/j.biopsych.2013.05.004
Camara, E., Rodriguez-Fornells, A., & Münte, T. F. (2008). Functional connectivity of reward processing in the brain. *Frontiers in human neuroscience, 2*, 19. doi: 10.3389/neuro.09.019.2008
Camara, E., Rodriguez-Fornells, A., Ye, Z., & Münte, T. F. (2009). Reward networks in the brain as captured by connectivity measures. *Frontiers in neuroscience, 3*, 350-362. doi: 10.3389/neuro.01.034.2009
Campbell, A. (2013). The evolutionary psychology of women's aggression. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*(1631). doi: 10.1098/rstb.2013.0078
Campese, V. D., Sears, R. M., Moscarello, J. M., Diaz-Mataix, L., Cain, C. K., & LeDoux, J. E. (2016). The Neural Foundations of Reaction and Action in Aversive Motivation. In E. H. Simpson & P. D. Balsam (Eds.), *Behavioral Neuroscience of Motivation* (pp. 171-195). Cham: Springer International Publishing.
Caramaschi, D., de Boer, S. F., de Vries, H., & Koolhaas, J. M. (2008). Development of violence in mice through repeated victory along with changes in prefrontal cortex neurochemistry. *Behavioural Brain Research, 189*(2), 263-272. doi: http://dx.doi.org/10.1016/j.bbr.2008.01.003
Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and Indirect Aggression During Childhood and Adolescence: A Meta-Analytic Review of Gender Differences, Intercorrelations, and Relations to Maladjustment. *Child Development, 79*(5), 1185-1229. doi: 10.1111/j.1467-8624.2008.01184.x
Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, 100*(9), 5497-5502. doi: 10.1073/pnas.0935845100

Carré, J. M., Campbell, J. A., Lozoya, E., Goetz, S. M. M., & Welker, K. M. (2013). Changes in testosterone mediate the effect of winning on subsequent aggressive behaviour. *Psychoneuroendocrinology, 38*(10), 2034-2041. doi: http://dx.doi.org/10.1016/j.psyneuen.2013.03.008

Carré, J. M., Hyde, L. W., Neumann, C. S., Viding, E., & Hariri, A. R. (2013). The neural signatures of distinct psychopathic traits. *Social Neuroscience, 8*(2), 122-135. doi: 10.1080/17470919.2012.703623

Carré, J. M., McCormick, C. M., & Hariri, A. R. (2011). The social neuroendocrinology of human aggression. *Psychoneuroendocrinology, 36*(7), 935-944. doi: 10.1016/j.psyneuen.2011.02.001

Carré, J. M., & Mehta, P. H. (2011). Importance of considering testosterone–cortisol interactions in predicting human aggression and dominance. *Aggressive Behavior, 37*(6), 489-491. doi: 10.1002/ab.20407

Carré, J. M., & Olmstead, N. A. (2015). Social neuroendocrinology of human aggression: Examining the role of competition-induced testosterone dynamics. *Neuroscience, 286*, 171-186. doi: 10.1016/j.neuroscience.2014.11.029

Carré, J. M., Putnam, S. K., & McCormick, C. M. (2009). Testosterone responses to competition predict future aggressive behaviour at a cost to reward in men. *Psychoneuroendocrinology, 34*(4), 561-570. doi: http://dx.doi.org/10.1016/j.psyneuen.2008.10.018

Carver, C. S. (2004). Negative Affects Deriving From the Behavioral Approach System. *Emotion, 4*(1), 3-22. doi: 10.1037/1528-3542.4.1.3

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin, 135*(2), 183-204. doi: 10.1037/a0013965

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319-333. doi: 10.1037/0022-3514.67.2.319

Caseras, F. X., Fullana, M. A., Riba, J., Barbanoj, M. J., Aluja, A., & Torrubia, R. (2006). Influence of individual differences in the Behavioral Inhibition System and stimulus content (fear versus blood-disgust) on affective startle reflex modulation. *Biol Psychol, 72*(3), 251-256. doi: 10.1016/j.biopsycho.2005.10.009

Catani, M., Jones, D. K., Donato, R., & Ffytche, D. H. (2003). Occipito-temporal connections in the human brain. *Brain, 126*(9), 2093-2107. doi: 10.1093/brain/awg203

Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition, 36*, 426-433. doi: http://dx.doi.org/10.1016/j.concog.2015.03.012

Cha, J., Fekete, T., Siciliano, F., Biezonski, D., Greenhill, L., Pliszka, S. R., . . . Posner, J. (2015). Neural Correlates of Aggression in Medication-Naive Children with ADHD: Multivariate Analysis of Morphometry and Tractography. [Original Article]. *Neuropsychopharmacology, 40*(7), 1717-1725. doi: 10.1038/npp.2015.18

Champaneri, S., Xu, X., Carnethon, M. R., Bertoni, A. G., Seeman, T., DeSantis, A. S., . . . Golden, S. H. (2013). Diurnal salivary cortisol is associated with body mass index and waist circumference: The multiethnic study of atherosclerosis. *Obesity, 21*(1), E56-E63. doi: 10.1002/oby.20047

Chaux, E., & Castellanos, M. (2015). Money and age in schools: Bullying and power imbalances. *Aggressive Behavior, 41*(3), 280-293. doi: 10.1002/ab.21558

Cherek, D. R. (1981). Effects of smoking different doses of nicotine on human aggressive behavior. *Psychopharmacology, 75*(4), 339-345. doi: 10.1007/bf00435849

Cherek, D. R., Moeller, F. G., Schnapp, W., & Dougherty, D. M. (1997). Studies of violent and nonviolent male parolees: I. Laboratory and psychometric measurements of aggression. *Biol Psychiatry, 41*(5), 514-522.

Chester, D. S., & DeWall, C. N. (2016). The pleasure of revenge: retaliatory aggression arises from a neural imbalance toward reward. *Social Cognitive and Affective Neuroscience, 11*(7), 1173-1182. doi: 10.1093/scan/nsv082

Cho, Y. T., Fromm, S., Guyer, A. E., Detloff, A., Pine, D. S., Fudge, J. L., & Ernst, M. (2012). Nucleus accumbens, thalamus and insula connectivity during incentive anticipation in typical adults and adolescents. *NeuroImage, 66C*, 508-521. doi: 10.1016/j.neuroimage.2012.10.013

Christov-Moore, L., Simpson, E. A., Coudé, G., Grigaityte, K., Iacoboni, M., & Ferrari, P. F. (2014). Empathy: Gender effects in brain and behavior. *Neuroscience & Biobehavioral Reviews, 46, Part 4*, 604-627. doi: http://dx.doi.org/10.1016/j.neubiorev.2014.09.001

Churchwell, J. C., Carey, P. D., Ferrett, H. L., Stein, D. J., & Yurgelun-Todd, D. A. (2012). Abnormal Striatal Circuitry and Intensified Novelty Seeking among Adolescents Who Abuse Methamphetamine and Cannabis. *Developmental Neuroscience, 34*(4), 310-317.

Cikara, M., Botvinick, M. M., & Fiske, S. T. (2011). Us Versus Them. *Psychological Science, 22*(3), 306-313. doi: 10.1177/0956797610397667

Clausen, A. N., Youngren, W., Sisante, J.-F. V., Billinger, S. A., Taylor, C., & Aupperle, R. L. (2016). Combat PTSD and Implicit Behavioral Tendencies for Positive Affective Stimuli: A Brief Report. [Original Research]. *Frontiers in Psychology, 7*(758). doi: 10.3389/fpsyg.2016.00758

Clinard, C. T., Barnes, A. K., Adler, S. G., & Cooper, M. A. (2016). Winning agonistic encounters increases testosterone and androgen receptor expression in Syrian hamsters. *Hormones and Behavior, 86*, 27-35. doi: http://dx.doi.org/10.1016/j.yhbeh.2016.09.002

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience, 9*(9), 1289-1302. doi: 10.1093/scan/nst106

Coccaro, E. F., Berman, M. E., & Kavoussi, R. J. (1997). Assessment of life history of aggression: development and psychometric characteristics. *Psychiatry Research, 73*(3), 147-157.

Coccaro, E. F., McCloskey, M. S., Fitzgerald, D. A., & Phan, K. L. (2007). Amygdala and Orbitofrontal Reactivity to Social Threat in Individuals with Impulsive Aggression. *Biological Psychiatry, 62*(2), 168-178. doi: http://dx.doi.org/10.1016/j.biopsych.2006.08.024

Coccaro, E. F., Sripada, C. S., Yanowitch, R. N., & Phan, K. L. (2011). Corticolimbic Function in Impulsive Aggressive Behavior. *Biological Psychiatry, 69*(12), 1153-1159. doi: 10.1016/j.biopsych.2011.02.032

Cohen, D., & Nisbett, R. E. (1994). Self-Protection and the Culture of Honor: Explaining Southern Violence. *Personality and Social Psychology Bulletin, 20*(5), 551-567. doi: 10.1177/0146167294205012

Conzelmann, A., Mucha, R. F., Jacob, C. P., Weyers, P., Romanos, J., Gerdes, A. B., . . . Pauli, P. (2009). Abnormal affective responsiveness in attention-deficit/hyperactivity disorder: subtype differences. *Biol Psychiatry, 65*(7), 578-585. doi: 10.1016/j.biopsych.2008.10.038

Cooper, R. M., & Langton, S. R. H. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour Research and Therapy, 44*(9), 1321-1329. doi: 10.1016/j.brat.2005.10.004

Corradi-Dell'Acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2012). Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Social Cognitive and Affective Neuroscience, 8*(4), 424-431. doi: 10.1093/scan/nss014

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2011). Felt and Seen Pain Evoke the Same Local Patterns of Cortical Activity in Insular and Cingulate Cortex. *The Journal of Neuroscience, 31*(49), 17996-18006. doi: 10.1523/jneurosci.2686-11.2011

Costumero, V., Barrós-Loscertales, A., Bustamante, J. C., Ventura-Campos, N., Fuentes, P., & Ávila, C. (2013). Reward sensitivity modulates connectivity among reward brain areas during processing of anticipatory reward cues. *European Journal of Neuroscience, 38*(3), 2399-2407. doi: 10.1111/ejn.12234

Coupland, N. J., Sustrik, R. A., Ting, P., Li, D., Hartfeil, M., Singh, A. J., & Blair, R. J. (2004). Positive and negative affect differentially influence identification of facial emotions. *Depression and Anxiety, 19*(1), 31-34. doi: 10.1002/da.10136

Craig, M. C., Catani, M., Deeley, Q., Latham, R., Daly, E., Kanaan, R., . . . Murphy, D. G. M. (2009). Altered connections on the road to psychopathy. *Mol Psychiatry, 14*(10), 946-953.

Cristofori, I., Zhong, W., Mandoske, V., Chau, A., Krueger, F., Strenziok, M., & Grafman, J. (2016). Brain Regions Influencing Implicit Violent Attitudes: A Lesion-Mapping Study. *The Journal of Neuroscience, 36*(9), 2757-2768. doi: 10.1523/jneurosci.2975-15.2016

Crockett, M. J., Apergis-Schoute, A., Herrmann, B., Lieberman, M. D., Müller, U., Robbins, T. W., & Clark, L. (2013). Serotonin Modulates Striatal Responses to Fairness and Retaliation in Humans. *The Journal of Neuroscience, 33*(8), 3505-3513. doi: 10.1523/jneurosci.2761-12.2013

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin, 122*(1), 5-37. doi: 10.1037/0033-2909.122.1.5

Cunningham-Bussel, A. C., Root, J. C., Butler, T., Tuescher, O., Pan, H., Epstein, J., . . . Silbersweig, D. (2009). Diurnal cortisol amplitude and fronto-limbic activity in response to stressful stimuli. *Psychoneuroendocrinology, 34*(5), 694-704. doi: 10.1016/j.psyneuen.2008.11.011

Cunningham, R. L., Lumia, A. R., & McGinnis, M. Y. (2012). Androgen Receptors, Sex Behaviour, and Aggression. *Neuroendocrinology, 96*(2), 131-140. doi: 10.1159/000337663

D'Angelo, G. M., Lazar, N. A., Zhou, G., Eddy, W. F., Morris, J. C., & Sheline, Y. I. (2012). Bootstrapping GEE models for fMRI regional connectivity. *NeuroImage, 63*(4), 1890-1900. doi: 10.1016/j.neuroimage.2012.08.036

da Cunha-Bang, S., Hjordt, L. V., Perfalk, E., Beliveau, V., Bock, C., Lehel, S., . . . Knudsen, G. M. (in press). Serotonin 1B Receptor Binding Is Associated With Trait Anger and Level of Psychopathy in Violent Offenders. *Biological Psychiatry*. doi: http://dx.doi.org/10.1016/j.biopsych.2016.02.030

da Cunha-Bang, S., MacDonald Fisher, P., Hjordt, L. V., Perfalk, E., Skibsted, A. P., Bock, C., . . . Knudsen, G. M. (in press). Violent offenders respond to provocations with high amygdala and striatal reactivity. *Social Cognitive and Affective Neuroscience*.

da Cunha-Bang, S., Mc Mahon, B., MacDonald Fisher, P., Jensen, P. S., Svarer, C., & Moos Knudsen, G. (2016). High trait aggression in men is associated with low 5-HT levels, as indexed by 5-HT 4 receptor binding. *Social Cognitive and Affective Neuroscience, 11*(4), 548-555. doi: 10.1093/scan/nsv140

da Cunha-Bang, S., Stenbæk, D. S., Holst, K., Licht, C. L., Jensen, P. S., Frokjaer, V. G., . . . Knudsen, G. M. (2013). Trait aggression and trait impulsivity are not related to frontal cortex 5-HT2A receptor binding in healthy individuals. *Psychiatry Research: Neuroimaging, 212*(2), 125-131. doi: http://dx.doi.org/10.1016/j.pscychresns.2012.09.007

Dabbs, J. M., Jr., & Hargrove, M. F. (1997). Age, testosterone, and behavior among female prison inmates. *Psychosomatic Medicine, 59*(5), 477-480.

Dambacher, F., Sack, A. T., Lobbestael, J., Arntz, A., Brugman, S., & Schuhmann, T. (2014). Out of control: Evidence for anterior insula involvement in motor impulsivity and reactive aggression. *Social Cognitive and Affective Neuroscience, 10*(4), 508-516. doi: 10.1093/scan/nsu077

Dambacher, F., Schuhmann, T., Lobbestael, J., Arntz, A., Brugman, S., & Sack, A. T. (2015a). No Effects of Bilateral tDCS over Inferior Frontal Gyrus on Response Inhibition and Aggression. *PLoS ONE, 10*(7), e0132170. doi: 10.1371/journal.pone.0132170

Dambacher, F., Schuhmann, T., Lobbestael, J., Arntz, A., Brugman, S., & Sack, A. T. (2015b). Reducing proactive aggression through non-invasive brain stimulation. *Social Cognitive and Affective Neuroscience, 10*(10), 1303-1309. doi: 10.1093/scan/nsv018

Dart, R. A. (1953). The Predatory Transition from Ape to Man. *International Anthropological and Linguistic Review, 1*(4), 201-219.

Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the Neural Circuitry of Emotion Regulation--A Possible Prelude to Violence. *Science, 289*(5479), 591-594. doi: 10.1126/science.289.5479.591

Davis, J. R., & Reyna, C. (2015). Seeing red: How perceptions of social status and worth influence hostile attributions and endorsement of aggression. *British Journal of Social Psychology, 54*(4), 728-747. doi: 10.1111/bjso.12109

Davis, M., Falls, W. A., Campeau, S., & Kim, M. (1993). Fear-potentiated startle: a neural and pharmacological analysis. *Behav Brain Res, 58*(1-2), 175-198.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for amultidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126.

De Dreu, C. K. W., Scholte, H. S., van Winden, F. A. A. M., & Ridderinkhof, K. R. (2015). Oxytocin tempers calculated greed but not impulsive defense in predator–prey contests. *Social Cognitive and Affective Neuroscience, 10*(5), 721-728. doi: 10.1093/scan/nsu109

de Jong, T. R., Beiderbeck, D. I., & Neumann, I. D. (2014). Measuring Virgin Female Aggression in the Female Intruder Test (FIT): Effects of Oxytocin, Estrous Cycle, and Anxiety. *PLoS ONE, 9*(3), e91701. doi: 10.1371/journal.pone.0091701

Decety, J., Chen, C., Harenski, C., & Kiehl, K. A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience, 7*, 489. doi: 10.3389/fnhum.2013.00489

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *NeuroImage, 23*(2), 744-751. doi: 10.1016/j.neuroimage.2004.05.025

Deisseroth, K. (2011). Optogenetics. [Commentary]. *Nature Methods, 8*(1), 26-29. doi: 10.1038/nmeth.f.324

Delgado, M. R., Schotter, A., Ozbay, E. Y., & Phelps, E. A. (2008). Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions. *Science, 321*(5897), 1849-1852. doi: 10.1126/science.1158860

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Denson, T. F., Mehta, P. H., & Ho Tan, D. (2013). Endogenous testosterone and cortisol jointly influence reactive aggression in women. *Psychoneuroendocrinology, 38*(3), 416-424. doi: 10.1016/j.psyneuen.2012.07.003

Denson, T. F., Ronay, R., von Hippel, W., & Schira, M. M. (2013). Endogenous testosterone and cortisol modulate neural responses during induced anger control. *Social Neuroscience, 8*(2), 165-177. doi: 10.1080/17470919.2012.655425

Derntl, B., Windischberger, C., Robinson, S., Kryspin-Exner, I., Gur, R. C., Moser, E., & Habel, U. (2009). Amygdala activity to fear and anger in healthy young males is associated with testosterone. *Psychoneuroendocrinology, 34*(5), 687-693. doi: 10.1016/j.psyneuen.2008.11.007

Derntl, B., Windischberger, C., Robinson, S., Lamplmayr, E., Kryspin-Exner, I., Gur, R. C., . . . Habel, U. (2008). Facial emotion recognition and amygdala activation are associated with menstrual cycle phase. *Psychoneuroendocrinology, 33*(8), 1031-1040. doi: http://dx.doi.org/10.1016/j.psyneuen.2008.04.014

DeWall, C. N., Anderson, C. A., & Bushman, B. J. (2011). The general aggression model: Theoretical extensions to violence. *Psychology of Violence, 1*, 245-258. doi: 10.1037/a0023842

DeWall, C. N., Finkel, E. J., Lambert, N. M., Slotter, E. B., Bodenhausen, G. V., Pond, R. S., . . . Fincham, F. D. (2013). The voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations. *Aggressive Behavior, 39*(6), 419-439. doi: 10.1002/ab.21496

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189-228. doi: 10.1214/ss/1032280214

Diekhof, E. K., Kaps, L., Falkai, P., & Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude - an activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia, 50*, 1252-1266. doi: 10.1016/j.neuropsychologia.2012.02.007

Díez-Cirarda, M., Ojeda, N., Peña, J., Cabrera-Zubizarreta, A., Gómez-Beldarrain, M. Á., Gómez-Esteban, J. C., & Ibarretxe-Bilbao, N. (2015). Neuroanatomical Correlates of Theory of Mind Deficit in Parkinson's Disease: A Multimodal Imaging Study. *PLoS ONE, 10*(11), e0142234. doi: 10.1371/journal.pone.0142234

Dijkman, C. I. M. P. C., & Devries, M. W. M. D. (1987). The Social Ecology of Anxiety Theoretical and Quantitative Perspectives. *Journal of Nervous & Mental Disease, 175*(9), 550-557.

Dixson, A. F. (1980). Androgens and aggressive behavior in primates: A review. *Aggressive Behavior, 6*(1), 37-67. doi: 10.1002/1098-2337(1980)6:1<37::aid-ab2480060106>3.0.co;2-7

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven, CT, US: Yale University Press.

Dougherty, D. D., Rauch, S. L., Deckersbach, T., & et al. (2004). Ventromedial prefrontal cortex and amygdala dysfunction during an anger induction positron emission tomography study in patients with major depressive disorder with anger attacks. *Archives of general psychiatry, 61*(8), 795-804. doi: 10.1001/archpsyc.61.8.795

Dougherty, D. M., Bjork, J. M., Moeller, F. G., & Swann, A. C. (1997). The Influence of Menstrual-Cycle Phase on the Relationship Between Testosterone and Aggression. *Physiology & Behavior, 62*(2), 431-435. doi: http://dx.doi.org/10.1016/S0031-9384(97)88991-3

Dragunow, M., & Faull, R. (1989). The use of c-fos as a metabolic marker in neuronal pathway tracing. *Journal of Neuroscience Methods, 29*(3), 261-265. doi: http://dx.doi.org/10.1016/0165-0270(89)90150-7

Dreher, J.-C., Dunne, S., Pazderska, A., Frodl, T., Nolan, J. J., & O'Doherty, J. P. (2016). Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males. *Proceedings of the National Academy of Sciences, 113*(41), 11633-11638. doi: 10.1073/pnas.1608085113

Driessen, M., Herrmann, J., Stahl, K., & et al. (2000). Magnetic resonance imaging volumes of the hippocampus and the amygdala in women with borderline personality disorder and early traumatization. *Archives of general psychiatry, 57*(12), 1115-1122. doi: 10.1001/archpsyc.57.12.1115

Ducharme, S., Hudziak, J. J., Botteron, K. N., Ganjavi, H., Lepage, C., Collins, D. L., . . . Karama, S. (2011). Right Anterior Cingulate Cortical Thickness and Bilateral Striatal Volume Correlate with Child Behavior Checklist Aggressive Behavior Scores in Healthy Children. *Biological Psychiatry, 70*(3), 283-290. doi: http://dx.doi.org/10.1016/j.biopsych.2011.03.015

Duke, A. A., Bègue, L., Bell, R., & Eisenlohr-Moul, T. (2013). Revisiting the serotonin–aggression relation in humans: A meta-analysis. *Psychological Bulletin, 139*(5), 1148-1172. doi: 10.1037/a0031544

Dušková, M., Sosvorová, L., Vítků, J., Jandíková, H., Rácz, B., Chlupáčová, T., . . . Stárka, L. (2015). Changes in the Concentrations of Corticoid Metabolites – The Effect of Stress, Diet and Analytical Method. *Prague Medical Report, 116*(4), 268-278. doi: 10.14712/23362936.2015.66

Dvash, J., Gilam, G., Ben-Ze'ev, A., Hendler, T., & Shamay-Tsoory, S. G. (2010). The envious brain: The neural basis of social comparison. *Human brain mapping, 31*(11), 1741-1750. doi: 10.1002/hbm.20972

Ebisch, S. J. H., Ferri, F., & Gallese, V. (2014). Touching Moments: Desire Modulates the Neural Anticipation of Active Romantic Caress. [Original Research]. *Frontiers in Behavioral Neuroscience, 8*. doi: 10.3389/fnbeh.2014.00060

Edwards, E. R., & Mottarella, K. (2014). Perceptions of the Previously Convicted. *International Journal of Offender Therapy and Comparative Criminology, 59*(12), 1358-1377. doi: 10.1177/0306624x14536899

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage, 36*(3), 511-521. doi: http://dx.doi.org/10.1016/j.neuroimage.2007.03.060

Eisenegger, C., Haushofer, J., & Fehr, E. (2011). The role of testosterone in social interaction. *Trends in Cognitive Sciences, 15*(6), 263-271. doi: http://dx.doi.org/10.1016/j.tics.2011.04.008

Eisenegger, C., Naef, M., Snozzi, R., Heinrichs, M., & Fehr, E. (2010). Prejudice and truth about the effect of testosterone on human bargaining behaviour. [10.1038/nature08711]. *Nature, 463*(7279), 356-359. doi: 10.1038/nature08711

Elgot, J., Halliday, J., & Rankin, J. (2016). Steven Woolfe 'smiling and well' after alleged fight with Ukip MEP, *The Guardian*. Retrieved from http://www.theguardian.com/politics/2016/oct/06/ukip-leadership-favourite-steven-woolfe-collapses-in-european-parliament

Elliott, F. A. (1988). Violence: A Product of Biosocial Interactions. *Journal of the American Academy of Psychiatry and the Law Online, 16*(2), 131-143.

Elnazer, H. Y., & Baldwin, D. S. (2014). Investigation of Cortisol Levels in Patients with Anxiety Disorders: A Structured Review. In C. M. Pariante & M. D. Lapiz-Bluhm (Eds.), *Behavioral Neurobiology of Stress-related Disorders* (pp. 191-216). Berlin, Heidelberg: Springer Berlin Heidelberg.

Else-Quest, N. M., Higgins, A., Allison, C., & Morton, L. C. (2012). Gender differences in self-conscious emotional experience: A meta-analysis. *Psychological Bulletin, 138*(5), 947-981. doi: 10.1037/a0027930

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment, 26*(2), 419-432. doi: 10.1037/a0035569

Emmerling, F., Schuhmann, T., Lobbestael, J., Arntz, A., Brugman, S., & Sack, A. T. (2016). The Role of the Insular Cortex in Retaliation. *PLoS ONE, 11*(4), e0152000. doi: 10.1371/journal.pone.0152000

Eron, L. D. (1987). The development of aggressive behavior from the perspective of a developing behaviorism. *American Psychologist, 42*(5), 435-442. doi: 10.1037/0003-066x.42.5.435

Eron, L. D., Huesmann, L. R., & Zelli, A. (1991). The Role of Parental Variables in the Learning of Aggression. In D. J. Pepler & K. H. Rubin (Eds.), *The Development and Treatment of Childhood Aggression* (pp. 169-188). Hillsdale, NJ: Lawrence Erlbaum Associates.

Falkner, A. L., Grosenick, L., Davidson, T. J., Deisseroth, K., & Lin, D. (2016). Hypothalamic control of male aggression-seeking behavior. *Nature Neuroscience, 19*(4), 596-604. doi: 10.1038/nn.4264
http://www.nature.com/neuro/journal/v19/n4/abs/nn.4264.html#supplementary-information

Falkner, A. L., & Lin, D. (2014). Recent advances in understanding the role of the hypothalamic circuit during aggression. *Frontiers in Systems Neuroscience, 8*. doi: 10.3389/fnsys.2014.00168

Fanning, J. R., Berman, M. E., & Long, J. M. (2014). P3 and provoked aggressive behavior. *Social Neuroscience, 9*(2), 118-129. doi: 10.1080/17470919.2013.866596

Faria, M. A. (2013). Violence, mental illness, and the brain – A brief history of psychosurgery: Part 3 – From deep brain stimulation to amygdalotomy for violent behavior, seizures, and pathological aggression in humans. *Surgical Neurology International, 4*, 91. doi: 10.4103/2152-7806.115162

Farrington, D. P. (1993). Childhood origins of teenage antisocial behaviour and adult social dysfunction. *Journal of the Royal Society of Medicine, 86*(1), 13-17.

Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the General Aggression Model. *Aggression and Violent Behavior, 17*(3), 220-228. doi: https://doi.org/10.1016/j.avb.2012.02.007

Ferguson, C. J., & Rueda, S. M. (2009). Examining the validity of the modified Taylor competitive reaction time test of aggression. *Journal of Experimental Criminology, 5*(2), 121. doi: 10.1007/s11292-009-9069-5

Ferguson, C. J., Rueda, S. M., Cruz, A. M., Ferguson, D. E., Fritz, S., & Smith, S. M. (2008). Violent Video Games and Aggression. *Criminal Justice and Behavior, 35*(3), 311-332. doi: doi:10.1177/0093854807311719

Fernando, A. B., Murray, J. E., & Milton, A. L. (2013). The amygdala: securing pleasure and avoiding pain. *Frontiers in Behavioral Neuroscience, 7*, 190. doi: 10.3389/fnbeh.2013.00190

Ferrari, P. F., Van Erp, A. M. M., Tornatzky, W., & Miczek, K. A. (2003). Accumbal dopamine and serotonin in anticipation of the next aggressive episode in rats. *European Journal of Neuroscience, 17*(2), 371-378. doi: 10.1046/j.1460-9568.2003.02447.x

Ficks, C. A., & Waldman, I. D. (2014). Candidate Genes for Aggression and Antisocial Behavior: A Meta-analysis of Association Studies of the 5HTTLPR and MAOA-uVNTR. *Behavior Genetics, 44*(5), 427-444. doi: 10.1007/s10519-014-9661-y

Filbey, F. M., & Dunlop, J. (2014). Differential reward network functional connectivity in cannabis dependent and non-dependent users. *Drug and alcohol dependence, 140*, 101-111. doi: 10.1016/j.drugalcdep.2014.04.002

Fisher, B. W., Gardella, J. H., & Teurbe-Tolon, A. R. (2016). Peer Cybervictimization Among Adolescents and the Associated Internalizing and Externalizing Problems: A Meta-Analysis. [journal article]. *Journal of Youth and Adolescence, 45*(9), 1727-1743. doi: 10.1007/s10964-016-0541-z

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., & Falk, A. (2007). Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum. *Science, 318*(5854), 1305-1308. doi: 10.1126/science.1145876

Forgays, D. G., Forgays, D. K., & Spielberger, C. D. (1997). Factor structure of the state-trait anger expression inventory. *Journal of personality assessment, 69*(3), 497-507.

Forsdyke, D. R. (2015). Wittgenstein's Certainty is Uncertain: Brain Scans of Cured Hydrocephalics Challenge Cherished Assumptions. [journal article]. *Biological Theory, 10*(4), 336-342. doi: 10.1007/s13752-015-0219-x

Fountas, K. N., & Smith, J. R. (2007). Historical evolution of stereotactic amygdalotomy for the management of severe aggression. *Journal of Neurosurgery, 106*(4), 710-713. doi: 10.3171/jns.2007.106.4.710

Frankle, W. G., Lombardo, I., New, A. S., Goodman, M., Talbot, P. S., Huang, Y., . . . Siever, L. J. (2005). Brain Serotonin Transporter Distribution in Subjects With Impulsive Aggressivity: A Positron Emission Study With [11C]McN 5652. *American Journal of Psychiatry, 162*(5), 915-923. doi: 10.1176/appi.ajp.162.5.915

Franzini, A., Broggi, G., Cordella, R., Dones, I., & Messina, G. (2013). Deep-Brain Stimulation for Aggressive and Disruptive Behavior. *World Neurosurgery, 80*(3–4), S29.e11-S29.e14. doi: http://dx.doi.org/10.1016/j.wneu.2012.06.038

Furl, N., Gallagher, S., & Averbeck, B. B. (2012). A Selective Emotional Decision-Making Bias Elicited by Facial Expressions. *PLoS ONE, 7*(3), e33461. doi: 10.1371/journal.pone.0033461

Gable, S. L., Reis, H. T., & Elliot, A. J. (2000). Behavioral activation and inhibition in everyday life. *Journal of Personality and Social Psychology, 78*(6), 1135-1149. doi: 10.1037/0022-3514.78.6.1135

Gagnon, J., Aubin, M., Emond, F. C., Derguy, S., Bessette, M., & Jolicoeur, P. (2016). Neural mechanisms underlying attribution of hostile intention in nonaggressive individuals: An ERP study. *International Journal of Psychophysiology, 110*, 153-162. doi: http://dx.doi.org/10.1016/j.ijpsycho.2016.08.007

Gagnon, J., Aubin, M., Emond, F. C., Derguy, S., Brochu, A. F., Bessette, M., & Jolicoeur, P. (2016). An ERP study on hostile attribution bias in aggressive and nonaggressive individuals. *Aggressive Behavior*. doi: 10.1002/ab.21676

Gallardo-Pujol, D., Kramp, U., García-Forero, C., Pérez-Ramírez, M., & Andrés-Pueyo, A. (2006). Assessing aggressiveness quickly and efficiently: the Spanish adaptation of Aggression Questionnaire-Refined version. *European Psychiatry, 21*(7), 487-494.

Gallardo-Pujol, D., Andrés-Pueyo, A., & Maydeu-Olivares, A. (2013). MAOA genotype, social exclusion and aggression: an experimental test of a gene–environment interaction. *Genes, Brain and Behavior, 12*(1), 140-145. doi: 10.1111/j.1601-183X.2012.00868.x

Gan, G., Preston-Campbell, R. N., Moeller, S. J., Steinberg, J. L., Lane, S. D., Maloney, T., . . . Alia-Klein, N. (2016). Reward vs. Retaliation—the Role of the Mesocorticolimbic Salience Network in Human Reactive Aggression. *Frontiers in Behavioral Neuroscience, 10*. doi: 10.3389/fnbeh.2016.00179

Gan, G., Sterzer, P., Marxen, M., Zimmermann, U. S., & Smolka, M. N. (2015). Neural and Behavioral Correlates of Alcohol-Induced Aggression Under Provocation. [Original Article]. *Neuropsychopharmacology, 40*(13), 2886-2896. doi: 10.1038/npp.2015.141

Gansler, D. A., McLaughlin, N. C. R., Iguchi, L., Jerram, M., Moore, D. W., Bhadelia, R., & Fulwiler, C. (2009). A multivariate approach to aggression and the orbital frontal cortex in psychiatric patients. *Psychiatry Research: Neuroimaging, 171*(3), 145-154. doi: http://dx.doi.org/10.1016/j.pscychresns.2008.03.007

Garandeau, C. F., Lee, I. A., & Salmivalli, C. (2014). Inequality Matters: Classroom Status Hierarchy and Adolescents' Bullying. *Journal of Youth and Adolescence, 43*(7), 1123-1133. doi: 10.1007/s10964-013-0040-4

García, A. M., Sedeño, L., Herrera Murcia, E., Couto, B., & Ibáñez, A. (2017). A Lesion-Proof Brain? Multidimensional Sensorimotor, Cognitive, and Socio-Affective Preservation Despite Extensive Damage in a Stroke Patient. [Original Research]. *Frontiers in Aging Neuroscience, 8*(335). doi: 10.3389/fnagi.2016.00335

Geniole, S. N., Carré, J. M., & McCormick, C. M. (2011). State, not trait, neuroendocrine function predicts costly reactive aggression in men after social exclusion and inclusion. *Biological Psychology, 87*(1), 137-145. doi: 10.1016/j.biopsycho.2011.02.020

Giancola, P. R., & Parrott, D. J. (2008). Further evidence for the validity of the Taylor Aggression Paradigm. *Aggressive Behavior, 34*(2), 214-229. doi: 10.1002/ab.20235
Gilbert, F., & Daffern, M. (2010). Integrating contemporary aggression theory with violent offender treatment: How thoroughly do interventions target violent behavior? *Aggression and Violent Behavior, 15*(3), 167-180. doi: https://doi.org/10.1016/j.avb.2009.11.003
Gladue, B. A. (1991). Aggressive behavioral characteristics, hormones, and sexual orientation in men and women. *Aggressive Behavior, 17*(6), 313-326. doi: 10.1002/1098-2337(1991)17:6<313::aid-ab2480170603>3.0.co;2-z
Gläscher, J. (2009). Visualization of Group Inference Data in Functional Neuroimaging. *Neuroinformatics, 7*(1), 73-82. doi: 10.1007/s12021-008-9042-x
Glass, J. M., Williams, D. A., Fernandez-Sanchez, M.-L., Kairys, A., Barjola, P., Heitzeg, M. M., . . . Schmidt-Wilcke, T. (2011). Executive function in chronic pain patients and healthy controls: different cortical activation during response inhibition in fibromyalgia. *The journal of pain : official journal of the American Pain Society, 12*, 1219-1229. doi: 10.1016/j.jpain.2011.06.007
Glenn, A. L., Raine, A., Schug, R. A., Gao, Y., & Granger, D. A. (2011). Increased testosterone-to-cortisol ratio in psychopathy. *Journal of Abnormal Psychology, 120*(2), 389-399. doi: 10.1037/a0021407
Glenn, A. L., Raine, A., Yaralian, P. S., & Yang, Y. (2010). Increased Volume of the Striatum in Psychopathic Individuals. *Biological Psychiatry, 67*(1), 52-58. doi: 10.1016/j.biopsych.2009.06.018
Glick, A. R. (2015). The role of serotonin in impulsive aggression, suicide, and homicide in adolescents and adults: a literature review. *International Journal of Adolescent Medicine and Health* (Vol. 27, pp. 143).
Gluth, S., Rieskamp, J., & Büchel, C. (2013). Classic EEG motor potentials track the emergence of value-based decisions. *NeuroImage, 79*, 394-403. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.05.005
Goetz, S. M. M., Tang, L., Thomason, M. E., Diamond, M. P., Hariri, A. R., & Carré, J. M. (2014). Testosterone Rapidly Increases Neural Reactivity to Threat in Healthy Men: A Novel Two-Step Pharmacological Challenge Paradigm. *Biological Psychiatry, 76*(4), 324-331. doi: 10.1016/j.biopsych.2014.01.016
Goldstein, J. H., Davis, R. W., & Herman, D. (1975). Escalation of aggression: Experimental studies. *Journal of Personality and Social Psychology, 31*(1), 162-170. doi: 10.1037/h0076241
Gopal, A., Clark, E., Allgair, A., D'Amato, C., Furman, M., Gansler, D. A., & Fulwiler, C. (2013). Dorsal/ventral parcellation of the amygdala: Relevance to impulsivity and aggression. *Psychiatry Research: Neuroimaging, 211*(1), 24-30. doi: http://dx.doi.org/10.1016/j.pscychresns.2012.10.010
Gordis, E., Granger, D., Susman, E., & Trickett, P. (2006). Asymmetry between salivary cortisol and α-amylase reactivity to stress: Relation to aggressive behavior in adolescents. *Psychoneuroendocrinology, 31*(8), 976-987. doi: 10.1016/j.psyneuen.2006.05.010
Gospic, K., Mohlin, E., Fransson, P., Petrovic, P., Johannesson, M., & Ingvar, M. (2011). Limbic Justice—Amygdala Involvement in Immediate Rejection in the Ultimatum Game. *PLoS Biol, 9*(5), e1001054. doi: 10.1371/journal.pbio.1001054
Goswami, R., Dufort, P., Tartaglia, M. C., Green, R. E., Crawley, A., Tator, C. H., . . . Davis, K. D. (2016). Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions. [journal article]. *Brain Structure and Function, 221*(4), 1911-1925. doi: 10.1007/s00429-015-1012-0
Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H., & Salazar, A. (1996). Frontal lobe injuries, violence, and aggression a report of the vietnam head injury study. *Neurology, 46*(5), 1231-1231.
Graham, K., Tremblay, P. F., Wells, S., Pernanen, K., Purcell, J., & Jelley, J. (2006). Harm, Intent, and the Nature of Aggressive Behavior. *Assessment, 13*(3), 280-296. doi: 10.1177/1073191106288180
Granger, D. A., Kivlighan, K. T., Fortunato, C., Harmon, A. G., Hibel, L. C., Schwartz, E. B., & Whembolua, G.-L. (2007). Integration of salivary biomarkers into developmental and behaviorally-oriented research: Problems and solutions for collecting specimens. *Physiology & Behavior, 92*(4), 583-590. doi: http://dx.doi.org/10.1016/j.physbeh.2007.05.004
Grecucci, A., Giorgetta, C., Bonini, N., & Sanfey, A. G. (2013). Reappraising social emotions: the role of inferior frontal gyrus, temporo-parietal junction and insula in interpersonal emotion regulation. [Original Research]. *Frontiers in Human Neuroscience, 7*. doi: 10.3389/fnhum.2013.00523
Gregg, T. R., & Siegel, A. (2001). Brain structures and neurotansmitters regulating aggression in cats: implications for human aggression. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 25*(1), 91-140. doi: http://dx.doi.org/10.1016/S0278-5846(00)00150-0
Greitemeyer, T., & Sagioglou, C. (2016). Subjective socioeconomic status causes aggression: A test of the theory of social deprivation. *Journal of Personality and Social Psychology, 111*(2), 178-194. doi: 10.1037/pspi0000058
Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage, 43*(3), 509-520. doi: 10.1016/j.neuroimage.2008.07.065

Gross, C. T., & Canteras, N. S. (2012). The many paths to fear. *Nature Reviews Neuroscience, 13*(9), 651-658. doi: 10.1038/nrn3301
Gu, Y., Hu, X., Pan, W., Yang, C., Wang, L., Li, Y., & Chen, A. (2016). Neural Activities Underlying the Feedback Express Salience Prediction Errors for Appetitive and Aversive Stimuli. *Scientific Reports, 6*, 34032. doi: 10.1038/srep34032
Gul, A., & Ahmad, H. (2014). Displaced aggression predicts switching deficits in people with temporal lobe epilepsy. *Epilepsy & Behavior, 41*, 109-113. doi: http://dx.doi.org/10.1016/j.yebeh.2014.09.044
Haapasalo, J., & Tremblay, R. E. (1994). Physically aggressive boys from ages 6 to 12: Family background, parenting behavior, and prediction of delinquency. *Journal of Consulting and Clinical Psychology, 62*(5), 1044-1052. doi: 10.1037/0022-006x.62.5.1044
Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology, 35*(1), 4-26. doi: 10.1038/npp.2009.129
Halasz, J., Liposits, Z., Meelis, W., Kruk, M. R., & Haller, J. C. A. (2002). Hypothalamic attack area-mediated activation of the forebrain in aggression. *Neuroreport, 13*(10), 1267-1270.
Halevy, N. (2017). Preemptive strikes: Fear, hope, and defensive aggression. *Journal of Personality and Social Psychology, 112*(2), 224-237. doi: 10.1037/pspi0000077
Halko, M.-L., Hlushchuk, Y., Hari, R., & Schürmann, M. (2009). Competing with peers: Mentalizing-related brain activity reflects what is at stake. *NeuroImage, 46*(2), 542-548. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.01.063
Haller, J., Ráczkevy-Deák, G., Gyimesine, K. P., Szakhmary, A., Farkas, I., & Vegh, J. (2014). Cardiac autonomic functions and the emergence of violence in a highly realistic model of social conflict in humans. [Original Research]. *Frontiers in Behavioral Neuroscience, 8*. doi: 10.3389/fnbeh.2014.00364
Hamann, J. M., Dayan, E., Hummel, F. C., & Cohen, L. G. (2014). Baseline frontostriatal-limbic connectivity predicts reward-based memory formation. *Human brain mapping, 35*(12), 5921-5931. doi: 10.1002/hbm.22594
Hare, B. (2017). Survival of the Friendliest: Homo sapiens Evolved via Selection for Prosociality. *Annual review of psychology, 68*(1), null. doi: doi:10.1146/annurev-psych-010416-044201
Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and Individual Differences, 35*(5), 995-1005. doi: 10.1016/S0191-8869(02)00313-6
Harrison, L. A., Ahn, C., & Adolphs, R. (2015). Exploring the Structure of Human Defensive Responses from Judgments of Threat Scenarios. *PLoS ONE, 10*(8), e0133682. doi: 10.1371/journal.pone.0133682
Hawkins, K. A., & Trobst, K. K. (2000). Frontal lobe dysfunction and aggression: Conceptual issues and research findings. *Aggression and Violent Behavior, 5*(2), 147-157. doi: http://dx.doi.org/10.1016/S1359-1789(98)00033-0
Heller, A. S., Fox, A. S., Wing, E. K., McQuisition, K. M., Vack, N. J., & Davidson, R. J. (2015). The Neurodynamics of Affect in the Laboratory Predicts Persistence of Real-World Emotional Responses. *The Journal of Neuroscience, 35*(29), 10503-10509. doi: 10.1523/jneurosci.0569-15.2015
Hellhammer, J., Fries, E., Schweisthal, O. W., Schlotz, W., Stone, A. A., & Hagemann, D. (2007). Several daily measurements are necessary to reliably assess the cortisol rise after awakening: State- and trait components. *Psychoneuroendocrinology, 32*(1), 80-86. doi: http://dx.doi.org/10.1016/j.psyneuen.2006.10.005
Henckens, M. J. A. G., Klumpers, F., Everaerd, D., Kooijman, S. C., van Wingen, G. A., & Fernández, G. (2016). Interindividual differences in stress sensitivity: basal and stress-induced cortisol levels differentially predict neural vigilance processing under stress. *Social Cognitive and Affective Neuroscience, 11*(4), 663-673. doi: 10.1093/scan/nsv149
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2-3), 61-83. doi: 10.1017/s0140525x0999152x
Hermans, E. J., Putman, P., Baas, J. M., Gecks, N. M., Kenemans, J. L., & van Honk, J. (2007). Exogenous testosterone attenuates the integrated central stress response in healthy young women. *Psychoneuroendocrinology, 32*(8-10), 1052-1061. doi: 10.1016/j.psyneuen.2007.08.006
Hermans, E. J., Putman, P., Baas, J. M., Koppeschaar, H. P., & van Honk, J. (2006). A Single Administration of Testosterone Reduces Fear-Potentiated Startle in Humans. *Biological Psychiatry, 59*(9), 872-874. doi: http://dx.doi.org/10.1016/j.biopsych.2005.11.015
Hermans, E. J., Ramsey, N. F., & van Honk, J. (2008). Exogenous Testosterone Enhances Responsiveness to Social Threat in the Neural Circuitry of Social Aggression in Humans. *Biological Psychiatry, 63*(3), 263-270. doi: 10.1016/j.biopsych.2007.05.013
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. doi: doi:10.1024/0170-1789.24.4.311

Hillbrand, M., Waite, B. M., Miller, D. S., Spitz, R. T., & Lingswiler, V. M. (2000). Serum Cholesterol Concentrations and Mood States in Violent Psychiatric Patients: An Experience Sampling Study. *Journal of Behavioral Medicine, 23*(6), 519-529. doi: 10.1023/a:1005551418922

Hillman, K. L. (2013). Cost-benefit analysis: the first real rule of fight club? [Review]. *Frontiers in Neuroscience, 7*. doi: 10.3389/fnins.2013.00248

Ho, N.-F., Chong, J. S. X., Koh, H. L., Koukouna, E., Lee, T.-S., Fung, D., . . . Zhou, J. (2015). Intrinsic Affective Network Is Impaired in Children with Attention-Deficit/Hyperactivity Disorder. *PLoS ONE, 10*(9), e0139018. doi: 10.1371/journal.pone.0139018

Hofman, D., & Schutter, D. J. L. G. (2009). Inside the wire: Aggression and functional interhemispheric connectivity in the human brain. *Psychophysiology, 46*(5), 1054-1058. doi: 10.1111/j.1469-8986.2009.00849.x

Hofman, D., Terburg, D., van Wielink, L., & Schutter, D. J. L. G. (2013). Coalescence of dominance motivation and responses to facial anger in resting-state and event-related electrophysiology. *NeuroImage, 79*, 138-144. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.04.088

Holz, N., Boecker, R., Buchmann, A. F., Blomeyer, D., Baumeister, S., Hohmann, S., . . . Laucht, M. (2016). Evidence for a Sex-Dependent MAOA× Childhood Stress Interaction in the Neural Circuitry of Aggression. *Cerebral Cortex, 26*(3), 904-914. doi: 10.1093/cercor/bhu249

Hoptman, M. J., D'Angelo, D., Catalano, D., Mauro, C. J., Shehzad, Z. E., Kelly, A. M. C., . . . Milham, M. P. (2010). Amygdalofrontal Functional Disconnectivity and Aggression in Schizophrenia. *Schizophrenia Bulletin, 36*(5), 1020-1028. doi: 10.1093/schbul/sbp012

Hornung, J.-P. (2003). The human raphe nuclei and the serotonergic system. *Journal of Chemical Neuroanatomy, 26*(4), 331-343. doi: http://dx.doi.org/10.1016/j.jchemneu.2003.10.002

Hortensius, R., Schutter, D. J. L. G., & Harmon-Jones, E. (2011). When anger leads to aggression: induction of relative left frontal cortical activity with transcranial direct current stimulation increases the anger–aggression relationship. *Social Cognitive and Affective Neuroscience, 7*(3), 342-347. doi: 10.1093/scan/nsr012

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology, 72*(2), 408.

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage, 84*, 1061-1069. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.12.075

Hyde, L. W., Byrd, A. L., Votruba-Drzal, E., Hariri, A. R., & Manuck, S. B. (2014). Amygdala reactivity and negative emotionality: Divergent correlates of antisocial personality and psychopathy traits in a community sample. *Journal of Abnormal Psychology, 123*(1), 214-224. doi: 10.1037/a0035467

Iyer, A., Lindner, A., Kagan, I., & Andersen, R. A. (2010). Motor Preparatory Activity in Posterior Parietal Cortex is Modulated by Subjective Absolute Value. *PLoS Biol, 8*(8), e1000444. doi: 10.1371/journal.pbio.1000444

Izuma, K., Saito, D. N., & Sadato, N. (2008). Processing of Social and Monetary Rewards in the Human Striatum. *Neuron, 58*(2), 284-294. doi: 10.1016/j.neuron.2008.03.020

Jacobson, L., Javitt, D. C., & Lavidor, M. (2011). Activation of Inhibition: Diminishing Impulsive Behavior by Direct Current Stimulation over the Inferior Frontal Gyrus. *Journal of cognitive neuroscience, 23*(11), 3380-3387. doi: 10.1162/jocn_a_00020

Josephs, R. A., Sellers, J. G., Newman, M. L., & Mehta, P. H. (2006). The Mismatch Effect: When Testosterone and Status Are at Odds. *Journal of Personality and Social Psychology, 90*(6), 999-1013. doi: 10.1037/0022-3514.90.6.999

Juhász, C., Behen, M. E., Muzik, O., Chugani, D. C., & Chugani, H. T. (2001). Bilateral Medial Prefrontal and Temporal Neocortical Hypometabolism in Children with Epilepsy and Aggression. *Epilepsia, 42*(8), 991-1001. doi: 10.1046/j.1528-1157.2001.042008991.x

Jung, Y.-C. a., Ku, J. c., Namkoong, K. a., Lee, W. c., Kim, S. I. c., & Kim, J.-J. a. b. (2010). Human orbitofrontal-striatum functional connectivity modulates behavioral persistence. *Neuroreport, 21*(7), 502-506.

Kaas, J. H. (2013). The Evolution of Brains from Early Mammals to Humans. *Wiley interdisciplinary reviews. Cognitive science, 4*(1), 33-45. doi: 10.1002/wcs.1206

Karlsgodt, K. H., Bato, A. A., Blair, M. A., DeRosse, P., Szeszko, P. R., & Malhotra, A. K. (2015). White matter microstructure in the executive network associated with aggression in healthy adolescents and young adults. *Social Cognitive and Affective Neuroscience, 10*(9), 1251-1256. doi: 10.1093/scan/nsv015

Kätsyri, J., Hari, R., Ravaja, N., & Nummenmaa, L. (2013). The Opponent Matters: Elevated fMRI Reward Responses to Winning Against a Human Versus a Computer Opponent During Interactive Video Game Playing. *Cerebral Cortex, 23*(12), 2829-2839. doi: 10.1093/cercor/bhs259

Keatley, D. A., Allom, V., & Mullan, B. (2017). The effects of implicit and explicit self-control on self-reported aggression. *Personality and Individual Differences, 107*, 154-158. doi: http://dx.doi.org/10.1016/j.paid.2016.11.046

Keevil, B., MacDonald, P., Macdowall, W., Lee, D., & Wu, F. (2013). Salivary testosterone measurement by liquid chromatography tandem mass spectrometry in adult males and females. *Annals of Clinical Biochemistry: An international journal of biochemistry and laboratory medicine, 51*(3), 368-378. doi: 10.1177/0004563213506412

Keller, J., Young, C. B., Kelley, E., Prater, K., Levitin, D. J., & Menon, V. (2013). Trait anhedonia is associated with reduced reactivity and connectivity of mesolimbic and paralimbic reward pathways. *Journal of Psychiatric Research, 47*(10), 1319-1328. doi: 10.1016/j.jpsychires.2013.05.015

Keune, P. M., van der Heiden, L., Várkuti, B., Konicar, L., Veit, R., & Birbaumer, N. (2012). Prefrontal brain asymmetry and aggression in imprisoned violent offenders. *Neuroscience letters, 515*(2), 191-195. doi: http://dx.doi.org/10.1016/j.neulet.2012.03.058

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. *PLoS Biol, 4*(8), e233. doi: 10.1371/journal.pbio.0040233

Kircher, T., Pohl, A., Krach, S., Thimm, M., Schulte-Rüther, M., Anders, S., & Mathiak, K. (2013). Affect-specific activation of shared networks for perception and execution of facial expressions. *Social Cognitive and Affective Neuroscience, 8*(4), 370-377. doi: 10.1093/scan/nss008

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine, 61*(2), 154-162.

Kivikangas, J. M., Kätsyri, J., Järvelä, S., & Ravaja, N. (2014). Gender Differences in Emotional Responses to Cooperative and Competitive Game Play. *PLoS ONE, 9*(7), e100318. doi: 10.1371/journal.pone.0100318

Kivlighan, K. T., Granger, D. A., & Booth, A. (2005). Gender differences in testosterone and cortisol response to competition. *Psychoneuroendocrinology, 30*(1), 58-71. doi: http://dx.doi.org/10.1016/j.psyneuen.2004.05.009

Knutson, B., Katovich, K., & Suri, G. (2014). Inferring affect from fMRI data. *Trends in Cognitive Sciences, 18*(8), 422-428. doi: 10.1016/j.tics.2014.04.006

Kolla, N. J., Matthews, B., Wilson, A. A., Houle, S., Michael Bagby, R., Links, P., . . . Meyer, J. H. (2015). Lower Monoamine Oxidase-A Total Distribution Volume in Impulsive and Violent Male Offenders with Antisocial Personality Disorder and High Psychopathic Traits: An [lsqb]11C[rsqb] Harmine Positron Emission Tomography Study. [Original Article]. *Neuropsychopharmacology, 40*(11), 2596-2603. doi: 10.1038/npp.2015.106

Konoshenko, M. Y., Timoshenko, T. V., & Plyusnina, I. Z. (2013). c-Fos activation and intermale aggression in rats selected for behavior toward humans. *Behavioural Brain Research, 237*, 103-106. doi: http://dx.doi.org/10.1016/j.bbr.2012.09.022

Kopsida, E., Berrebi, J., Petrovic, P., & Ingvar, M. (2016). Testosterone Administration Related Differences in Brain Activation during the Ultimatum Game. *Frontiers in Neuroscience, 10*. doi: 10.3389/fnins.2016.00066

Kose, S., Steinberg, J. L., Moeller, F. G., Gowin, J. L., Zuniga, E., Kamdar, Z. N., . . . Lane, S. D. (2015). Neural Correlates of Impulsive Aggressive Behavior in Subjects With a History of Alcohol Dependence. *Behavioral Neuroscience, 129*(2), 183-196. doi: 10.1037/bne0000038

Kotowicz, Z. (2007). The strange case of Phineas Gage. *History of the Human Sciences, 20*(1), 115-131. doi: 10.1177/0952695106075178

Koziol, L. F., Budding, D., Andreasen, N., D'Arrigo, S., Bulgheroni, S., Imamizu, H., . . . Yamazaki, T. (2014). Consensus Paper: The Cerebellum's Role in Movement and Cognition. [journal article]. *The Cerebellum, 13*(1), 151-177. doi: 10.1007/s12311-013-0511-x

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLoS ONE, 3*(7), e2597. doi: 10.1371/journal.pone.0002597

Krämer, U. M. (2008). *The Neural Basis of Human Aggression*. Göttingen: Sierke.

Krämer, U. M., Büttner, S., Roth, G., & Münte, T. F. (2008). Trait Aggressiveness Modulates Neurophysiological Correlates of Laboratory-induced Reactive Aggression in Humans. *Journal of cognitive neuroscience, 20*(8), 1464-1477. doi: 10.1162/jocn.2008.20103

Krämer, U. M., Jansma, H., Tempelmann, C., & Münte, T. F. (2007). Tit-for-tat: the neural basis of reactive aggression. *NeuroImage, 38*(1), 203-211. doi: 10.1016/j.neuroimage.2007.07.029

Krämer, U. M., Kopyciok, R., Richter, S., Rodriguez-Fornells, A., & Münte, T. F. (2011). The Role of Executive Functions in the Control of Aggressive Behavior. *Frontiers in Psychology, 2*(152). doi: 10.3389/fpsyg.2011.00152

Krämer, U. M., Riba, J., Richter, S., & Münte, T. F. (2011). An fMRI Study on the Role of Serotonin in Reactive Aggression. *PLoS ONE, 6*(11), e27668. doi: 10.1371/journal.pone.0027668

Kraus, M. W., Horberg, E. J., Goetz, J. L., & Keltner, D. (2011). Social Class Rank, Threat Vigilance, and Hostile Reactivity. *Personality and Social Psychology Bulletin, 37*(10), 1376-1388. doi: 10.1177/0146167211410987

Kringelbach, M. L. (2005). The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience, 6*(9), 691-702. doi: 10.1038/nrn1747

Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. *Brain structure & function, 214*, 519-534. doi: 10.1007/s00429-010-0255-z

Kuznetsova, A., Christensen, R. H. B., Bavay, C., & Brockhoff, P. B. (2015). Automated mixed ANOVA modeling of sensory and consumer data. *Food Quality and Preference, 40, Part A*, 31-38. doi: http://dx.doi.org/10.1016/j.foodqual.2014.08.004

Lackner, C. L., Marshall, W. J., Santesso, D. L., Dywan, J., Wade, T., & Segalowitz, S. J. (2014). Adolescent anxiety and aggression can be differentially predicted by electrocortical phase reset variables. *Brain and Cognition, 89*, 90-98. doi: http://dx.doi.org/10.1016/j.bandc.2013.10.004

Lamb, R., Annetta, L., Hoston, D., Shapiro, M., & Matthews, B. (2017). Examining human behavior in video games: The development of a computational model to measure aggression. *Social Neuroscience*, 1-17. doi: 10.1080/17470919.2017.1318777

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502. doi: http://dx.doi.org/10.1016/j.neuroimage.2010.10.014

Lang, P. J., & Bradley, M. M. (2013). Appetitive and Defensive Motivation: Goal-Directed or Goal-Determined? *Emotion Review, 5*(3), 230-234. doi: 10.1177/1754073913477511

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Technical manual and affective ratings. *Gainesville, FL: The Center for Research in Psychophysiology, University of Florida*.

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

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

Lawrence, C. (2006). Measuring Individual Responses to Aggression-Triggering Events: Development of the Situational Triggers of Aggressive Responses (STAR) Scale. *Aggressive Behavior, 32*, 241-252.

LeDoux, J. E. (2003). The Emotional Brain, Fear, and the Amygdala. [journal article]. *Cellular and Molecular Neurobiology, 23*(4), 727-738. doi: 10.1023/a:1025048802629

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences, 111*(8), 2871-2878. doi: 10.1073/pnas.1400335111

Lee, B. X. (2016). Causes and cures IX: Consequences of violence. *Aggression and Violent Behavior, 30*, 110-114. doi: http://dx.doi.org/10.1016/j.avb.2016.06.013

Lee, R., Arfanakis, K., Evia, A. M., Fanning, J., Keedy, S., & Coccaro, E. F. (2016). White Matter Integrity Reductions in Intermittent Explosive Disorder. *Neuropsychopharmacology, 41*(11), 2697-2703. doi: 10.1038/npp.2016.74

Leknes, S., Lee, M., Berna, C., Andersson, J., & Tracey, I. (2011). Relief as a Reward: Hedonic and Neural Responses to Safety from Pain. *PLoS ONE, 6*(4), e17870. doi: 10.1371/journal.pone.0017870

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software; Vol 1, Issue 1 (2016)*. doi: 10.18637/jss.v069.i01

Leutgeb, V., Wabnegger, A., Leitner, M., Zussner, T., Scharmüller, W., Klug, D., & Schienle, A. (2016). Altered cerebellar-amygdala connectivity in violent offenders: A resting-state fMRI study. *Neuroscience letters, 610*, 160-164. doi: http://dx.doi.org/10.1016/j.neulet.2015.10.063

Lewis, D. O. (1992). From Abuse to Violence: Psychophysiological Consequences of Maltreatment. *Journal of the American Academy of Child & Adolescent Psychiatry, 31*(3), 383-391. doi: http://dx.doi.org/10.1097/00004583-199205000-00001

Li, Y., Yuan, K., Cai, C., Feng, D., Yin, J., Bi, Y., . . . Tian, J. (2015). Reduced frontal cortical thickness and increased caudate volume within fronto-striatal circuits in young adult smokers. *Drug and Alcohol Dependence, 151*, 211-219. doi: http://dx.doi.org/10.1016/j.drugalcdep.2015.03.023

Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—A meta-analysis of neuroimaging studies. *Behavioural Brain Research, 225*(1), 341-347. doi: 10.1016/j.bbr.2011.06.022

Liddell, B. J., Brown, K. J., Kemp, A. H., Barton, M. J., Das, P., Peduto, A., . . . Williams, L. M. (2005). A direct brainstem-amygdala-cortical 'alarm' system for subliminal signals of fear. *NeuroImage, 24*(1), 235-243. doi: 10.1016/j.neuroimage.2004.08.016

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. doi: 10.1002/(sici)1098-2337(1999)25:5<331::aid-ab2>3.0.co;2-1

Lieberman, M. D. (2007). Social Cognitive Neuroscience: A Review of Core Processes. *Annual review of psychology, 58*(1), 259-289. doi: doi:10.1146/annurev.psych.58.110405.085654

Ligneul, R., Obeso, I., Ruff, C. C., & Dreher, J.-C. (2016). Dynamical Representation of Dominance Relationships in the Human Rostromedial Prefrontal Cortex. *Current Biology, 26*(23), 3107-3115. doi: http://dx.doi.org/10.1016/j.cub.2016.09.015

Lin, A., Adolphs, R., & Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsr006

Lin, D., Boyle, M. P., Dollar, P., Lee, H., Lein, E. S., Perona, P., & Anderson, D. J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. *Nature, 470*(7333), 221-226. doi: 10.1038/nature09736

Little, K., Olsson, C. A., Youssef, G. J., Whittle, S., Simmons, J. G., Yücel, M., . . . Allen, N. B. (2015). Linking the serotonin transporter gene, family environments, hippocampal volume and depression onset: A prospective imaging gene × environment analysis. *Journal of Abnormal Psychology, 124*(4), 834-849. doi: 10.1037/abn0000101

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience and biobehavioral reviews, 35*, 1219-1236. doi: 10.1016/j.neubiorev.2010.12.012

Lobbestael, J., Cousijn, J., Brugman, S., & Wiers, R. W. (2016). Approach and avoidance towards aggressive stimuli and its relation to reactive and proactive aggression. *Psychiatry Research, 240*, 196-201. doi: http://dx.doi.org/10.1016/j.psychres.2016.04.038

Lofgren, J. L. S., Erdman, S. E., Hewes, C., Wong, C., King, R., Chavarria, T. E., . . . Maurer, K. J. (2012). Castration Eliminates Conspecific Aggression in Group-Housed CD1 Male Surveillance Mice (Mus musculus). *Journal of the American Association for Laboratory Animal Science : JAALAS, 51*(5), 594-599.

Look, A. E., McCloskey, M. S., & Coccaro, E. F. (2015). Verbal versus physical aggression in Intermittent Explosive Disorder. *Psychiatry Research, 225*(3), 531-539. doi: http://dx.doi.org/10.1016/j.psychres.2014.11.052

Lorenz, K. (2005). *On Aggression*. London and New York: Routledge.

Lotze, M., Veit, R., Anders, S., & Birbaumer, N. (2007). Evidence for a different role of the ventral and dorsal medial prefrontal cortex for social reactive aggression: An interactive fMRI study. *NeuroImage, 34*, 470-478. doi: 10.1016/j.neuroimage.2006.09.028

Löw, A., Weymar, M., & Hamm, A. O. (2015). When Threat Is Near, Get Out of Here. *Psychological Science, 26*(11), 1706-1716. doi: doi:10.1177/0956797615597332

Lowe, N. C., & May, D. C. (2011). Responses to Scenarios That May Provoke Acts of Conflict and Aggression Among the General Public: An Exploratory Study. *Journal of Interpersonal Violence, 26*(8), 1606-1627. doi: doi:10.1177/0886260510370595

Ludvig, E. A. (2003). Why Pinker Needs Behaviorism: A Critique of "The Blank Slate". *Behavior and Philosophy, 31*, 139-143.

Ma, J., & Kanwal, J. (2014). Stimulation of the basal and central amygdala in the mustached bat triggers echolocation and agonistic vocalizations within multimodal output. *Frontiers in Physiology, 5*(55). doi: 10.3389/fphys.2014.00055

Macdonald, P. R., Owen, L. J., Wu, F. C., Macdowall, W., Keevil, B. G., & team, f. t. N. (2011). A Liquid Chromatography–Tandem Mass Spectrometry Method for Salivary Testosterone with Adult Male Reference Interval Determination. *Clinical Chemistry, 57*(5), 774-775. doi: 10.1373/clinchem.2010.154484

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *J Abnorm Psychol, 95*(1), 15-20.

Madipakkam, A. R., Rothkirch, M., Guggenmos, M., Heinz, A., & Sterzer, P. (2015). Gaze Direction Modulates the Relation between Neural Responses to Faces and Visual Awareness. *The Journal of Neuroscience, 35*(39), 13287-13299. doi: 10.1523/jneurosci.0815-15.2015

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

Manuck, S. B., Marsland, A. L., Flory, J. D., Gorka, A., Ferrell, R. E., & Hariri, A. R. (2010). Salivary testosterone and a trinucleotide (CAG) length polymorphism in the androgen receptor gene predict amygdala reactivity in men. *Psychoneuroendocrinology, 35*(1), 94-104. doi: 10.1016/j.psyneuen.2009.04.013

Marchewka, A., Wypych, M., Moslehi, A., Riegel, M., Michałowski, J. M., & Jednoróg, K. (2016). Arousal Rather than Basic Emotions Influence Long-Term Recognition Memory in Humans. [Original Research]. *Frontiers in Behavioral Neuroscience, 10*(198). doi: 10.3389/fnbeh.2016.00198

Marcus, B., & Schütz, A. (2005). Who Are the People Reluctant to Participate in Research? Personality Correlates of Four Different Types of Nonresponse as Inferred from Self- and Observer Ratings. *Journal of Personality, 73*(4), 959-984. doi: 10.1111/j.1467-6494.2005.00335.x

Marcus, D. K., Zeigler-Hill, V., Mercer, S. H., & Norris, A. L. (2014). The psychology of spite and the measurement of spitefulness. *Psychological Assessment, 26*(2), 563-574. doi: 10.1037/a0036039

Margulies, D. S., Vincent, J. L., Kelly, C., Lohmann, G., Uddin, L. Q., Biswal, B. B., . . . Petrides, M. (2009). Precuneus shares intrinsic functional architecture in humans and monkeys. *Proceedings of the National Academy of Sciences, 106*(47), 20069-20074. doi: 10.1073/pnas.0905314106

Marsh, A. A., Finger, E. C., Fowler, K. A., Adalio, C. J., Jurkowitz, I. T. N., Schechter, J. C., . . . Blair, R. J. R. (2013). Empathic responsiveness in amygdala and anterior cingulate cortex in youths with psychopathic traits. *Journal of Child Psychology and Psychiatry, 54*(8), 900-910. doi: 10.1111/jcpp.12063

Mason, M., Magee, J. C., & Fiske, S. T. (2014). Neural Substrates of Social Status Inference: Roles of Medial Prefrontal Cortex and Superior Temporal Sulcus. *Journal of cognitive neuroscience, 26*(5), 1131-1140. doi: 10.1162/jocn_a_00553

Matsui, F., Koh, E., Yamamoto, K., Sugimoto, K., Sin, H.-S., Maeda, Y., . . . Namiki, M. (2009). Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) Assay for Simultaneous Measurement of Salivary Testosterone and Cortisol in Healthy Men for Utilization in the Diagnosis of Late-onset Hypogonadism in Males. *Endocrine Journal, 56*(9), 1083-1093. doi: 10.1507/endocrj.K09E-186

May, A. C., Stewart, J. L., Tapert, S. F., & Paulus, M. P. (2014). The Effect of Age on Neural Processing of Pleasant Soft Touch Stimuli. *Frontiers in Behavioral Neuroscience, 8*. doi: 10.3389/fnbeh.2014.00052

McCloskey, M. S., Phan, K. L., Angstadt, M., Fettich, K. C., Keedy, S., & Coccaro, E. F. (2016). Amygdala hyperactivation to angry faces in intermittent explosive disorder. *Journal of Psychiatric Research, 79*, 34-41. doi: http://dx.doi.org/10.1016/j.jpsychires.2016.04.006

McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: the modern role of fMRI. *The Neuroscientist, 10*, 260-268. doi: 10.1177/1073858404263526

McFarquhar, M. (2016). Testable Hypotheses for Unbalanced Neuroimaging Data. *Frontiers in Neuroscience, 10*, 270. doi: 10.3389/fnins.2016.00270

McHenry, J., Carrier, N., Hull, E., & Kabbaj, M. (2014). Sex differences in anxiety and depression: Role of testosterone. *Frontiers in Neuroendocrinology, 35*(1), 42-57. doi: http://dx.doi.org/10.1016/j.yfrne.2013.09.001

McIntosh, A. R. (1999). Mapping Cognition to the Brain Through Neural Interactions. *Memory, 7*(5-6), 523-548. doi: 10.1080/096582199387733

McNaughton, N., & Corr, P. J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience & Biobehavioral Reviews, 28*(3), 285-305. doi: http://dx.doi.org/10.1016/j.neubiorev.2004.03.005

McTeague, L. M., Gruss, L. F., & Keil, A. (2015). Aversive learning shapes neuronal orientation tuning in human visual cortex. [Article]. *Nature Communications, 6*, 7823. doi: 10.1038/ncomms8823

Mehta, P. H., & Beer, J. (2009). Neural Mechanisms of the Testosterone–Aggression Relation: The Role of Orbitofrontal Cortex. *Journal of cognitive neuroscience, 22*(10), 2357-2368. doi: 10.1162/jocn.2009.21389

Mehta, P. H., Jones, A. C., & Josephs, R. A. (2008). The social endocrinology of dominance: Basal testosterone predicts cortisol changes and behavior following victory and defeat. *Journal of Personality and Social Psychology, 94*(6), 1078-1093. doi: 10.1037/0022-3514.94.6.1078

Micai, M., Kavussanu, M., & Ring, C. (2015). Executive Function is Associated With Antisocial Behavior and Aggression in Athletes. *Journal of Sport and Exercise Psychology, 37*(5), 469-476. doi: 10.1123/jsep.2015-0021

Micieli, R., Rios, A. L. L., Aguilar, R. P., Posada, L. F. B., & Hutchison, W. D. (2016). Single-unit analysis of the human posterior hypothalamus and red nucleus during deep brain stimulation for aggressivity. *Journal of Neurosurgery*, 1-7. doi: 10.3171/2016.4.jns141704

Miczek, K. A., Takahashi, A., Gobrogge, K. L., Hwa, L. S., & de Almeida, R. M. M. (2015). Escalated aggression in animal models: shedding new light on mesocorticolimbic circuits. *Current Opinion in Behavioral Sciences, 3*, 90-95. doi: http://dx.doi.org/10.1016/j.cobeha.2015.02.007

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the Hemodynamic Response: Effects of Presentation Rate, Sampling Procedure, and the Possibility of Ordering Brain Activity Based on Relative Timing. *NeuroImage, 11*(6), 735-759. doi: http://dx.doi.org/10.1006/nimg.2000.0568

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. [Review]. *Annual Review of Neuroscience, 24*, 167-202. doi: 10.1146/annurev.neuro.24.1.167

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*(2), 246-268. doi: 10.1037/0033-295x.102.2.246

Miskovic, V., & Keil, A. (2014). Escape from harm: linking affective vision and motor responses during active avoidance. *Social Cognitive and Affective Neuroscience, 9*(12), 1993-2000. doi: 10.1093/scan/nsu013

Mobbs, D., Marchant, J. L., Hassabis, D., Seymour, B., Tan, G., Gray, M., . . . Frith, C. D. (2009). From Threat to Fear: The Neural Organization of Defensive Fear Systems in Humans. *The Journal of Neuroscience, 29*(39), 12236-12243. doi: 10.1523/jneurosci.2378-09.2009

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., . . . Frith, C. D. (2007). When Fear Is Near: Threat Imminence Elicits Prefrontal-Periaqueductal Gray Shifts in Humans. *Science, 317*(5841), 1079-1083. doi: 10.1126/science.1144298

Mogg, K., Bradley, B., Miles, F., & Dixon, R. (2004). Time course of attentional bias for threat scenes: Testing the vigilance-avoidance hypothesis. *Cognition and Emotion, 18*(5), 689-700. doi: 10.1080/02699930341000158

Molenberghs, P., Trautwein, F.-M., Böckler, A., Singer, T., & Kanske, P. (2016). Neural correlates of metacognitive ability and of feeling confident: a large-scale fMRI study. *Social Cognitive and Affective Neuroscience, 11*(12), 1942-1951. doi: 10.1093/scan/nsw093

Montoya, E. R., Terburg, D., Bos, P. A., & van Honk, J. (2011). Testosterone, cortisol, and serotonin as key regulators of social aggression: A review and theoretical perspective. *Motivation and Emotion, 36*(1), 65-73. doi: 10.1007/s11031-011-9264-3

Montoya, E. R., van Honk, J., Bos, P. A., & Terburg, D. (2015). Dissociated neural effects of cortisol depending on threat escapability. *Human brain mapping, 36*(11), 4304-4316. doi: 10.1002/hbm.22918

Motzkin, J. C., Newman, J. P., Kiehl, K. A., & Koenigs, M. (2011). Reduced Prefrontal Connectivity in Psychopathy. *The Journal of Neuroscience, 31*(48), 17348-17357. doi: 10.1523/jneurosci.4215-11.2011

Mujica-Parodi, L. R., Carlson, J. M., Cha, J., & Rubin, D. (2014). The fine line between 'brave' and 'reckless': Amygdala reactivity and regulation predict recognition of risk. *NeuroImage, 103*, 1-9. doi: http://dx.doi.org/10.1016/j.neuroimage.2014.08.038

Muller, M. N., & Wrangham, R. W. (2004). Dominance, aggression and testosterone in wild chimpanzees: a test of the 'challenge hypothesis'. *Animal Behaviour, 67*(1), 113-123. doi: http://dx.doi.org/10.1016/j.anbehav.2003.03.013

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience, 9*(11), 856-869. doi: 10.1038/nrn2478

Nair, A., Keown, C. L., Datko, M., Shih, P., Keehn, B., & Müller, R.-A. (2014). Impact of methodological variables on functional connectivity findings in autism spectrum disorders. *Human brain mapping, 35*, 4035-4048. doi: 10.1002/hbm.22456

Nehrenberg, D. L., Sheikh, A., & Ghashghaei, H. T. (2013). Identification of neuronal loci involved with displays of affective aggression in NC900 mice. *Brain Structure and Function, 218*(4), 1033-1049. doi: 10.1007/s00429-012-0445-y

Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nature Reviews Neuroscience, 8*(7), 536-546. doi: 10.1038/nrn2174

Neumann, I. D., Veenema, A. H., & Beiderbeck, D. I. (2010). Aggression and anxiety: social context and neurobiological links. *Frontiers in Behavioral Neuroscience, 4*, 12. doi: 10.3389/fnbeh.2010.00012

New, A. S., Buchsbaum, M. S., Hazlett, E. A., Goodman, M., Koenigsberg, H. W., Lo, J., . . . Siever, L. J. (2004). Fluoxetine increases relative metabolic rate in prefrontal cortex in impulsive aggression. *Psychopharmacology, 176*(3), 451-458. doi: 10.1007/s00213-004-1913-8

New, A. S., Hazlett, E. A., Buchsbaum, M. S., & et al. (2002). Blunted prefrontal cortical 18fluorodeoxyglucose positron emission tomography response to meta-chlorophenylpiperazine in impulsive aggression. *Archives of general psychiatry, 59*(7), 621-629. doi: 10.1001/archpsyc.59.7.621

New, A. S., Hazlett, E. A., Buchsbaum, M. S., Goodman, M., Mitelman, S. A., Newmark, R., . . . Siever, L. J. (2007). Amygdala-Prefrontal Disconnection in Borderline Personality Disorder. *Neuropsychopharmacology, 32*(7), 1629-1640.

New, A. S., Hazlett, E. A., Newmark, R. E., Zhang, J., Triebwasser, J., Meyerson, D., . . . Buchsbaum, M. S. (2009). Laboratory Induced Aggression: A Positron Emission Tomography Study of Aggressive Individuals with Borderline Personality Disorder. *Biological Psychiatry, 66*(12), 1107-1114. doi: 10.1016/j.biopsych.2009.07.015

Nichols, A. L., & Maner, J. K. (2008). The Good-Subject Effect: Investigating Participant Demand Characteristics. *The Journal of General Psychology, 135*(2), 151-166. doi: 10.3200/genp.135.2.151-166

Niederkofler, V., Asher, Tedi E., Okaty, Benjamin W., Rood, Benjamin D., Narayan, A., Hwa, Lara S., . . . Dymecki, Susan M. (2016). Identification of Serotonergic Neuronal Modules that Affect Aggressive Behavior. *Cell Reports, 17*(8), 1934-1949. doi: 10.1016/j.celrep.2016.10.063

Nieuwenhuys, R. (2012). The insular cortex: a review. *Progress in brain research, 195*, 123-163. doi: 10.1016/B978-0-444-53860-4.00007-6

Niv, S., Ashrafulla, S., Tuvblad, C., Joshi, A., Raine, A., Leahy, R., & Baker, L. A. (2015). Childhood EEG frontal alpha power as a predictor of adolescent antisocial behavior: A twin heritability study. *Biological Psychology, 105*, 72-76. doi: http://dx.doi.org/10.1016/j.biopsycho.2014.11.010

Nivette, A. E., Eisner, M., Malti, T., & Ribeaud, D. (2014). Sex differences in aggression among children of low and high gender inequality backgrounds: A comparison of gender role and sexual selection theories. *Aggressive Behavior, 40*(5), 451-464. doi: 10.1002/ab.21530

Nolte, T., Bolling, D. Z., Hudac, C., Fonagy, P., Mayes, L. C., & Pelphrey, K. A. (2013). Brain Mechanisms Underlying the Impact of Attachment-Related Stress on Social Cognition. [Original Research]. *Frontiers in Human Neuroscience, 7*. doi: 10.3389/fnhum.2013.00816

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social cognitive and affective neuroscience, 7*, 604-609. doi: 10.1093/scan/nss055

Oliveira, T., Gouveia, M. J., & Oliveira, R. F. (2009). Testosterone responsiveness to winning and losing experiences in female soccer players. *Psychoneuroendocrinology, 34*(7), 1056-1064. doi: http://dx.doi.org/10.1016/j.psyneuen.2009.02.006

Ostrowsky, M. K. (2016). Sports Fans, Alcohol Use, and Violent Behavior. *Trauma, Violence, & Abuse*. doi: 10.1177/1524838016663937

Oyegbile, T. O., & Marler, C. A. (2005). Winning fights elevates testosterone levels in California mice and enhances future ability to win fights. *Hormones and Behavior, 48*(3), 259-267. doi: 10.1016/j.yhbeh.2005.04.007

Pabayo, R., Molnar, B. E., & Kawachi, I. (2014). The Role of Neighborhood Income Inequality in Adolescent Aggression and Violence. *Journal of Adolescent Health, 55*(4), 571-579. doi: http://dx.doi.org/10.1016/j.jadohealth.2014.04.012

Pandey, M. (2017, January 30). Senior scientists in a bare-knuckle fight inside Bhabha Atomic Research Centre's (BARC), *Mumbai Mirror*. Retrieved from http://mumbaimirror.indiatimes.com/mumbai/cover-story/senior-scientists-in-a-bare-knuckle-fight-inside-barc/articleshow/56863934.cms

Panksepp, J. (1971). Aggression elicited by electrical stimulation of the hypothalamus in albino rats. *Physiology & Behavior, 6*(4), 321-329. doi: http://dx.doi.org/10.1016/0031-9384(71)90163-6

Panksepp, J. (2011). Cross-Species Affective Neuroscience Decoding of the Primal Affective Experiences of Humans and Related Animals. *PLoS ONE, 6*(9), e21236. doi: 10.1371/journal.pone.0021236

Parasuraman, R., & Galster, S. (2013). Sensing, assessing, and augmenting threat detection: behavioral, neuroimaging, and brain stimulation evidence for the critical role of attention. [Review]. *Frontiers in Human Neuroscience, 7*. doi: 10.3389/fnhum.2013.00273

Pardini, D. A., Raine, A., Erickson, K., & Loeber, R. (2014). Lower Amygdala Volume in Men is Associated with Childhood Aggression, Early Psychopathic Traits, and Future Violence. *Biological Psychiatry, 75*(1), 73-80. doi: http://dx.doi.org/10.1016/j.biopsych.2013.04.003

Parkinson, C., & Wheatley, T. (2012). Relating Anatomical and Social Connectivity: White Matter Microstructure Predicts Emotional Empathy. *Cerebral Cortex, 24*(3), 614-625. doi: 10.1093/cercor/bhs347

Parks, M. J., Osgood, D. W., Felson, R. B., Wells, S., & Graham, K. (2013). Third Party Involvement in Barroom Conflicts. *Aggressive Behavior, 39*(4), 257-268. doi: 10.1002/ab.21475

Parmley, M., & Zhang, F. (2015). Your Face Says It All: Closeness and Perception of Emotional Expressions Among Females. *The Journal of Social Psychology, 155*(2), 127-142. doi: 10.1080/00224545.2014.972313

Parrott, D. J., & Zeichner, A. (2002). Effects of alcohol and trait anger on physical aggression in men. *Journal of Studies on Alcohol, 63*(2), 196-204.

Parsey, R. V., Oquendo, M. A., Simpson, N. R., Ogden, R. T., Van Heertum, R., Arango, V., & Mann, J. J. (2002). Effects of sex, age, and aggressive traits in man on brain serotonin 5-HT1A receptor binding potential measured by PET using [C-11]WAY-100635. *Brain Research, 954*(2), 173-182. doi: http://dx.doi.org/10.1016/S0006-8993(02)03243-2

Passamonti, L., Rowe, J. B., Ewbank, M., Hampshire, A., Keane, J., & Calder, A. J. (2008). Connectivity from the ventral anterior cingulate to the amygdala is modulated by appetitive motivation in response to facial signals of aggression. *NeuroImage, 43*(3), 562-570. doi: 10.1016/j.neuroimage.2008.07.045

Passamonti, L., Rowe, J. B., Schwarzbauer, C., Ewbank, M. P., von dem Hagen, E., & Calder, A. J. (2009). Personality Predicts the Brain's Response to Viewing Appetizing Foods: The Neural Basis of a Risk Factor for Overeating. *The Journal of Neuroscience, 29*(1), 43-51.

Pawliczek, C. M., Derntl, B., Kellermann, T., Gur, R. C., Schneider, F., & Habel, U. (2013). Anger under Control: Neural Correlates of Frustration as a Function of Trait Aggression. *PLoS ONE, 8*(10), e78503. doi: 10.1371/journal.pone.0078503

Pawliczek, C. M., Derntl, B., Kellermann, T., Kohn, N., Gur, R. C., & Habel, U. (2013). Inhibitory control and trait aggression: Neural and behavioral insights using the emotional stop signal task. *NeuroImage, 79*, 264-274. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.04.104

Payne, B. K., & Chappell, A. (2008). Using Student Samples in Criminological Research. *Journal of Criminal Justice Education, 19*(2), 175-192. doi: 10.1080/10511250802137226

Perach-Barzilay, N., Tauber, A., Klein, E., Chistyakov, A., Ne'eman, R., & Shamay-Tsoory, S. G. (2013). Asymmetry in the dorsolateral prefrontal cortex and aggressive behavior: A continuous theta-burst magnetic stimulation study. *Social Neuroscience, 8*(2), 178-188. doi: 10.1080/17470919.2012.720602

Perogamvros, I., Owen, L. J., Newell-Price, J., Ray, D. W., Trainer, P. J., & Keevil, B. G. (2009). Simultaneous measurement of cortisol and cortisone in human saliva using liquid chromatography–tandem mass spectrometry: Application in basal and stimulated conditions. *Journal of Chromatography B, 877*(29), 3771-3775. doi: http://dx.doi.org/10.1016/j.jchromb.2009.09.014

Pessoa, L. (2010). Emotion and cognition and the amygdala: From “what is it?” to “what's to be done?”. *Neuropsychologia, 48*(12), 3416-3429. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2010.06.038

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance. [10.1038/nrn2920]. *Nat Rev Neurosci, 11*(11), 773-783. doi: http://www.nature.com/nrn/journal/v11/n11/suppinfo/nrn2920_S1.html

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cognitive Brain Research, 15*(1), 31-45. doi: 10.1016/S0926-6410(02)00214-8

Peterson, C. K., Shackman, A. J., & Harmon-Jones, E. (2008). The role of asymmetrical frontal cortical activity in aggression. *Psychophysiology, 45*(1), 86-92. doi: 10.1111/j.1469-8986.2007.00597.x

Philipp-Wiegmann, F., Rösler, M., Römer, K. D., Schneider, M., Baumgart, S., & Retz, W. (2011). Reduced Cortical Inhibition in Violent Offenders: A Study with Transcranial Magnetic Stimulation. *Neuropsychobiology, 64*(2), 86-92.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking Press.

Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., . . . Fava, M. (2009). Reduced Caudate and Nucleus Accumbens Response to Rewards in Unmedicated Individuals With Major Depressive Disorder. *American Journal of Psychiatry, 166*(6), 702-710. doi: 10.1176/appi.ajp.2008.08081201

Platje, E., Vermeiren, R. R. J. M., Raine, A., Doreleijers, T. A. H., Keijsers, L. G. M. T., Branje, S. J. T., . . . Jansen, L. M. C. (2013). A longitudinal biosocial study of cortisol and peer influence on the development of adolescent antisocial behavior. *Psychoneuroendocrinology, 38*(11), 2770-2779. doi: http://dx.doi.org/10.1016/j.psyneuen.2013.07.006

Polosan, M., Baciu, M., Cousin, E., Perrone, M., Pichat, C., & Bougerol, T. (2011). An fMRI study of the social competition in healthy subjects. *Brain and Cognition, 77*(3), 401-411. doi: 10.1016/j.bandc.2011.08.018

Popma, A., Vermeiren, R., Geluk, C. A. M. L., Rinne, T., van den Brink, W., Knol, D. L., . . . Doreleijers, T. A. H. (2007). Cortisol Moderates the Relationship between Testosterone and Aggression in Delinquent Male Adolescents. *Biological Psychiatry, 61*(3), 405-411. doi: http://dx.doi.org/10.1016/j.biopsych.2006.06.006

Porges, E. C., & Decety, J. (2013). Violence as a source of pleasure or displeasure is associated with specific functional connectivity with the nucleus accumbens. *Frontiers in human neuroscience, 7*, 447. doi: 10.3389/fnhum.2013.00447

Porsch, R. M., Middeldorp, C. M., Cherny, S. S., Krapohl, E., Beijsterveldt, C. E. M. v., Loukola, A., . . . Bartels, M. (2016). Longitudinal heritability of childhood aggression. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171*(5), 697-707. doi: 10.1002/ajmg.b.32420

Potegal, M., Hebert, M., DeCoster, M., & Meyerhoff, J. L. (1996). Brief, high-frequency stimulation of the corticomedial amygdala induces a delayed and prolonged increase of aggressiveness in male Syrian golden hamsters. *Behavioral Neuroscience, 110*(2), 401-412. doi: 10.1037/0735-7044.110.2.401

Poustka, L., Maras, A., Hohm, E., Fellinger, J., Holtmann, M., Banaschewski, T., . . . Laucht, M. (2010). Negative association between plasma cortisol levels and aggression in a high-risk community sample of adolescents. *Journal of Neural Transmission, 117*(5), 621-627. doi: 10.1007/s00702-010-0386-7

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*(7), 916-931. doi: http://dx.doi.org/10.1016/S0306-4530(02)00108-7

Puhalla, A. A., Ammerman, B. A., Uyeji, L. L., Berman, M. E., & McCloskey, M. S. (2016). Negative urgency and reward/punishment sensitivity in intermittent explosive disorder. *Journal of Affective Disorders, 201*, 8-14. doi: http://dx.doi.org/10.1016/j.jad.2016.04.045

Rachlin, H. (2004). The Behavioral Economics of Violence. *Annals of the New York Academy of Sciences, 1036*(1), 325-335. doi: 10.1196/annals.1330.020

Rademacher, L., Krach, S., Kohls, G., Irmak, A., Gründer, G., & Spreckelmeyer, K. N. (2010). Dissociation of neural networks for anticipation and consumption of monetary and social rewards. *NeuroImage, 49*(4), 3276-3285. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.10.089

Radke, S., Volman, I., Mehta, P., van Son, V., Enter, D., Sanfey, A., . . . Roelofs, K. (2015). Testosterone biases the amygdala toward social threat approach. *Science Advances, 1*(5). doi: 10.1126/sciadv.1400074

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

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

Raine, A., Lencz, T., Bihrle, S., LaCasse, L., & Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Archives of general psychiatry, 57*(2), 119-127. doi: 10.1001/archpsyc.57.2.119

Raine, A., Meloy, J. R., Bihrle, S., Stoddard, J., Lacasse, L., & Buchsbaum, M. S. (1998). Reduced prefrontal and increased subcortical brain functioning assessed using positron emission tomography in predatory and affective murderers. *Behavioral Sciences & the Law, 16*(3), 319-332. doi: 10.1002/(sici)1099-0798(199822)16:3<319::aid-bsl311>3.0.co;2-g

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

Raine, A., Yang, Y., Narr, K. L., & Toga, A. W. (2011). Sex differences in orbitofrontal gray as a partial explanation for sex differences in antisocial personality. *Mol Psychiatry, 16*(2), 227-236.

Ramirez, F., Moscarello, J. M., LeDoux, J. E., & Sears, R. M. (2015). Active Avoidance Requires a Serial Basal Amygdala to Nucleus Accumbens Shell Circuit. *The Journal of Neuroscience, 35*(8), 3470-3477. doi: 10.1523/jneurosci.1331-14.2015

Rasmussen, K. (2016). Entitled vengeance: A meta-analysis relating narcissism to provoked aggression. *Aggressive Behavior, 42*(4), 362-379. doi: 10.1002/ab.21632

Rausch, J., Gäbel, A., Nagy, K., Kleindienst, N., Herpertz, S. C., & Bertsch, K. (2015). Increased testosterone levels and cortisol awakening responses in patients with borderline personality disorder: Gender and trait aggressiveness matter. *Psychoneuroendocrinology, 55*, 116-127. doi: http://dx.doi.org/10.1016/j.psyneuen.2015.02.002

Reed, L. I., DeScioli, P., & Pinker, S. A. (2014). The Commitment Function of Angry Facial Expressions. *Psychological Science, 25*(8), 1511-1517. doi: 10.1177/0956797614531027

Rinck, M., & Becker, E. S. (2006). Spider fearful individuals attend to threat, then quickly avoid it: Evidence from eye movements. *Journal of Abnormal Psychology, 115*(2), 231-238. doi: 10.1037/0021-843x.115.2.231

Rinck, M., & Becker, E. S. (2007). Approach and avoidance in fear of spiders. *J Behav Ther Exp Psychiatry, 38*(2), 105-120. doi: 10.1016/j.jbtep.2006.10.001

Ritter, D., & Eslea, M. (2005). Hot Sauce, toy guns, and graffiti: A critical account of current laboratory aggression paradigms. *Aggressive Behavior, 31*(5), 407-419. doi: 10.1002/ab.20066

Riva, P., Gabbiadini, A., Romero Lauro, L. J., Andrighetto, L., Volpato, C., & Bushman, B. J. (2016). Neuromodulation can reduce aggressive behavior elicited by violent video games. [journal article]. *Cognitive, Affective, & Behavioral Neuroscience*, 1-8. doi: 10.3758/s13415-016-0490-8

Riva, P., Romero Lauro, L. J., DeWall, C. N., Chester, D. S., & Bushman, B. J. (2014). Reducing aggressive responses to social exclusion using transcranial direct current stimulation. *Social Cognitive and Affective Neuroscience, 10*(3), 352-356. doi: 10.1093/scan/nsu053

Rivier, C., & Rivest, S. (1991). Effect of stress on the activity of the hypothalamic-pituitary-gonadal axis: peripheral and central mechanisms. *Biology of Reproduction, 45*(4), 523-532. doi: 10.1095/biolreprod45.4.523

Rodehacke, S., Mennigen, E., Müller, K. U., Ripke, S., Jacob, M. J., Hübner, T., . . . Smolka, M. N. (2014). Interindividual Differences in Mid-Adolescents in Error Monitoring and Post-Error Adjustment. *PLoS ONE, 9*(2), e88957. doi: 10.1371/journal.pone.0088957

Roelofs, K., Elzinga, B. M., & Rotteveel, M. (2005). The effects of stress-induced cortisol responses on approach-avoidance behavior. *Psychoneuroendocrinology, 30*(7), 665-677. doi: 10.1016/j.psyneuen.2005.02.008

Roelofs, K., Minelli, A., Mars, R. B., van Peer, J., & Toni, I. (2009). On the neural control of social emotional behavior. *Social Cognitive and Affective Neuroscience, 4*(1), 50-58. doi: 10.1093/scan/nsn036

Rolls, E. T., & Grabenhorst, F. (2008). The orbitofrontal cortex and beyond: From affect to decision-making. *Progress in Neurobiology, 86*(3), 216-244. doi: http://dx.doi.org/10.1016/j.pneurobio.2008.09.001

Rost, K., Stahel, L., & Frey, B. S. (2016). Digital Social Norm Enforcement: Online Firestorms in Social Media. *PLoS ONE, 11*(6), e0155923. doi: 10.1371/journal.pone.0155923

Rubinow, D. R., Roca, C. A., Schmidt, P. J., Danaceau, M. A., Putnam, K., Cizza, G., . . . Nieman, L. (2005). Testosterone Suppression of CRH-Stimulated Cortisol in Men. *Neuropsychopharmacology, 30*(10), 1906-1912.

Rudebeck, P. H., Walton, M. E., Millette, B. H. P., Shirley, E., Rushworth, M. F. S., & Bannerman, D. M. (2007). Distinct contributions of frontal areas to emotion and social behaviour in the rat. *European Journal of Neuroscience, 26*(8), 2315-2326. doi: 10.1111/j.1460-9568.2007.05844.x

Rudorf, S., & Hare, T. A. (2014). Interactions between Dorsolateral and Ventromedial Prefrontal Cortex Underlie Context-Dependent Stimulus Valuation in Goal-Directed Choice. *The Journal of Neuroscience, 34*(48), 15988-15996. doi: 10.1523/jneurosci.3192-14.2014

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature reviews. Neuroscience, advance on*. doi: 10.1038/nrn3776

Sabatinelli, D., McTeague, L. M., Dhamala, M., Frank, D. W., Wanger, T. J., & Adhikari, B. M. (2014). Reduced Medial Prefrontal–Subcortical Connectivity in Dysphoria: Granger Causality Analyses of Rapid Functional Magnetic Resonance Imaging. *Brain Connectivity*. doi: 10.1089/brain.2013.0186

Saleem, M., Anderson, C. A., & Barlett, C. P. (2015). Assessing Helping and Hurting Behaviors Through the Tangram Help/Hurt Task. *Personality and Social Psychology Bulletin, 41*(10), 1345-1362. doi: 10.1177/0146167215594348

Sano, K., Mayanagi, Y., Sekino, H., Ogashiwa, M., & Ishijima, B. (1970). Results of stimulation and destruction of the posterior hypothalamus in man. *Journal of Neurosurgery, 33*(6), 689-707. doi: 10.3171/jns.1970.33.6.0689

Sapolsky, R. M. (1982). The endocrine stress-response and social status in the wild baboon. *Hormones and Behavior, 16*(3), 279-292. doi: http://dx.doi.org/10.1016/0018-506X(82)90027-7

Sapolsky, R. M. (1991). Testicular function, social rank and personality among wild baboons. *Psychoneuroendocrinology, 16*(4), 281-293. doi: 10.1016/0306-4530(91)90015-l

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions. *Endocrine Reviews, 21*(1), 55-89. doi: doi:10.1210/edrv.21.1.0389

Scarpa, A., & Raine, A. (2007). Biosocial bases of violence. In D. J. Flannery, A. T. Vazsonyi & I. D. Waldman (Eds.), *The Cambridge handbook of violent behavior and aggression* (pp. 151-169). New York, NY, US: Cambridge University Press.

Schiffer, B., Müller, B. W., Scherbaum, N., & et al. (2011). Disentangling structural brain alterations associated with violent behavior from those associated with substance use disorders. *Archives of general psychiatry, 68*(10), 1039-1049. doi: 10.1001/archgenpsychiatry.2011.61

Schlüter, T., Winz, O., Henkel, K., Eggermann, T., Mohammadkhani-Shali, S., Dietrich, C., . . . Vernaleken, I. (2016). MAOA-VNTR polymorphism modulates context-dependent dopamine release and aggressive behavior in males. *NeuroImage, 125*, 378-385. doi: http://dx.doi.org/10.1016/j.neuroimage.2015.10.031

Schlüter, T., Winz, O., Henkel, K., Prinz, S., Rademacher, L., Schmaljohann, J., . . . Vernaleken, I. (2013). The Impact of Dopamine on Aggression: An [^{18}F]-FDOPA PET Study in Healthy Males. *The Journal of Neuroscience, 33*(43), 16889-16896. doi: 10.1523/jneurosci.1398-13.2013

Schmahmann, J. D., Weilburg, J. B., & Sherman, J. C. (2007). The neuropsychiatry of the cerebellum — insights from the clinic. [journal article]. *The Cerebellum, 6*(3), 254-267. doi: 10.1080/14734220701490995

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality, 19*(7), 595-605. doi: 10.1002/per.554

Schneider, M., Retz, W., Freitag, C., Irsch, J., Graf, P., Retz-Junginger, P., & Rösier, M. (2007). Impaired cortical inhibition in adult ADHD patients: a study with transcranial magnetic stimulation*. In M. Gerlach, J. Deckert, K. Double & E. Koutsilieri (Eds.), *Neuropsychiatric Disorders An Integrative Approach* (pp. 303-309). Vienna: Springer Vienna.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9-34. doi: http://dx.doi.org/10.1016/j.neubiorev.2014.01.009

Sescousse, G., Caldú, X., Segura, B., & Dreher, J.-C. (2013). Processing of primary and secondary rewards: A quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 37*(4), 681-696. doi: 10.1016/j.neubiorev.2013.02.002

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience, 8*(4), 300-311. doi: 10.1038/nrn2119

Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., . . . Rauch, S. L. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry, 48*(1), 43-50. doi: 10.1016/S0006-3223(00)00251-1

Shivers, M., & Edwards, D. A. (1978). Hypothalamic destruction and mouse aggression. [journal article]. *Physiological Psychology, 6*(4), 485-487. doi: 10.3758/bf03326753

Siegel, A., Roeling, T. A. P., Gregg, T. R., & Kruk, M. R. (1999). Neuropharmacology of brain-stimulation-evoked aggression. *Neuroscience & Biobehavioral Reviews, 23*(3), 359-389. doi: http://dx.doi.org/10.1016/S0149-7634(98)00040-2

Simons, J. S., Wills, T. A., Emery, N. N., & Spelman, P. J. (2015). Keep calm and carry on: Maintaining self-control when intoxicated, upset, or depleted. *Cognition and Emotion*, 1-15. doi: 10.1080/02699931.2015.1069733

Sims, T. B., Neufeld, J., Johnstone, T., & Chakrabarti, B. (2014). Autistic traits modulate frontostriatal connectivity during processing of rewarding faces. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsu010

Simunovic, D., Mifune, N., & Yamagishi, T. (2013). Preemptive strike: An experimental study of fear-based aggression. *Journal of Experimental Social Psychology, 49*(6), 1120-1123. doi: https://doi.org/10.1016/j.jesp.2013.08.003

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences, 13*(8), 334-340. doi: 10.1016/j.tics.2009.05.001

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for Pain Involves the Affective but not Sensory Components of Pain. *Science, 303*(5661), 1157-1162. doi: 10.1126/science.1093535

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466-469. doi: 10.1038/nature04271

Slap, J. W. (1967). Freud's View on Pleasure and Aggression. *Journal of the American Psychoanalytic Association, 15*(2), 370-375. doi: 10.1177/000306516701500208

Smeets, T., Dziobek, I., & Wolf, O. T. (2009). Social cognition under stress: Differential effects of stress-induced cortisol elevations in healthy young men and women. *Hormones and Behavior, 55*(4), 507-513. doi: http://dx.doi.org/10.1016/j.yhbeh.2009.01.011

Smith, D. V., Clithero, J. A., Boltuck, S. E., & Huettel, S. A. (2014). Functional connectivity with ventromedial prefrontal cortex reflects subjective value for social rewards. *Social cognitive and affective neuroscience*. doi: 10.1093/scan/nsu005

Soloff, P. H., Abraham, K., Burgess, A., Ramaseshan, K., Chowdury, A., & Diwadkar, V. A. (2017). Impulsivity and aggression mediate regional brain responses in Borderline Personality Disorder: An fMRI study. *Psychiatry Research: Neuroimaging, 260*, 76-85. doi: http://dx.doi.org/10.1016/j.pscychresns.2016.12.009

Soloff, P. H., Abraham, K., Ramaseshan, K., Burgess, A., & Diwadkar, V. A. (2017). Hyper-modulation of brain networks by the amygdala among women with Borderline Personality Disorder: Network signatures of

affective interference during cognitive processing. *Journal of Psychiatric Research, 88*, 56-63. doi: http://dx.doi.org/10.1016/j.jpsychires.2016.12.016

Soloff, P. H., White, R., Omari, A., Ramaseshan, K., & Diwadkar, V. A. (2015). Affective context interferes with brain responses during cognitive processing in borderline personality disorder: fMRI evidence. *Psychiatry Research, 233*(1), 23-35. doi: 10.1016/j.pscychresns.2015.04.006

Sperduti, M., Guionnet, S., Fossati, P., & Nadel, J. (2014). Mirror Neuron System and Mentalizing System connect during online social interaction. [journal article]. *Cognitive Processing, 15*(3), 307-316. doi: 10.1007/s10339-014-0600-x

Spielberg, J. M., Forbes, E. E., Ladouceur, C. D., Worthman, C. M., Olino, T. M., Ryan, N. D., & Dahl, R. E. (2015). Pubertal testosterone influences threat-related amygdala–orbitofrontal cortex coupling. *Social Cognitive and Affective Neuroscience, 10*(3), 408-415. doi: 10.1093/scan/nsu062

Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. [Review]. *Nat Neurosci, 18*(5), 620-627. doi: 10.1038/nn.3982

Stanton, S. J., & Edelstein, R. S. (2009). The physiology of women's power motive: Implicit power motivation is positively associated with estradiol levels in women. *Journal of Research in Personality, 43*(6), 1109-1113. doi: https://doi.org/10.1016/j.jrp.2009.08.002

Stanton, S. J., Wirth, M. M., Waugh, C. E., & Schultheiss, O. C. (2009). Endogenous testosterone levels are associated with amygdala and ventromedial prefrontal cortex responses to anger faces in men but not women. *Biological Psychology, 81*(2), 118-122. doi: 10.1016/j.biopsycho.2009.03.004

Stevens, M. C., & Haney-Caron, E. (2012). Comparison of brain volume abnormalities between ADHD and conduct disorder in adolescence. *Journal of Psychiatry & Neuroscience : JPN, 37*(6), 389-398. doi: 10.1503/jpn.110148

Strobel, A., Beauducel, A., Debener, S., & Brocke, B. (2001). Eine deutschsprachige Version des BIS/BAS-Fragebogens von Carver und White. *Zeitschrift für Differentielle und Diagnostische Psychologie, 22*(3), 216-227. doi: doi:10.1024//0170-1789.22.3.216

Strüber, D., Lück, M., & Roth, G. (2008). Sex, aggression and impulse control: An integrative account. *Neurocase, 14*(1), 93-121. doi: 10.1080/13554790801992743

Sundram, F., Deeley, Q., Sarkar, S., Daly, E., Latham, R., Craig, M., . . . Murphy, D. G. M. (2012). White matter microstructural abnormalities in the frontal lobe of adults with antisocial personality disorder. *Cortex, 48*(2), 216-229. doi: 10.1016/j.cortex.2011.06.005

Sussman, T. J., Weinberg, A., Szekely, A., Hajcak, G., & Mohanty, A. (2016). Here Comes Trouble: Prestimulus Brain Activity Predicts Enhanced Perception of Threat. *Cerebral Cortex*. doi: 10.1093/cercor/bhw104

Takahashi, A., Nagayasu, K., Nishitani, N., Kaneko, S., & Koide, T. (2014). Control of Intermale Aggression by Medial Prefrontal Cortex Activation in the Mouse. *PLoS ONE, 9*(4), e94657. doi: 10.1371/journal.pone.0094657

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., & Okubo, Y. (2009). When Your Gain Is My Pain and Your Pain Is My Gain: Neural Correlates of Envy and Schadenfreude. *Science, 323*(5916), 937-939. doi: 10.1126/science.1165604

Tavris, C. (1984). On the wisdom of counting to ten: Personal and social dangers of anger expression. *Review of Personality & Social Psychology, 5*, 170-191.

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality, 35*(2), 297-310. doi: 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. doi: 10.1016/1359-1789(95)00014-3

Terburg, D., & van Honk, J. (2013). Approach–Avoidance versus Dominance–Submissiveness: A Multilevel Neural Framework on How Testosterone Promotes Social Status. *Emotion Review, 5*(3), 296-302. doi: 10.1177/1754073913477510

Thijssen, S., Ringoot, A. P., Wildeboer, A., Bakermans-Kranenburg, M. J., El Marroun, H., Hofman, A., . . . White, T. (2015). Brain morphology of childhood aggressive behavior: A multi-informant study in school-age children. [journal article]. *Cognitive, Affective, & Behavioral Neuroscience, 15*(3), 564-577. doi: 10.3758/s13415-015-0344-9

Thomaes, S., Bushman, B. J., Stegge, H., & Olthof, T. (2008). Trumping Shame by Blasts of Noise: Narcissism, Self-Esteem, Shame, and Aggression in Young Adolescents. *Child Development, 79*(6), 1792-1801. doi: 10.1111/j.1467-8624.2008.01226.x

Tibon, R., & Levy, D. A. (2015). Striking a balance: analyzing unbalanced event-related potential data. *Frontiers in Psychology, 6*, 555. doi: 10.3389/fpsyg.2015.00555

Tidoni, E., & Candidi, M. (2016). Causal Approaches to Study Intention Understanding from Action Observation. [General Commentary]. *Frontiers in Behavioral Neuroscience, 10*. doi: 10.3389/fnbeh.2016.00013

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

Tolkien, J. R. R. (1954a). *The Fellowship of the Ring*. London: George Allen & Unwin.

Tolkien, J. R. R. (1954b). *The Two Towers*. London: George Allen & Unwin.

Tonnaer, F. M., Cima, M. P., & Arntz, A. P. (2016). Executive (Dys)Functioning and Impulsivity as Possible Vulnerability Factors for Aggression in Forensic Patients. *Journal of Nervous & Mental Disease, 204*(4), 280-286.

Tulogdi, A., Biro, L., Barsvari, B., Stankovic, M., Haller, J., & Toth, M. (2015). Neural mechanisms of predatory aggression in rats—Implications for abnormal intraspecific aggression. *Behavioural Brain Research, 283*, 108-115. doi: http://dx.doi.org/10.1016/j.bbr.2015.01.030

Twenge, J. M., Baumeister, R. F., Tice, D. M., & Stucke, T. S. (2001). If you can't join them, beat them: Effects of social exclusion on aggressive behavior. *Journal of Personality and Social Psychology, 81*(6), 1058-1069. doi: 10.1037/0022-3514.81.6.1058

Unger, Elizabeth K., Burke, Kenneth J., Jr., Yang, Cindy F., Bender, Kevin J., Fuller, Patrick M., & Shah, Nirao M. (2015). Medial Amygdalar Aromatase Neurons Regulate Aggression in Both Sexes. *Cell Reports, 10*(4), 453-462. doi: 10.1016/j.celrep.2014.12.040

Vaidyanathan, U., Patrick, C. J., & Bernat, E. M. (2009). Startle reflex potentiation during aversive picture viewing as an indicator of trait fear. *Psychophysiology, 46*(1), 75-85. doi: 10.1111/j.1469-8986.2008.00751.x

van de Giessen, E., Rosell, D. R., Thompson, J. L., Xu, X., Girgis, R. R., Ehrlich, Y., . . . Siever, L. J. (2014). Serotonin transporter availability in impulsive aggressive personality disordered patients: A PET study with [11C]DASB. *Journal of Psychiatric Research, 58*, 147-154. doi: http://dx.doi.org/10.1016/j.jpsychires.2014.07.025

van den Bos, W., Cohen, M. X., Kahnt, T., & Crone, E. A. (2012). Striatum–Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning. *Cerebral Cortex, 22*(6), 1247-1255.

van den Bos, W., Golka, P. J. M., Effelsberg, D., & McClure, S. M. (2013). Pyrrhic victories: the need for social status drives costly competitive behavior. *Frontiers in Neuroscience, 7*, 189. doi: 10.3389/fnins.2013.00189

Van den Stock, J., Hortensius, R., Sinke, C., Goebel, R., & de Gelder, B. (2015). Personality traits predict brain activation and connectivity when witnessing a violent conflict. [Article]. *Scientific Reports, 5*, 13779. doi: 10.1038/srep13779

van der Meij, L., Klauke, F., Moore, H. L., Ludwig, Y. S., Almela, M., & van Lange, P. A. M. (2015). Football Fan Aggression: The Importance of Low Basal Cortisol and a Fair Referee. *PLoS ONE, 10*(4), e0120103. doi: 10.1371/journal.pone.0120103

Van Doesum, N. J., Van Prooijen, J.-W., Verburgh, L., & Van Lange, P. A. M. (2016). Social Hostility in Soccer and Beyond. *PLoS ONE, 11*(4), e0153577. doi: 10.1371/journal.pone.0153577

van Erp, A. M. M., & Miczek, K. A. (2000). Aggressive Behavior, Increased Accumbal Dopamine, and Decreased Cortical Serotonin in Rats. *The Journal of Neuroscience, 20*(24), 9320-9325.

van Goozen, S. H. M., Fairchild, G., Snoek, H., & Harold, G. T. (2007). The evidence for a neurobiological model of childhood antisocial behavior. *Psychological Bulletin, 133*(1), 149-182. doi: 10.1037/0033-2909.133.1.149

van Holst, R. J., Chase, H. W., & Clark, L. (2014). Striatal connectivity changes following gambling wins and near-misses: Associations with gambling severity. *NeuroImage: Clinical, 5*(0), 232-239. doi: 10.1016/j.nicl.2014.06.008

van Honk, J., Harmon-Jones, E., Morgan, B. E., & Schutter, D. J. L. G. (2010). Socially Explosive Minds: The Triple Imbalance Hypothesis of Reactive Aggression. *Journal of Personality, 78*(1), 67-94. doi: 10.1111/j.1467-6494.2009.00609.x

van Honk, J., Montoya, E. R., Bos, P. A., van Vugt, M., & Terburg, D. (2012). New evidence on testosterone and cooperation. *Nature, 485*(7399), E4-E5. doi: 10.1038/nature11136

van Honk, J., & Schutter, D. J. L. G. (2007). Testosterone Reduces Conscious Detection of Signals Serving Social Correction: Implications for Antisocial Behavior. *Psychological Science, 18*(8), 663-667. doi: 10.1111/j.1467-9280.2007.01955.x

van Honk, J., Tuiten, A., de Haan, E., van den Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition and Emotion, 15*(3), 279-297. doi: 10.1080/02699930126112

van Honk, J., Tuiten, A., Hermans, E., Putnam, P., Koppeschaar, H., Thijssen, J., . . . van Doornen, L. (2001). A single administration of testosterone induces cardiac accelerative responses to angry faces in healthy young women. *Behavioral Neuroscience, 115*(1), 238-242. doi: 10.1037/0735-7044.115.1.238

van Honk, J., Tuiten, A., Verbaten, R., van den Hout, M., Koppeschaar, H., Thijssen, J., & de Haan, E. (1999). Correlations among Salivary Testosterone, Mood, and Selective Attention to Threat in Humans. *Hormones and Behavior, 36*(1), 17-24. doi: http://dx.doi.org/10.1006/hbeh.1999.1521

van Honk, J., Will, G.-J., Terburg, D., Raub, W., Eisenegger, C., & Buskens, V. (2016). Effects of Testosterone Administration on Strategic Gambling in Poker Play. [Article]. *Scientific Reports, 6*, 18096. doi: 10.1038/srep18096

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2014). Social cognition and the cerebellum: A meta-analysis of over 350 fMRI studies. *NeuroImage, 86*, 554-572. doi: http://dx.doi.org/10.1016/j.neuroimage.2013.09.033

van Wingen, G., Zylicz, S., Pieters, S., Mattern, C., Verkes, R. J., Buitelaar, J., & Fernández, G. (2008). Testosterone Increases Amygdala Reactivity in Middle-Aged Women to a Young Adulthood Level. *Neuropsychopharmacology, 34*(3), 539-547. doi: 10.1038/npp.2008.2

Vasquez, E. A., Pedersen, W. C., Bushman, B. J., Kelley, N. J., Demeestere, P., & Miller, N. (2013). Lashing Out after Stewing over Public Insults: The Effects of Public Provocation, Provocation Intensity, and Rumination on Triggered Displaced Aggression. *Aggressive Behavior, 39*(1), 13-29. doi: 10.1002/ab.21453

Vassos, E., Collier, D. A., & Fazel, S. (2014). Systematic meta-analyses and field synopsis of genetic association studies of violence and aggression. [Original Article]. *Molecular Psychiatry, 19*(4), 471-477. doi: 10.1038/mp.2013.31

Veit, R., Lotze, M., Sewing, S., Missenhardt, H., Gaber, T., & Birbaumer, N. (2010). Aberrant social and cerebral responding in a competitive reaction time paradigm in criminal psychopaths. *NeuroImage, 49*, 3365-3372. doi: 10.1016/j.neuroimage.2009.11.040

Velicer, W. F., Govia, J. M., Cherico, N. P., & Corriveau, D. P. (1985). Item format and the structure of the buss-durkee hostility inventory. *Aggressive Behavior, 11*(1), 65-82.

Verleger, R., Jaskowski, P., & Wauschkuhn, B. (1994). Suspense and surprise: On the relationship between expectancies and P3. *Psychophysiology, 31*(4), 359-369. doi: 10.1111/j.1469-8986.1994.tb02444.x

Viau, V. (2002). Functional Cross-Talk Between the Hypothalamic-Pituitary-Gonadal and -Adrenal Axes. *Journal of Neuroendocrinology, 14*(6), 506-513. doi: 10.1046/j.1365-2826.2002.00798.x

Vigil-Colet, A., Ruiz-Pamies, M., Anguiano-Carrasco, C., & Lorenzo-Seva, U. (2012). The impact of social desirability on psychometric measures of aggression. *Psicothema, 24*(2), 310-315.

Volkow, N. D., Tancredib, L. R., Grant, C., Gillespie, H., Valentine, A., Mullani, N., . . . Hollister, L. (1995). Brain glucose metabolism in violent psychiatric patients: a preliminary study. *Psychiatry Research: Neuroimaging, 61*(4), 243-253. doi: http://dx.doi.org/10.1016/0925-4927(95)02671-J

Völlm, B., Richardson, P., Stirling, J., Elliott, R., Dolan, M., Chaudhry, I., . . . Deakin, B. (2004). Neurobiological substrates of antisocial and borderline personality disorder: preliminary results of a functional fMRI study. *Criminal Behaviour and Mental Health, 14*(1), 39-54. doi: 10.1002/cbm.559

Volman, I., Toni, I., Verhagen, L., & Roelofs, K. (2011). Endogenous Testosterone Modulates Prefrontal-Amygdala Connectivity during Social Emotional Behavior. *Cerebral Cortex, 21*(10), 2282-2290. doi: 10.1093/cercor/bhr001

Von Der Heide, R. J., Skipper, L. M., Klobusicky, E., & Olson, I. R. (2013). Dissecting the uncinate fasciculus: disorders, controversies and a hypothesis. *Brain, 136*(6), 1692-1707. doi: 10.1093/brain/awt094

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*(12), 585-594. doi: 10.1016/j.tics.2005.10.011

Waltes, R., Chiocchetti, A. G., & Freitag, C. M. (2016). The neurobiological basis of human aggression: A review on genetic and epigenetic mechanisms. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 171*(5), 650-675. doi: 10.1002/ajmg.b.32388

Wan, X., Cheng, K., & Tanaka, K. (2016). The Neural System of Post-Decision Evaluation in Rostral Frontal Cortex during Problem Solving Tasks. *eneuro*. doi: 10.1523/eneuro.0188-16.2016

Wang, Q., Verweij, E. W. E., Krugers, H. J., Joels, M., Swaab, D. F., & Lucassen, P. J. (2013). Distribution of the glucocorticoid receptor in the human amygdala; changes in mood disorder patients. *Brain Structure and Function, 219*(5), 1615-1626. doi: 10.1007/s00429-013-0589-4

Ward, J. T., Krohn, M. D., & Gibson, C. L. (2013). The Effects of Police Contact on Trajectories of Violence. *Journal of Interpersonal Violence, 29*(3), 440-475. doi: 10.1177/0886260513505219

Waters, H. R., Hyder, A. A., Rajkotia, Y., Basu, S., & Butchart, A. (2005). The costs of interpersonal violence—an international review. *Health Policy, 73*(3), 303-315. doi: http://dx.doi.org/10.1016/j.healthpol.2004.11.022

Webster, G. D., DeWall, C. N., Pond, R. S., Deckman, T., Jonason, P. K., Le, B. M., . . . Bator, R. J. (2014). The brief aggression questionnaire: psychometric and behavioral evidence for an efficient measure of trait aggression. *Aggressive Behavior, 40*(2), 120-139. doi: 10.1002/ab.21507

Weiland, B. J., Welsh, R. C., Yau, W.-Y. W., Zucker, R. A., Zubieta, J.-K., & Heitzeg, M. M. (2013). Accumbens functional connectivity during reward mediates sensation-seeking and alcohol use in high-risk youth. *Drug and Alcohol Dependence, 128*(1–2), 130-139. doi: 10.1016/j.drugalcdep.2012.08.019

Weldon, A. L., Hagan, M., Van Meter, A., Jacobs, R. H., Kassel, M. T., Hazlett, K. E., . . . Langenecker, S. A. (2015). Stress Response to the Functional Magnetic Resonance Imaging Environment in Healthy Adults Relates to the Degree of Limbic Reactivity during Emotion Processing. *Neuropsychobiology, 71*(2), 85-96.

Welker, K. M., Norman, R. E., Goetz, S., Moreau, B. J. P., Kitayama, S., & Carré, J. M. (2016). Preliminary evidence that testosterone's association with aggression depends on self-construal. *Hormones and Behavior*. doi: 10.1016/j.yhbeh.2016.10.014

Weymar, M., & Schwabe, L. (2016). Amygdala and emotion: the bright side of it. [Frontiers Commentary]. *Frontiers in Neuroscience, 10*. doi: 10.3389/fnins.2016.00224

White, S. F., Brislin, S. J., Sinclair, S., & Blair, R. J. R. (2014). Punishing unfairness: rewarding or the organization of a reactively aggressive response? *Human brain mapping, 35*, 2137-2147. doi: 10.1002/hbm.22316

White, S. F., VanTieghem, M., Brislin, S. J., Sypher, I., Sinclair, S., Pine, D. S., . . . Blair, R. J. R. (2015). Neural Correlates of the Propensity for Retaliatory Behavior in Youths With Disruptive Behavior Disorders. *American Journal of Psychiatry, 173*(3), 282-290. doi: 10.1176/appi.ajp.2015.15020250

Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain connectivity, 2*, 125-141. doi: 10.1089/brain.2012.0073

Widom, C. S. (1989). Child abuse, neglect, and adult behavior: research design and findings on criminality, violence, and child abuse. *American Journal of Orthopsychiatry, 59*(3), 355.

Wiecko, F. M. (2010). Research Note: Assessing the validity of college samples: Are students really that different? *Journal of Criminal Justice, 38*(6), 1186-1190. doi: http://dx.doi.org/10.1016/j.jcrimjus.2010.09.007

Wilkowski, B. M., & Robinson, M. D. (2010). The Anatomy of Anger: An Integrative Cognitive Model of Trait Anger and Reactive Aggression. *Journal of Personality, 78*(1), 9-38. doi: 10.1111/j.1467-6494.2009.00607.x

Williams, K. D. (2007). Ostracism. *Annual review of psychology, 58*(1), 425-452. doi: doi:10.1146/annurev.psych.58.110405.085641

Williams, R. (2009, May 17). Zidane shocker - France, 2006, *The Guardian*. Retrieved from http://www.theguardian.com/football/2009/may/17/seven-deadly-sins-zidane-materazzi-pride-world-cup

Winecoff, A., Clithero, J. A., Carter, R. M., Bergman, S. R., Wang, L., & Huettel, S. A. (2013). Ventromedial prefrontal cortex encodes emotional value. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 33*, 11032-11039. doi: 10.1523/JNEUROSCI.4317-12.2013

Wirth, M. M., & Schultheiss, O. C. (2007). Basal testosterone moderates responses to anger faces in humans. *Physiology & Behavior, 90*(2-3), 496-505. doi: 10.1016/j.physbeh.2006.10.016

Wiswede, D., Taubner, S., Münte, T. F., Roth, G., Strüber, D., Wahl, K., & Krämer, U. M. (2011). Neurophysiological correlates of laboratory-induced aggression in young men with and without a history of violence. *PloS one, 6*, e22599. doi: 10.1371/journal.pone.0022599

Witte, A. V., Flöel, A., Stein, P., Savli, M., Mien, L.-K., Wadsak, W., . . . Lanzenberger, R. (2009). Aggression is related to frontal serotonin-1A receptor distribution as revealed by PET in healthy subjects. *Human brain mapping, 30*(8), 2558-2570. doi: 10.1002/hbm.20687

Woodworth, C. H. (1971). Attack elicited in rats by electrical stimulation of the lateral hypothalamus. *Physiology & Behavior, 6*(4), 345-353. doi: http://dx.doi.org/10.1016/0031-9384(71)90166-1

Wu, Y., Eisenegger, C., Zilioli, S., Watson, N. V., & Clark, L. (2016). Comparison of clear and narrow outcomes on testosterone levels in social competition. *Hormones and Behavior*. doi: http://dx.doi.org/10.1016/j.yhbeh.2016.05.016

Yang, D. Y. J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews, 51*, 263-275. doi: http://dx.doi.org/10.1016/j.neubiorev.2015.01.020

Yang, Y., Raine, A., Narr, K. L., Colletti, P., & Toga, A. W. (2009). Localization of Deformations Within the Amygdala in Individuals With Psychopathy. *Archives of general psychiatry, 66*(9), 986-994. doi: 10.1001/archgenpsychiatry.2009.110

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral Disinhibition: Liability for Externalizing Spectrum Disorders and Its Genetic and Environmental Relation to Response Inhibition Across Adolescence. *Journal of Abnormal Psychology, 118*(1), 117-130. doi: 10.1037/a0014657

Yu, Q., Teixeira, C. M., Mahadevia, D., Huang, Y., Balsam, D., Mann, J. J., . . . Ansorge, M. S. (2014). Dopamine and serotonin signaling during two sensitive developmental periods differentially impact adult aggressive and affective behaviors in mice. *Mol Psychiatry, 19*(6), 688-698. doi: 10.1038/mp.2014.10

Yu, R., Mobbs, D., Seymour, B., Rowe, J. B., & Calder, A. J. (2014). The neural signature of escalating frustration in humans. *Cortex, 54*, 165-178. doi: http://dx.doi.org/10.1016/j.cortex.2014.02.013

Zaki, J., & Ochsner, K. (2012). The neuroscience of empathy: progress, pitfalls and promise. [10.1038/nn.3085]. *Nat Neurosci, 15*(5), 675-680.

Zeichner, A., Frey, F. C., Parrott, D. J., & Butryn, M. F. (1999). Measurement of Laboratory Aggression: A New Response-Choice Paradigm. *Psychological Reports, 85*(3_suppl), 1229-1237. doi: doi:10.2466/pr0.1999.85.3f.1229

Zhang, L., Kerich, M., Schwandt, M. L., Rawlings, R. R., McKellar, J. D., Momenan, R., . . . George, D. T. (2013). Smaller right amygdala in Caucasian alcohol-dependent male patients with a history of intimate partner violence: a volumetric imaging study. *Addiction Biology, 18*(3), 537-547. doi: 10.1111/j.1369-1600.2011.00381.x

Zilioli, S., Ponzi, D., Henry, A., & Maestripieri, D. (2014). Testosterone, Cortisol and Empathy: Evidence for the Dual-Hormone Hypothesis. *Adaptive Human Behavior and Physiology*, 1-13. doi: 10.1007/s40750-014-0017-x

Zilioli, S., & Watson, N. V. (2014). Testosterone across successive competitions: Evidence for a 'winner effect' in humans? *Psychoneuroendocrinology, 47*(0), 1-9. doi: http://dx.doi.org/10.1016/j.psyneuen.2014.05.001

Zilles, K., Palomero-Gallagher, N., & Amunts, K. (2013). Development of cortical folding during evolution and ontogeny. *Trends in Neurosciences, 36*(5), 275-284. doi: http://dx.doi.org/10.1016/j.tins.2013.01.006

Zimmerman, Y., Eijkemans, M. J. C., Coelingh Bennink, H. J. T., Blankenstein, M. A., & Fauser, B. C. J. M. (2013). The effect of combined oral contraception on testosterone levels in healthy women: a systematic review and meta-analysis. *Human Reproduction Update, 20*(1), 76-105. doi: 10.1093/humupd/dmt038

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know Your Place: Neural Processing of Social Hierarchy in Humans. *Neuron, 58*(2), 273-283. doi: http://dx.doi.org/10.1016/j.neuron.2008.01.025