The twentieth century will be remembered as a century marked by violence. It burdens us with its legacy of mass destruction, of violence inflicted on a scale never seen and never possible before in human history. But this legacy – the result of new technology in the service of ideologies of hate – is not the only one we carry, nor that we must face up to. (…) We owe our children – the most vulnerable citizens in any society – a life free from violence and fear. In order to ensure this, we must be tireless in our efforts not only to attain peace, justice and prosperity for countries, but also for communities and members of the same family. We must address the roots of violence. Only then will we transform the past century’s legacy from a crushing burden into a cautionary lesson.

CHAPTER 1

Background

Aggression is of all times, species and cultures, and has puzzled many researchers and policy makers for the fact that it poses severe problems to society (Buss & Duntley, 2006; Buss & Shackelford, 1997; Goldstein, 1998; Sluyter & Schalkwyk, 2003; Stone, 1976). For example, aggression may result in violent crimes, which can have a major short- and long-term impact on its victims (Butchart & Mikton, 2014; World report on violence and health: summary, 2002). Social scientists have investigated aggression for a long time, resulting in many theories, underlying social, biological and psychological mechanisms, as well as in treatment programs to prevent or reduce problematic aggressive behaviour (Butchart & Mikton, 2014; World report on violence and health: summary, 2002). With advancements in forensic neuroscience in recent years, more attention is paid to specific characteristics of ‘the violent brain’, including its structural and functional neuro-anatomy, neurochemistry and neurocognition, including behaviour (Raine, 2013). Although all this research has resulted in many new scientific insights (Hirschtritt, Carroll, & Ross, 2018; Ogilvie, Stewart, Chan, & Shum, 2011; Raine & Yang, 2006), the translation of this knowledge to clinical forensic practice (e.g. to identify those at risk for re-offending and/or treat aggressive behaviour in offenders) still lags behind (Hirschtritt et al., 2018).

In the present thesis the general aim is to identify some of the neurocognitive building blocks of aggression that may have potential as targets for future interventions in offenders. More specifically, the focus lies on particular aspects of social cognition that have been pointed out as important factors contributing to aggression: interpretation biases in facial emotion perception and risky decision making (Bass & Nussbaum, 2010; Broomhall, 2005; Penton-Voak et al., 2013; Smeijers, Rinck, Bulten, Van den Heuvel, & Verkes, 2017). In the following sections, the main concepts in this thesis will be explained in more detail, including the links between them. Building on this foundation, the current topics of investigation will be summed up at the end of this introduction.

Aggression and violence

While in the scientific literature multiple, interchangeable definitions are applied to explain aggressive tendencies (Norlander & Eckhardt, 2005), ‘aggression’ is most often defined as: ‘any form of behaviour directed toward the goal of harming or injuring another living being who is motivated to avoid such treatment’ (Baron & Richardson, 1994). Aggression in its most extreme form is ‘violence’, which has severe harm as its goal (e.g. death). All violence is aggression, but many instances of aggression are not violent (Anderson & Bushman, 2002). In the present thesis the
main focus lies on the broad concept of aggression, which therefore also implicitly includes violence. Apart from these behavioural expressions of aggression, the concept is also more broadly linked to internal states and feelings, such as anger and hostility.

Aggression is often subdivided into aggression with a premeditated character to gain an intended advantage (proactive or instrumental aggression) and aggression as a result of perceived provocation/threat or intense emotion and lack of impulse control (hostile, reactive or impulsive aggression) (Anderson & Bushman, 2002; Dodge, 1991). Although the distinction between these two major types of aggression is insightful, it leaves the matter of underlying mechanisms of aggression largely unattended. One of the most well-known theoretical models that does provide an integrated overview of all components involved in eliciting aggression is called the ‘General Aggression Model’ (Anderson & Bushman, 2002; DeWall & Anderson, 2011). This model can be used to explain how aggression unfolds within one cycle of an ongoing social interaction, based on three main components: person and situation factors (i.e. personality traits, attitudes, beliefs), present internal states (i.e. cognition, arousal, affect), and outcomes of appraisal and decision-making processes (Anderson & Bushman, 2002; DeWall & Anderson, 2011). Neurocognitive processes play an important role in the second and third component.

Given the potential detrimental physical and mental consequences of aggression, large scale investments have been made in developing interventions for preventing or reducing aggressive behaviour. Unfortunately, these - usually cognitive behavioural therapy (CBT) interventions - appear to be only moderately successful (Jones et al., 2011; Scotto Rosato et al., 2012; Smeets et al., 2014), especially in violent offender populations (Genovés, Morales, & Sánchez-Meca, 2006; Herman, Rotunda, Williamson, & Vodanovich, 2014). The cognitive focus of these CBT programs, such as the Aggression Replacement Therapy (Goldstein, Glick, & Gibbs, 1998), is unilaterally geared towards investigating and influencing the explicit content of cognition (i.e. “what were you thinking, when...?”). Expanding this emphasis to more implicit, neuropsychological processes of cognition (i.e. “how were you thinking, when...?”) may provide information on missing links for therapy and enable therapists to improve the effect sizes of current CBT programs (Fishbein et al., 2009). For example, the effectiveness of CBT-programs depends on participants’ ability to (a) be aware of and responsive to potential future negative consequences, (b) inhibit inappropriate behavioural responses, and (c) understand and act on the benefits of deliberate and cautious decision making (Fishbein et al., 2009). Although much is known about classic neurocognitive processes in this regard, mostly concerning
executive functions (Hofmann, Schmeichel, & Baddeley, 2012; Morgan & Lilienfeld, 2000; Ogilvie et al., 2011), knowledge on the role of more implicit social cognitive aspects is still relatively lacking. This concept will be explained further later on in this introduction.

**Psychopathy**

Although the concept of psychopathy in itself is not a main focus of this thesis, it is closely related to aggression and also potentially relevant because this is a relatively intensely investigated construct with regard to its neurocognitive, neuroanatomical and neurofunctional components (Cummings, 2015; Hirschtritt et al., 2018; Wahlund & Kristiansson, 2009). Therefore, psychopathy is regarded as a separate secondary variable of interest in this thesis. In contrast to aggression, which is primarily a behavioural concept (i.e. a state), the term psychopathy refers to relatively robust personality characteristics (i.e. traits), such as egocentricity, impulsivity, lack of remorse and empathy, shallow affect, manipulativeness and persistent violation of social norms (Hare, Hart, & Harpur, 1991). In general, the psychopathy concept as defined by Hare (1980) is applied most. According to this definition, psychopathy is a constellation of affective, interpersonal and behavioural characteristics, that can be subdivided into two main factors. The first factor is portrayed by so-called callous, unemotional traits, while the second represents the antisocial, impulsive lifestyle. Many people with psychopathy tend to show aggressive and violent behaviour (Wahlund & Kristiansson, 2009). This mainly involves instrumental aggression in order to benefit from it (which is often related to the first factor), but also impulsive/reactive aggression as a result from disinhibition (which is in turn related to the second factor of the psychopathy construct) (Coccaro, Lee, & McCloskey, 2014; Reidy, Shelley-Tremblay, & Lilienfeld, 2011).

**Social Cognition**

Social cognition is regarded as a neurocognitive domain just as for example attention, memory and executive functions. The importance of social cognition for functioning in everyday life is acknowledged more and more, even though its assessment and treatment still is a novelty in neuropsychological clinical practice (Arioli, Crespi, & Canessa, 2018). The concept is mostly defined as: “all mental processes underlying social interactions” (Brothers, 1990), and can be roughly subdivided into three stages of information processing: (1) perception of social stimuli, (2) interpretation of social signals, experiencing and evaluating emotions, and (3) socially appropriate, reactive behaviour and decision making. Deficits in social cognition can be found in some neurological conditions, such as frontotemporal dementia or Parkinson’s disease (Christidi, Migliaccio, Santamaria-Garcia, Santangelo, & Trojsi, 2018), and also in
some psychiatric disorders, such as autism spectrum disorders or schizophrenia (Arioli et al., 2018; Kimoto, Makinodan, & Kishimoto, 2018; Zwick, 2017). One could easily reason that social cognitive skills could be impaired in antisocial, aggressive and/or psychopathic groups as well, since these groups are often characterised by troublesome social behaviour. Indeed, in different studies impairments were found in aspects of social cognition in antisocial groups (Blair, 2010; Marsh & Blair, 2008; Raine & Yang, 2006). In the present thesis the focus lies on two rather implicit, or so to say ‘intuitive’, social cognitive processes that have not yet been fully investigated in relation to aggression: facial emotion perception (important in the first stage of social information processing) and risky decision making (which is part of the third stage of social information processing).

Facial emotion perception
With respect to the perception of facial emotional expressions, individuals with antisocial traits appear to show deficits in recognizing some of the six basic emotional facial expressions, mostly in recognition of fear (Hoaken, Allaby, & Earle, 2007; Jusyte, Mayer, Künzel, Hautzinger, & Schönenberg, 2015; Marsh & Blair, 2008), which has been linked to callous unemotional traits in psychopathy in particular (Jusyte et al., 2015). In addition to such insensitivity in the perception of aforementioned expressions, it has been shown that people with proneness to aggression tend to be overly sensitive to recognise anger (Schönenberg, Louis, Mayer, & Jusyte, 2013), and, in fact, are biased to interpret neutral or ambiguous facial expressions as hostile, or angry (Schönenberg & Jusyte, 2014). This general tendency of aggressive individuals to perceive hostile intent in others, even when there are no clear outward signals of such hostile motives, is referred to as the ‘hostile attribution bias’ (Nasby, Hayden, & DePaulo, 1980). This hostile attribution bias has shown to be present in situations where intentions of story-characters needed to be rated by children with reactive aggressive dispositions (Dodge, 1980; Dodge, Price, Bachorowski, & Newman, 1990), but also in adult offenders’ perception of neutral or ambiguous facial expressions (Schönenberg & Jusyte, 2014; Smeijers et al., 2017) or body posture (Kret & de Gelder, 2013). In the latter situations, the bias is better described as a hostile interpretation bias, since it is constituted already in the first stage of social information processing, in contrast to the hostile attribution bias in the second stage.

Both types of bias are generally regarded as a reinforcing factor for aggression (Dodge, 2006), because misinterpretation or -attribution of ambiguous emotional cues can result in inappropriate social responses, potentially leading to an escalation into conflict, and finally acting aggressively or violently (Dodge & Schwartz, 1997).
In this scenario, the chance that the other person reacts aggressively in return increases as well, thereby creating a self-enhancing mechanism. If there indeed is a causal - or at least reinforcing – relation between perceiving facial expressions as hostile and aggression, the next step would be to investigate if reducing this emotion perception bias would also lessen aggressive tendencies. In fact, two studies have shown that this bias could be restored with respect to facial emotion interpretation using a computerised training (Penton-Voak et al., 2013; Schönenberg et al., 2014), and that this resulted in a decrease in anger and aggression as well (Penton-Voak et al., 2013).

**Risky decision making**

The process of decision making in general encompasses different cognitive skills, necessary for adaptive behaviour in a range of contexts. Successful decision making depends on an individual’s ability to identify a set of choice possibilities, integrate different variables like probability and value, and predict the consequences of a choice for our goals, whilst inhibiting the temptation of immediate rewarding outcomes (Coutlee & Huettel, 2012). Deliberate decision making is closely linked to executive functioning (EF), since it requires classic executive functions such as the ability to think and plan ahead, whilst overseeing potential consequences of decisions. However, some researchers suggest that decision making in daily life often is not so much a strategic, systematic, conscious process, but more an automatic and intuitive process, where emotional aspects play a substantial role (Naqvi, Shiv, & Bechara, 2006). The first, classic aspects of EF, are also referred to as ‘cool EF’ given the absence of emotional involvedness in the process, while the second, affect-driven processes are referred to as ‘hot EF’ (Zelazo & Carlson, 2012). It has been argued that the latter intuitive decision making processes may be more useful for decision making in daily social situations, where ambiguous and unforeseen circumstances often make it hard to follow a specific strategic plan (Bechara, Damasio, Damasio, & Anderson, 1994; Toplak, Sorge, Benoit, West, & Stanovich, 2010). Because of the uncertain outcome in such situations, these decision always involve a certain degree of risk. Hence, the term ‘risky decision making’ is often applied to describe these hot EF (Reyna & Huettel, 2014). Since risky decision making is typically relevant for social decision making, this process is not only linked to EF, but is also regarded as an essential element during the last stage of social cognitive information processing.

In the second part of the present study the focus is shifted towards the role of risky decision-making in relation to aggression. Knowledge about this relationship is essential, because impairments in decision making may have implications for the
capacity of aggressive individuals to actually apply learned knowledge and skills in daily situations. On risky decision making tasks, it has indeed been shown that some antisocial groups don’t show the tendency for 'safe intuition' that healthy controls do, and, in contrast, favour risky choice options. These studies on risky decision-making have mainly focussed on the relationship with either general delinquent behavioural aspects (Ames, Grenard, & Stacy, 2013; Bouchard, Brown, & Nadeau, 2012), or specific antisocial and psychopathic personality traits (Baker et al., 2013; Bettison, Mahmut, & Stevenson, 2013; Blair, Colledge, & Mitchell, 2001; Dean et al., 2013; Zimak, Suhr, & Bolinger, 2014), and have revealed inconsistent results. Focussing on behavioural symptoms, of which aggression is the most pronounced, instead of on syndromes, could help to verify some of these inconsistencies. This knowledge is also needed to enhance treatment programs which in turn can contribute to reducing recidivism rates of violent offenders. Although both hot and cool EF appear to be relatively stable constructs, there is also growing evidence that they are surprisingly malleable, with positive implications for interventions and prevention (Zelazo & Carlson, 2012). This concerns, for example, working memory training (Karbach & Verhaeghen, 2014; Melby-Lervag & Hulme, 2013; Olesen, Westerberg, & Klingberg, 2004) or other process-based EF training procedures that target general capacities such as inhibition or mental flexibility (Karbach & Verhaeghen, 2014), multi-domain training (e.g. video-game training) (Dovis, van der Oord, Wiers, & Prins, 2015; Gray, Robertson, Manches, & Rajendran, 2018; Schmiedek, Lövdén, & Lindenberger, 2010) and strategy-based training (Alfonso, Caracuel, Delgado-Pastor, & Verdejo-García, 2011), as well as indirect approaches such as intense physical exercise (Moreau, Kirk, & Waldie, 2017) and music training (Moreno et al., 2011). It must be said, however, that positive training effects do not always remain over time, the generalizability to other cognitive domains is limited and transfer to daily life still remains unclear (Karbach & Verhaeghen, 2014; Melby-Lervag & Hulme, 2013). If EF can be improved through such procedures in offender populations remains to be investigated.

**Neuroanatomical and neurochemical substrates**

Like most cognitive functions, social cognition is dependent on a complex interplay between multiple parts in the brain. Nevertheless, two important sections of the prefrontal cortex, the orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC), have been specifically related to social cognition and risky decision making in particular (Arioli et al., 2018; Bechara, Damasio, & Damasio, 2000; Jonker, Jonker, Scheltens, & Scherder, 2014; Reyna & Huettel, 2014). The OFC is located across the bottom of the frontal lobes, right above the eyes, and is directly connected to the amygdala (an important structure of the limbic system, which
CHAPTER 1

is essential for emotion) through the fasciculus uncinatus tract (Olson, McCoy, Klobusicky, & Ross, 2013). The vmPFC is located in the lowest parts of the inner sides of both frontal lobes, in between of the left and right orbitofrontal regions. Individuals with damage to the OFC and vmPFC show impaired decision making in daily life and on measures of hot EF, even though they tend to perform normal on classic neuropsychological tasks (Jonker et al., 2014). This can be explained by the fact that cool EF are mostly related to higher parts of the frontal lobes than hot EF, mainly the dorsolateral frontal regions (Wood & Worthington, 2017). Facial emotion perception has been related to a lot of structures, including occipital and temporal regions (including the amygdala), but all also to the OFC (Arioli et al., 2018).

When looking at aggression, some similarities arise with respect to underlying neuroanatomical substrates. Proactive aggression and psychopathy have been related mainly to decreased activity in the amygdala, resulting in lowered emotional arousal and thus less aversion driven, avoidant reaction mechanisms in response to the inflicted suffering of others (Blair, 2001; Blair, 2010). Simply put: these individuals are able to logically understand the negative impact of their behaviour on others, but are insensitive to this knowledge because it is not accompanied by an uncomfortable feeling, and thus they feel no real urge to stop their harmful behaviour. Reactive aggression, on the other hand, appears to be the result of increased amygdala activation in combination with decreased functioning of prefrontal cortical areas, including the OFC. This results in increased emotional reactivity combined with lacking inhibition of subsequent emotionally driven, impulsive behaviour (Blair, 2001; Blair & Lee, 2013; Patrick, 2008). In simple words: these individuals act violently because they are overpowered by their strong emotions (anger or sense of threat) and are unable to control themselves. It has been suggested that the problems in risky decision-making and reactive aggression both derive from this underlying orbitofrontal deficit (Bechara, Damasio, et al., 2000; Blair, 2001; Mehta & Beer, 2010). This could indicate a similar point of origin and, as a result, a shared vulnerability for both risky decision-making and reactive aggression. It is not yet clear whether an OFC deficit leads to poor decision making and to increased reactive aggression independent of each other, or if OFC disfunctioning leads to problems in cognition (such as decision making), which in turn lead to aggressive tendencies.

With regard to facial emotion perception there are also similarities when it comes to underlying neurological substrates. For example, the perception of aggressive facial expressions is associated with increased amygdala activation, and decreased activation in the vmPFC and OFC (Beaver, Lawrence, Passamonti, & Calder, 2008; Coccaro, McCloskey, Fitzgerald, & Phan, 2007), especially in individuals who show
problematic aggressive behaviour (Coccaro et al., 2007). Furthermore, in aggressive individuals, the connectivity between the amygdala and OFC during perception of angry faces is less strong than in healthy controls (Coccaro et al., 2007). These patterns are similar to those described above in reactive aggressive individuals (Blair, 2001; Blair & Lee, 2013; Patrick, 2008). This also seems to indicate a similar point of origin and, as a result, a shared vulnerability for reactive aggression when perceiving anger in facial expressions. There is some preliminary evidence that the training of facial emotional perception leads to a decline in aggressive tendencies (Penton-Voak et al., 2013), which suggests a causal relationship.

Causes of disfunctioning of orbitofrontal, ventromedial and dorsolateral prefrontal systems can be diverse. Both systems can be affected, for example, in patients who suffered from traumatic brain injury (TBI) (Wood & Worthington, 2017). Interestingly, prevalence rates of mild TBI in prisoners are striking (Farrer & Hedges, 2011), with average rates of 46.0 to 60.3% (Durand et al., 2017; Shiroma, Ferguson, & Pickelsimer, 2010) and peaks up to 94.7% (Woolhouse, McKinlay, & Grace, 2018). However, no numbers of TBI-prevalence in Dutch prisons are known.

Other factors related to the efficiency of the orbitofrontal circuit include neurochemical processes. It has been suggested, for example, that steroid hormones such as testosterone are key regulators in certain brain regions (Carré & Olmstead, 2015; Mehta & Beer, 2010), including the orbitofrontal cortex (Mehta & Beer, 2010; Op de Macks et al., 2016), the amygdala (Manuck et al., 2010; Radke et al., 2015; van Wingen, Mattern, Verkes, Buitelaar, & Fernández, 2010; Volman et al., 2016; Volman et al., 2016), and the striatum (Hermans et al., 2010; Montoya, Terburg, Bos, & van Honk, 2012; Op de Macks et al., 2011). Even more, high testosterone levels have been related to both (reactive) aggression (Carré, Campbell, Lozoya, Goetz, & Welker, 2013; Carré et al., 2017; Denson, Mehta, & Ho Tan, 2013; Mehta & Beer, 2010; Montoya et al., 2012; Popma et al., 2007) and risky decision making (Apicella, Dreber, & Mollerstrom, 2014; Mehta, Welker, Zilioli, & Carré, 2015; Op de Macks et al., 2016; Van Honk et al., 2004; Wagels et al., 2017). Inducing testosterone actually appears to result in decreased activation in the medial OFC, and increased aggressive behaviour during a decision making paradigm (Mehta & Beer, 2010). However, the testosterone-aggression-decision making relation is complex and inconsistently confirmed in earlier studies (Book, Starzyk, & Quinsey, 2001; Carré & Archer, 2018). These varying results may be due to the fact that the effect of testosterone on aggression and decision making is moderated by cortisol levels in such a way that high testosterone only increases aggression and impairs decision making in individuals with low cortisol levels, while the effect of testosterone is
often investigated without regard of cortisol levels (Van Honk, Harmon-Jones, Morgan, & Schutter, 2010). Multiple studies that used the testosterone/cortisol ratio confirm this ‘dual hormone hypothesis’, in relation to both aggression and risk taking (Denson et al., 2013; Mehta & Josephs, 2010; Mehta & Prasad, 2015; Mehta et al., 2015; Montoya et al., 2012; Popma et al., 2007; Terburg, Morgan, & van Honk, 2009), but none have investigated those altogether in one study.

**Aims and outline of this thesis**

As was described in the first part of this introduction, the general aim of the research in this thesis is to increase the knowledge of neurocognitive deficiencies that may underlie aggression and find potential targets to improve aggression treatment programs. Overseeing all of the above, important questions remain concerning the specific relationship between tendencies for risky decision making and types of aggression and psychopathic traits in offenders, including if this might be influenced by testosterone/cortisol levels and could be related to mild TBI. This is something that will be elaborated on in the first part of this thesis.

Also, it is important to further investigate if having a hostile interpretation bias is related to specific types of aggression and (violent) crimes in offenders and if this can be actively reversed through treatment. Ergo, in the second part of this thesis the focus shifts to the role of facial emotion perception in aggression and towards assessing the efficacy of an emotion perception training to reduce the hostile perception bias in a larger sample of incarcerated offenders.

More specifically the aims in the present thesis are to:

1. Systematically review the empirical literature on the potential relation between risky decision making and aggression to gain more understanding in the nature of this relation (Chapter 2).

2. Assess if prisoners perform worse than normal controls on measures of risky decision making (hot EF) and classic (cool) EF (Chapter 3).

3. Investigate if the tendency for risky decision making in prisoners is more closely related to impulsive/reactive aggression than to instrumental/proactive aggression (Chapter 3).
4. Determine if this relationship, when present, is mediated by neurochemical factors (testosterone and cortisol levels) (Chapter 3).

5. Explore if there is a relationship between risky decision making and psychopathic personality traits in a prison population (Chapter 4).

6. Assess the prevalence rates for (mild) TBI in Dutch prisoners and if TBI is related to measures of aggression and elevated tendencies for the making of risky decisions (Chapter 5).

7. Investigate if there is a hostile interpretation bias in facial emotion perception in violent offenders, and if this is related to aspects of impulsive/reactive aggression (Chapter 6).

8. Determine if a computerised training, designed to reduce hostile interpretation of facial emotion perception in ambiguous faces, is effective to restore a hostile interpretation bias in an adult prison population, and whether this leads to a subsequent decline of aggressive traits (Chapter 7).

Study design

The studies in this thesis are based on different empirical designs. First of all, a systematic literature review was performed. Second, an explorative cross-sectional study was conducted with male prisoners and a normal control group. This assessment consisted of a large neuropsychological test battery, measures of aggression, personality and saliva sampling. Both between subject and within subject analyses were conducted. Finally, a randomized controlled trial was performed with a double blind design.