

Resisting, desisting or persisting?

Neural correlates of antisocial and psychopathic pathways in early adulthood

Ilse H. van de Groep



The research described in this thesis was supported by the Ammodo Science Award 2017 for Social Sciences awarded to E.A. Crone.

Lay-out + typography: Wendy-Bour van Telgen
Cover design + chapter illustrations: Ilse van de Groep
Printing: Ipskamp Printing te Enschede
ISBN: 978-94-6473-035-7

© Ilse van de Groep, 2023

All rights are reserved. No part of this book may be reproduced, distributed, stored in a retrieval system, or transmitted in any form or by any means, without prior written permission of the author.

Resisting, Desisting or Persisting?
Neural correlates of antisocial and psychopathic pathways in early adulthood

Weerstaan, stoppen of volharden?
Neurale mechanismen van antisociale en psychopathische ontwikkelingspaden in de
vroege volwassenheid

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 6 april 2023 om 13.00 uur

door

Ilse Hetty van de Groep
geboren te Utrecht.

Promotiecommissie:

Promotoren:

prof.dr. E.A.M. Crone
prof.dr. A. Popma

Overige leden:

prof.dr. I.H.A. Franken
prof.dr. L.G.M.T. Keijsers
dr. A.C.K. van Duijvenvoorde

Copromotoren:

dr. M.G.N. Bos
prof.dr. L.M.C. Nauta-Jansen

TABLE OF CONTENTS

Chapter 1	General Introduction	7
Chapter 2	Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others	35
Chapter 3	Through the looking glass: the neural basis of self-concept in young adults with antisocial trajectories	75
Chapter 4	Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback	119
Chapter 5	Psychopathic traits are associated with reward-related striatal activity patterns during reinforcement learning for self and others	167
Chapter 6	General Discussion	195
Addendum		
	Nederlandstalige samenvatting (Summary in Dutch)	218
	Dankwoord (Acknowledgements in Dutch)	226
	Curriculum Vitae	231
	Publications and PhD Activities	232



General Introduction

Part of this chapter was preprinted in modified form as: van de Groep, I. H., Bos, M., Popma, A., Crone, E., & Nauta-Jansen, L. (2022). A neurocognitive model of early onset persistent and desistant antisocial behavior in early adulthood. <https://doi.org/10.31234/osf.io/sw73u>

Scope

Antisocial behavior, like aggression or non-compliance, violates the rights and wellbeing of others (Frick et al., 2018), and is costly for victims, perpetrators, and society at large (Moffitt, 2018; Romeo et al., 2006). A wealth of studies has shown that antisocial behavior peaks in adolescence (ages 10-18), and subsequently drops off during early adulthood (also known as young adulthood; ages 18-26), a pattern known as the age-crime curve (Jennings and Reingle, 2012, see figure 1). Accordingly, young adulthood has long been recognized as a turning point for antisocial behavior, including aggression (Moffitt, 2018; Nguyen & Loughran, 2018). For most antisocial youth, early adulthood is a period where they desist from antisocial and aggressive behavior (Bersani and Doherty, 2018; Moffitt, 1993; 2018) and find their place in our society, as a result of both psychosocial and neurobiological maturation (Cauffman & Steinberg, 2000), which is associated with rising well-being (Arnett et al., 2011). However, a small group with an early onset of antisocial behavior persist in their antisocial behavior and show life-course persistent antisocial behavior (Moffitt, 1993, 2018), resulting in societal dysfunctioning, and a wide range of problems later in life, including poor mental and physical health, substance abuse and involvement in crime (Brazil et al., 2018; Shaw & Gross, 2008).

One important factor that has been proposed to differentiate between persistent and desistant antisocial developmental trajectories concerns differential patterns of brain development (Moffitt, 1993, 2018). In line with this idea, recent studies have shown that life-course persistent antisocial behavior – but not desistant antisocial behavior - was characterized by differential cortical and subcortical brain structure (Carlisi et al., 2020, 2021). However, possible neural functional mechanisms that help explain *how* and *why* differences between persistent and desistant antisocial developmental trajectories arise remain largely elusive, especially in young adulthood. Therefore, the current dissertation examines several candidate neurocognitive mechanisms in young adults with persistent and desistant trajectories of antisocial behaviour, and typically developing young adults. In this thesis, I focus specifically on developmentally salient knowledge, skills, and self-understanding that young adults need to find their place in society: evaluating, acting upon, monitoring, and learning about and for themselves and others.

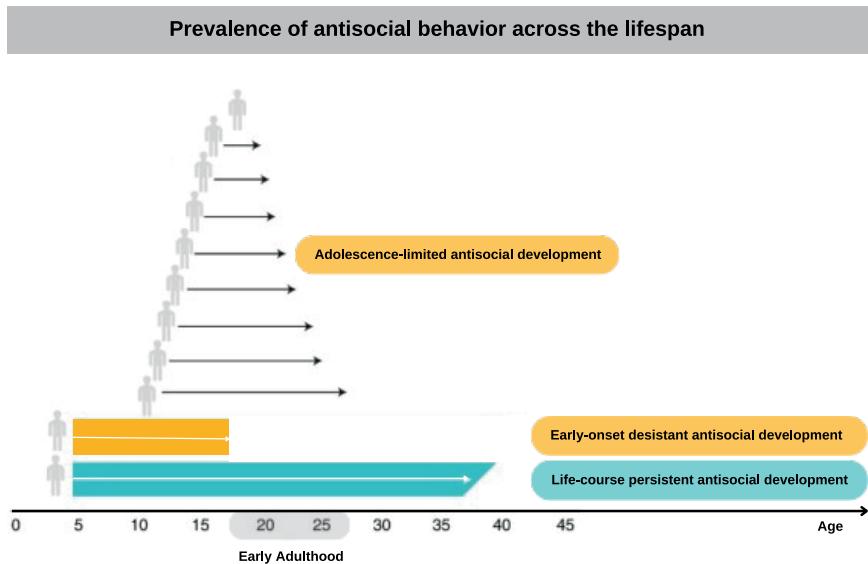


Figure 1. Schematic depiction of the prevalence of antisocial behavior across the lifespan. According to the developmental taxonomy of antisocial behavior (Moffitt, 1993, 2018), the age crime curve reveals two distinct developmental trajectories of antisocial behavior: Life-course persistent antisocial behavior and adolescence-limited antisocial behavior. Individuals with life-course persistent antisocial behavior start showing such behavior early in life, and continue throughout development. In contrast, individuals with adolescence-limited antisocial behavior only show age-normative antisocial behavior, and desist in early adulthood. Later research also identified other groups, like individuals who abstain from showing antisocial behavior throughout development, and childhood-limited/early onset desistant antisocial individuals, who also display antisocial behavior at an early age, but desist from antisocial behavior in adolescence and early adulthood. Adapted from Moffitt (2018).

1.2 The neurodevelopment of childhood-onset persistent and desistant trajectories of antisocial behavior in late adolescence

One of the most influential and well-cited theories on persistent and desistant antisocial development is the taxonomy proposed by Moffitt (1993, 2018), which outlined two distinct developmental trajectories of antisocial behavior. The first, known as life-course persistent antisocial behavior, emerges early in life, and is characterized by early neurodevelopmental problems, which are repeatedly amplified and reinforced by a high-risk social environment throughout development (Moffitt, 2018). The second trajectory, known as adolescence-limited antisocial behavior, emerges in adolescence, and is thought to result from a developmentally normative desire to feel more mature, but limited access to the desired privileges and responsibilities associated with adulthood (Monahan et al., 2013; Moffitt, 2018). To fill this gap, adolescents may show antisocial

behavior to impress and form bonds with their peers, and to gain autonomy from their parents (Moffitt, 2018). Longitudinal research generally supports the existence of these two antisocial developmental pathways, but over the past few years, additional developmental trajectories have also been identified (Piquero, 2008). For instance, in addition to these antisocial groups, there are also individuals who abstain from antisocial behavior throughout development (Moffitt, 2018; Monahan et al., 2013), either because they have early access to more mature social roles, or they fail to fit in with their peers and are therefore less likely to be pressured into delinquency (Moffitt, 2018). Finally, various studies have identified individuals with a childhood-onset of antisocial behavior who show conduct problems early in life, but desist in adolescence and early adulthood (Monahan et al., 2013; Bevilacqua et al., 2018; Moffitt, 2008), now known as the 'childhood-limited' taxon (Moffitt, 2008). Individuals with a childhood-limited desistant trajectory often develop into healthy, well-functioning young adults (Odgers et al., 2007, 2008), although a subset might develop internalizing and substance abuse problems in early adulthood (Moffitt et al., 2002; Odgers et al., 2007, 2008).

Previous studies have started to uncover functional neural mechanisms that underlie differences in behavioral functioning between individuals with early onset and persistent or desistant antisocial development, and abstainers, in adolescence. For instance, Cohn and colleagues (2013; 2015, 2016a) examined a subsample ($n = 150$) of a childhood arrestee cohort ($n = 364$) called RESIST at age 17 using a technique to non-invasively map and measure functional neural activity, called functional magnetic resonance imaging (fMRI). In the RESIST childhood arrestee cohort, which also forms the basis of the current dissertation, all adolescents were first arrested by the police before the age of twelve, in both rural and urban regions, for a range of offences including theft, vandalism, and violence (Cohn et al., 2013, 2015; Cohn, et al., 2016a; Domburgh, 2009; Pape et al., 2015). Note that most research on the characteristics and mechanisms underlying the persistence of antisocial behavior, including in the RESIST study, has been conducted in males (Eme, 2020; Moffitt, 2018; but see Freitag et al., 2018), in line with the finding that males are 10-14 times more likely to develop life-course persistent antisocial behavior than females (Eme, 2020).

As expected, the employed developmental cognitive neuroscience approach revealed several mechanisms contributing to persistence in adolescence, including deficient processing of feedback cues. First, using an adapted version of the monetary incentive delay task, Cohn et al. (2015) showed that persistence, but not desistance of antisocial behavior was associated with hypoactivity in the ventral striatum during reward processing, and with amygdala hyperreactivity during loss processing. These alterations raise the question whether such aberrant incentive processing underlie difficulties in (reinforcement) learning and decision making. Second, Cohn et al. (2013, 2016a) also revealed that some behavioral and neural patterns were not specific to persistent antisocial behavior, but instead underlie both persistent and desistant developmental groups. For

instance, both persistent and desistant antisocial behavior were characterized by neural hyperactivity during fear acquisition and extinction, compared to healthy controls. Together, these findings suggest that both persistent and desistant antisocial behavior may be associated with altered salience processing of negative (feedback) cues (Cohn et al., 2015) – although this pattern may be more prevalent (across different contexts) for persistent antisocial behavior.

Similar to an early onset of antisocial behavior, psychopathic personality traits (see Box 1) have also been associated with a more severe and persistent development of antisocial behavior (Frick et al., 2014; Cohn et al., 2015; 2016a; Stickle et al., 2009). Interestingly, separate dimensions of psychopathic traits, such as Callous-Unemotional, Grandiose-Manipulative and Impulsive-Irresponsible traits (Andershed et al., 2002) have been shown to differentially influence neurocognitive functioning in persistent and desistant antisocial development. For instance, in prior functional neuroimaging work in adolescents by Cohn and colleagues, neural hyperactivity during fear learning and extinction was positively associated with Impulsive-Irresponsible traits in persistent and desistant antisocial groups, while Callous-Unemotional traits were negatively associated with neural responses during reward processing and fear acquisition and extinction (Cohn et al., 2013, 2015, 2016a). In line with these findings, structural imaging studies also revealed distinct structural patterns in similar limbic brain areas (Insula, Amygdala) for adolescents characterized by Callous-Unemotional traits (Cohn, Viding, et al., 2016). Moreover, individuals with Callous-Unemotional traits and Grandiose-Manipulative traits also showed unique patterns of cortical white-matter structural connectivity, which likely influences the functionality of fronto-limbic brain areas (Pape et al., 2015). Notably, high levels of psychopathic traits do not only affect brain structure and functioning, but also affect associated (mal)adaptive behavior, where higher levels may be beneficial in some types of situations (e.g. fast-life strategy; quickly changing or hostile social contexts, Doerfler et al., 2021), and more negative in others (e.g. situations that require multi-tasking or processing of multiple information streams) (Baskin-Sommers & Newman, 2013; Doerfler et al., 2021). Combined with the empirical observation that early adulthood is a salient period for the development and influence of personality traits and personality disorders on aggression (Ostrov & Houston, 2008) - like psychopathy and Antisocial Personality Disorder (ASPD) – these findings suggest that considering psychopathic traits may offer additional clues and specificity about mechanisms underlying the persistence of antisocial behaviors into early adulthood, and help explain the observed heterogeneity in antisocial developmental trajectories (Brazil et al., 2018).

Even though important prior work focused on antisocial development in adolescence, relatively few studies have focused on the transition from adolescence into early adulthood, despite evidence that this developmental period is perhaps equally important to understand the (dis)continuation of antisocial behavior throughout the life course (Monahan et al., 2013; Taber-Thomas & Pérez-Edgar, 2015). This long-term developmental

approach, despite being practically challenging, is one of the only ways to examine persistent versus desistant trajectories in development (Moffitt, 1993; 2018). Before considering potential neurocognitive differences between these trajectories, and potential impairments or adaptations that are associated with these developmental pathways, I will now turn to describing typical neurodevelopmental changes in early adulthood.

Box 1 – Understanding the multidimensional development of psychopathic traits

Psychopathy is a personality construct characterized by difficulties in affective, interpersonal and behavioral domains (Carré et al., 2013; Nentjes et al., 2022). Although there are many different conceptualizations and operationalizations of psychopathy in the literature, most concur on the notion that psychopathy is multidimensional in nature (Lilienfeld, 2018). Research on the development of psychopathic traits often employs a conceptualization consisting of three dimensions: Callous-Unemotional traits (characterized by a lack of empathy, remorse and shallow affect), Impulsive-Irresponsible traits (characterized by impulsivity and irresponsibility) and Grandiose-Manipulative traits (characterized by lying, manipulating and a grandiose sense of self-worth) (Andershed, 2002), which has received ample empirical support (Lee & Kim, 2020). At the same time, many studies using this conceptualization have focused predominantly on either the total, global construct of psychopathy, or only one of the dimensions of psychopathy (Callous-Unemotional traits) (Lilienfeld, 2018). However, it has become increasingly clear that the different dimensions of psychopathy are often associated with different behavioral and neurological outcomes and mechanisms and thus potentially provide information above and beyond other dimensions (Carré et al., 2013; Lilienfeld, 2018). Coincidentally, in some situations, psychopathic dimensions may also interactively influence such outcomes and mechanisms (Lilienfeld, 2018) and show uniformity across dimensions (Garofalo et al., 2018). Although psychopathic traits are assumed to show a relatively stable pattern throughout development, recent developmental studies emphasize the potential for change in these traits in some individuals, and different expressions of these traits in changing social contexts when adolescents transition into early adulthood (Lee & Kim, 2020; Nentjes et al., 2022).

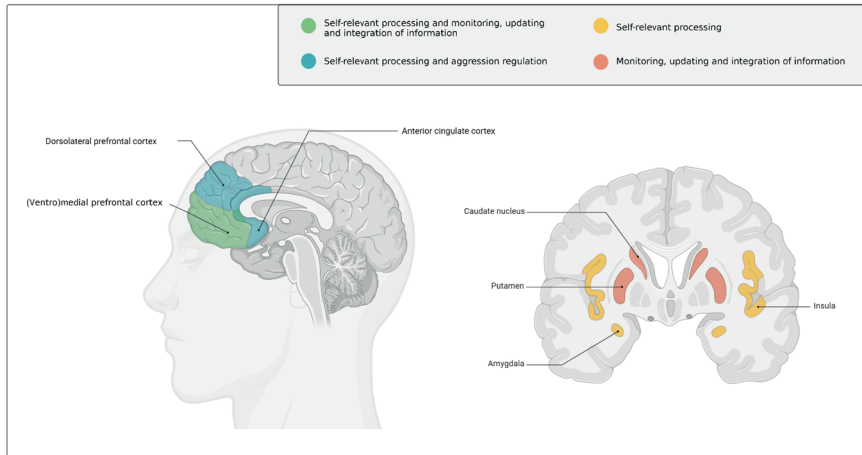


Figure 2. Schematic overview of the brain regions that undergo marked neurodevelopmental changes during the transition from adolescence into early adulthood and have been implicated in self-relevant and goal-related social information processing, behavioral regulation, monitoring and learning in early adulthood.

1.3 Neurodevelopmental changes in Early Adulthood

Early adulthood is a life period that is characterized by changes in social interactions, due to rapidly changing environments, social relationships, social roles, and social norms (Arnett, 2000; 2007; Sussman & Arnett, 2014). The exact age ranges for early adulthood are dependent on contextual factors such as societal norms and historical times, but general consensus is that early adulthood encompasses approximately the age ranges 18-26 (Sawyer et al., 2018). To navigate these contextual changes and ultimately effectively function as an adult in society, young adults need to develop knowledge, skills, and self-understanding to balance between environmental constraints and their own goals (Arnett 2000, Arnett et al. 2007). Developmentally distinctive to other developmental periods across the lifespan is that early adults focus more on themselves, individualistic goals (Nelson, 2021), and the development of various (social) identities that fit different social roles and contexts (Arnett, 2000). During early adulthood, individuals also grow more confident that they can achieve their goals (i.e., period of opportunities and possibilities, Arnett, 2000). These social and psychological changes during early adulthood are supported by ongoing brain development (Herting et al., 2018; Mills et al., 2016; Taber-Thomas & Pérez-Edgar, 2015; Tamnes et al., 2017). More specifically, early adulthood is marked by extensive structural changes in association cortices (i.e., areas that integrate and associate information from various sensory modalities) and frontolimbic systems, such as the prefrontal cortex (PFC) (in particular in the ventrolateral and dorsolateral PFC (vlPFC and dlPFC), as well as the ventromedial PFC extending into the anterior cingulate

cortex (ACC)), and subcortical limbic structures like the (ventral) striatum and the anterior insula (AI) (Gogtay et al., 2004; Taber-Thomas and Pérez-Edgar, 2015; see Figure 2).

Not only in structural, but also in functional brain development, early adulthood marks a period of transition. During the transition from adolescence into early adulthood, individuals show increased prefrontal functioning and enhanced connectivity between the PFC and subcortical structures (Taber-Thomas & Perez-Edgar, 2015). These neurobiological changes are thought to facilitate several processes that may be important for the development of and desistance from antisocial behavior, including integration of multiple streams of cortical and subcortical (social) information processing, appropriate behavioral selection, behavioral regulation (e.g. balancing between approach and avoidance tendencies) and future-oriented behavior (e.g. increased attention towards and opportunities to learn from negative (long-term) consequences) (Flechsengar et al., 2022; Monahan et al., 2013; Taber-Thomas & Pérez-Edgar, 2015). Together, these social, psychological, and neurobiological changes in early adulthood support adaptation to constantly changing environments (Flechsengar et al., 2022), while balancing these situational characteristics with self-relevant goals and motivations.

1.4. Social information processing model of aggressive and antisocial behavior

Studying the role of social cognitive processes contributing to aggressive and persistent antisocial behavior is important to understand the origin and maintenance of such behavior (Choe et al., 2015), as well as to improve assessment, and ultimately improve prevention and intervention efforts (Klein Tunte et al., 2019). The Social Information Processing model (SIP, Crick & Dodge, 1994; Dodge & Crick, 1990, see Figure 3) provides a theoretical framework to understand reactive aggression in social contexts in early adulthood. According to SIP, how aggressively someone responds to social cues depends on both the social cues themselves, as well as on how they are interpreted and processed. Over the past few decades, the study of social information processing has greatly informed the understanding of both aggression in specific social contexts, and the development and maintenance of persistent antisocial behavior (Stickle et al., 2009), especially in childhood and early adolescence (Bowen et al., 2016; Klein Tunte et al., 2019). According to the SIP model, social information is interpreted and responded to in six steps (Crick & Dodge, 1990; 1994): In the first step, individuals *attend to and encode information* from the current social situation, using internal (physiological and emotional states) and external cues (environmental stimuli). Second, individuals *give meaning to the information*, using contextual cues and organized knowledge from memory. During this step, individuals interpret the intent of others (e.g. hostile attribution bias (Klein Tunte et al., 2019; Smeijers et al., 2019)), and consider what the situation might mean for self and others (Galán et al., 2022). In the third step, individuals *set a specific goal* for the current situation. Fourth, individuals *identify potential responses* for the current situation (either accessed from

long-term memory based on previous stimulus-response associations, or newly generated). During the fifth step, individuals *evaluate whether the response* chosen in step 4 *is indeed the best to select and implement* (e.g. by considering the anticipated consequences of behavior). Finally, *the selected response is performed and monitored*. Together, the first three steps concern social cognitions about input, while the final three steps are social cognitions about output (Bowen et al., 2016; Galán et al., 2022). Note that the order of SIP is not sequential, but cyclic with multiple feedback loops and often simultaneous processes (Galán et al., 2022). Although most research on SIP has been done in children and adolescents (Klein Tunte et al., 2019; Bowen et al., 2016), this model may provide an important framework to understand the social-cognitive processes and the neural basis of social information processing and reactive aggression in early adulthood (Vitaro et al., 2006). Different parts of the model fit well with specific developmentally salient characteristics and changes of early adulthood and may provide important starting points for research into persistent and desistant antisocial trajectories.

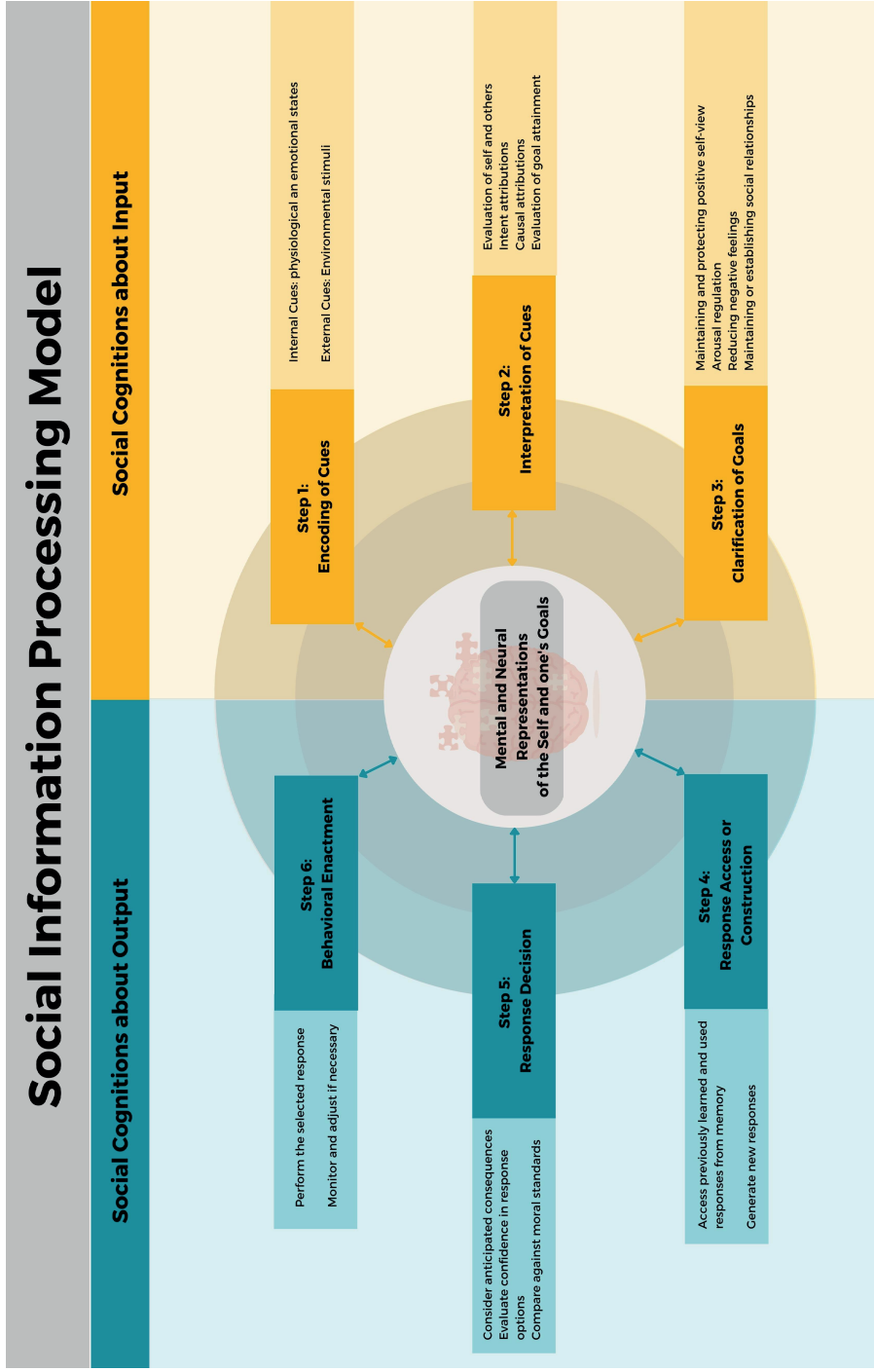


Figure 3. Adapted schematic depiction of Crick & Dodge's (1994) Social Information Processing Model.

Prior studies examining neural responses to social stimuli have mainly focused on separate – and usually the first – information processing steps (e.g. encoding and interpretation; Dodge et al., 2006). For example, many studies have examined neural responses to emotional faces and social threats in the amygdala, a deep subcortical emotion processing region (Adolphs, 2010; Bertsch et al., 2020; Bickart et al., 2014). These studies showed that the amygdala is involved in recognizing emotions from faces, and more generally contributes to the processing of emotional and socially relevant, salient information (see e.g. Sergerie et al., 2008; Adolphs, 2010 for reviews and meta-analyses). Notably, most research to date has focused on the perception and attribution (of intent) to others (i.e., the evaluation of others), largely ignoring the hypothesized role for internal, self-related processing (Crick & Dodge, 1990; 1994; Huesmann, 1998; but see Blankenstein et al., 2021), at least with regard to self-evaluation (i.e., SIP step 2). Moreover, only recently have researchers examined the interaction between multiple social information processing steps, including between the first three steps (encoding, interpretation, goal selection) and the final three steps (identification, selection, monitoring) (Dodge et al., 2022). In line with this notion, researchers have called for an extension of the SIP model to more closely integrate emotion and cognition across all SIP steps (Lemerise & Arsenio, 2000; Smeijers et al., 2020). Emotional processes (e.g., emotional experiences, emotional understanding, emotion recognition, and emotion regulation, Smeijers et al., 2020) are thought to reduce information processing demands and support goal-directed behavior (Lemerise & Arsenio, 2002; Smeijers et al., 2020). Indeed, impaired emotional processes do not only affect the first two – but rather all SIP steps, probably also interactively, and have been shown to contribute to (the development of) aggressive behavior (Lemerise & Arsenio, 2002; Smeijers, et al., 2020). For instance, emotionality may influence which information is accessible and therefore used to interpret social contexts and limit the evaluation and use of alternative response options (Smeijers et al., 2020), which may bias some individuals towards aggressive and antisocial responses. Interestingly, a recent systematic review suggests that considering distinct emotional features, like valence, can provide even more specificity to understand potential impairments in SIP (Smeijers et al., 2020).

This introduction chapter continues with background information for three candidate neurocognitive SIP mechanisms of interest that may differ between young adults with different antisocial trajectories: self-concept appraisal (**Chapter 2, 3**), social feedback processing and subsequent aggression regulation (**Chapter 2, 4**), and social reinforcement learning (**Chapter 5**).

1.5 Evaluating the self: is there a common neural substrate underlying internal and external self-evaluation?

Navigating complex and changing social contexts requires young adults to reflect upon themselves – on their thoughts, traits, and actions, and how these differ from those of others (Crone & Fuligni, 2020; Rodman et al., 2017). At the same time, they need to monitor

relevant social feedback from others that may influence such self-evaluations. Previous studies have implicated that the Medial Prefrontal Cortex (mPFC) is involved in both these internal and external self-evaluations (Denny et al., 2012; Flagan & Beer, 2013; Lieberman et al., 2019), potentially signaling an intertwined function in monitoring the self relative to others (Crone & Fuligni, 2020). However, this hypothesis has not been tested within the same individuals. Moreover, recent studies have raised concerns about the reliability and reproducibility of task-based fMRI studies (Elliot et al., 2020), highlighting that it is important to assess the robustness of findings, for research and clinical practice (Elliott et al., 2020; Kragel et al., 2021). Therefore, the first aim of the current dissertation is to replicate previous findings on the neural correlates of internal and external self-evaluations, and test to what extent their neural substrates overlap.

1.6 Evaluating the self, by oneself

In most individuals, their view of themselves (i.e., their self-concept) is generally positive and well-structured, with a more positive self-concept in some domains than in others (Crone et al., 2022). A positive self-concept appraisal has been associated with increased well-being (Rodman et al., 2017), and its development is largely shaped by previous social experiences and development of cognitive abilities (Crone et al., 2022; Harter, 2012). Cognitive and social development also enable self-appraisals to become more increasingly multifaceted and complex in early adulthood, which support both goal attainment and adaptation to changing social contexts (Crone et al., 2022).

Given the complexity of self-related thoughts and their sensitivity to biases, studying self-appraisals is inherently complex. To address this complexity, neuroscience studies have started to examine self-appraisal using trait-evaluation paradigms, by asking participants whether positive and negative trait statements fit with them, in different domains (van der Cruijssen et al., 2018; van der Cruijssen et al., 2017). Using these paradigms, previous research has repeatedly shown activity during self-appraisals in cortical midline areas (Denny et al., 2012; Northoff et al., 2006). In particular, the anterior mPFC often shows increased activity for self-related activation (Denny et al., 2012; Northoff et al., 2006). More detailed understanding of self-appraisals can also be acquired by examining self-appraisals in different contexts. Studies using this approach show that early adults differentiate in self-appraisal across different life domains (e.g. social, physical, academic domain (van der Cruijssen et al., 2018)), and these evaluations in diverging domains have been associated with different neural underpinnings (see van der Cruijssen et al., 2018). However, an important open question, which will be answered in the current dissertation, concerns whether early adults with a history of aggressive behavior and varying levels of psychopathic traits form similar or diverging neural responses to self-evaluation across domains and contexts, especially given that one's self-concept is shaped by prior - and in the case of persistent antisocial behavior often negative - social experiences (Harter, 2012; Veenstra et al., 2009).

1.7 Evaluation of social feedback about the self and subsequent aggression regulation

How individuals appraise themselves is not solely determined by internal, self-related processes. Rather, how people evaluate themselves is also influenced by external, social feedback from others (Crone et al., 2022). Social feedback can signal positive or negative information about oneself or one's behavior. Receiving social feedback is important for learning, imitation, and adaptation of social behavior (Zhang et al., 2022), and the pursuit and attainment of goals (Fishbach & Finkelstein, 2012). Social feedback can take many forms that differ depending on the number of people involved and the content or type of feedback (Rappaport & Barch, 2020). Most studies on the neural correlates of social feedback processing have focused on neural activation underlying social exclusion (i.e., negative social feedback by multiple individuals at the same time), which is often assessed using the Cyberball task (Williams & Jarvis, 2006); for meta-analyses see Cacioppo et al., 2013; Vijayakumar et al., 2017), by contrasting this to neural activation to social inclusion. Generally, negative social feedback triggers anger and frustration, which in turn leads to reactive aggression (Chester & DeWall, 2014; Dodge & Pettit, 2003). Studies using the Cyberball paradigm typically reveal that social exclusion evokes increased activity in cortical midline areas like the ACC, mPFC and AI (see Cacioppo et al., 2013; Vijayakumar et al., 2017).

The responses in the ACC, mPFC and AI have been interpreted as reflecting 'social pain' (Eisenberger & Lieberman, 2004) as they respond strongly to social rejection, or 'salience' as the ACC is also active when social feedback does not match prior expectations (Somerville et al., 2006). However, recent findings indicate that activity in these areas may not be specific to negative social feedback (i.e., not valence-specific), but instead reflect increased social salience of all stimuli that elicit affective responses, including positive feedback (Dalgleish et al., 2017; Perini et al., 2018). Moreover, the Cyberball inclusion condition is often considered to be a neutral, rather than a positive and rewarding condition (Rappaport & Barch, 2020) meaning that it includes only one salient event (Perini et al., 2018). Therefore, studies have introduced experimental tasks that do not only distinguish between negative and neutral feedback, but also positive feedback (Guyer et al., 2008; Kujawa et al., 2014; Silk et al., 2012). The Social Network Aggression Task (SNAT, Achterberg et al., 2016) is a task in which participants receive positive, negative, and neutral feedback from their peers, and subsequently get the opportunity to show or regulate aggressive behavior towards the sources of social feedback by sending a (not so) loud noise blast. Several studies in young adults employing the SNAT show that both positive and negative feedback elicit activity in the ACC, AI and (v)mPFC (Achterberg et al., 2016) compared to neutral feedback. Taken together, in typically developing young adults, receiving social feedback from others results in activity in a network of 'salience' brain regions, including the ACC and AI, that may signal importance of the events. However, young adults with a persistent antisocial development may not show this developmentally

normative pattern, given persistence has previously been associated with aberrant processing of feedback cues (Cohn et al., 2015). More specifically, from a SIP perspective, young adults with prior antisocial experiences may interpret negative and neutral as more hostile and indicative of rejection (Crick & Dodge, 1994; Dodge 2003; Brennan et al., 2018), and hence more salient and self-relevant, which might be reflected in neural hypersensitivity to cues signaling potential social rejection (Baskin-Sommers & Newman, 2013; Blair, 2013).

One way to protect our self-image after receiving salient social feedback information is by retaliation (Chester et al., 2018), which requires a combination of the input and output steps of the SIP model. Throughout development, people show various compensatory behaviors to maintain positive and coherent self-reviews, like blaming negative feedback on external sources, devaluation of feedback sources (Chester et al., 2018; DeWall et al., 2009), or retaliatory and aggressive behavior (Achterberg et al., 2016). Especially in the context of reactive aggression, such retaliatory behavior is thought to result from poor cognitive or behavioral control (Bertsch et al., 2020). Neuroimaging studies have identified several (lateral) fronto-parietal regions that are implicated in cognitive control of aggressive responses, including the dlPFC, vlPFC, ACC, and AI (Bertsch et al., 2020; Brockett et al., 2020; Crew et al., 2021; Grahn et al., 2008; Reidy et al., 2011; van Heukelum et al., 2021). In typically developing young adults, and throughout development, negative social feedback typically elicits more aggression than neutral and positive feedback (Achterberg et al., 2016, 2017, 2018, 2020; Dobbelaar et al., 2021, 2022), and stronger activity in the dlPFC has been associated with less reactive aggression following negative social feedback (Achterberg et al., 2016). If young adults with prior antisocial experiences interpret feedback as more hostile and indicative of rejection (Crick & Dodge, 1994; Dodge 2003; Brennan et al., 2018), they may show more subsequent aggression in contexts that may be interpreted as signaling social rejection (Blair, 2013; Baskin-Sommers & Newman, 2013). Therefore, an additional aim of the current dissertation is to disentangle whether young adults with different developmental trajectories of antisocial behavior and varying levels of psychopathic traits show differences in social feedback processing and subsequent aggression regulation.

1.8 Learning how behavior might benefit or harm self and others

Although people sometimes learn or change their behavior by receiving social, external feedback, it most often results from internal monitoring of whether actions result in the desired end states (Moskowitz, 2012). From a SIP perspective, learning such action-outcome associations forms the basis of potential (dominant) behavioral responses that will be selected from memory in the future, and helps young adults to evaluate whether they should select and implement a response, based on the anticipated consequences of their behavior, which both support the goal they wish to attain (i.e., SIP steps 3-6).

Learning for self and others has previously been examined using probabilistic, or reinforcement learning tasks. While performing these tasks, individuals are required to make a series of choices, where each option probabilistically results in positive, negative or neutral outcomes (Nussenbaum & Hartley, 2019). The probability of these outcomes can remain stable throughout the task, or change at specific moments (Nussenbaum & Hartley, 2019). Over the series of choices, individuals thus learn what the best option is, and when they should change their behavior (e.g., when an option is no longer rewarding). Typically developing young adults learn to repeat actions that result in positive, rewarding, and valuable outcomes from themselves, while negative outcomes like losses are often avoided (Carvalho et al., 2021). In contrast, individuals with psychopathic tendencies often show difficulties learning how to value their actions, which negatively influences their ability to make effective choices for themselves. For instance, previous studies have demonstrated that individuals with psychopathic tendencies often fail to change their choice behavior after repeated negative feedback (Finger et al., 2011; Murray et al., 2022; Oba et al., 2019). Moreover, even after their choices resulted in rewarding outcomes, they have been shown to change their behavior (Blair et al., 2013). Thus, individuals with psychopathic tendencies show difficulties developing an effective long-term learning strategy for themselves – both for negative outcomes (losses) and positive outcomes (rewards). However, despite evidence that individuals with psychopathic tendencies often act in disregard of others (Viding & McCrory, 2019), it remains unclear whether similar learning difficulties arise if young adults have to consider outcomes for self and others simultaneously (Monahan et al., 2013). In previous studies in typically developing (young) adults that focused on learning in social contexts, higher psychopathic traits were associated with reduced sensitivity for the outcomes of others (Cutler et al., 2021; O'Connell et al., 2021). However, it remains unclear whether and how these findings translate to more high-risk samples, like those including young adults with an early onset of antisocial behavior who are at higher risk of persistent antisocial development and other negative life-outcomes (Cohn et al., 2015; Moffitt, 2018; Murray et al., 2022; Odgers et al., 2008; Pauli & Lockwood, 2022).

A further way to examine the mechanistic underpinnings of reinforcement learning is by examining neural responses to learning signals. Prior studies using probabilistic learning tasks have shown repeatedly that the striatum is involved in tracking social reinforcement learning signals for self and others while individuals receive outcomes (Lockwood, 2016; Westhoff et al., 2021), particularly when receiving rewards (Dugré et al., 2018; Oldham et al., 2018), while losses are associated more closely with the dorsal striatum (Dugré et al., 2018; Murray et al., 2022). Therefore, a final aim of this dissertation is to examine social reinforcement learning and its neural correlates in young adults with varying levels of psychopathic traits.

1.9 Aim and outline of this thesis

The main aim of the current thesis is to examine (1) how young adults – with and without developmental histories of antisocial behavior - evaluate, act upon, monitor and learn *about* and *for* themselves and others, and (2) the neural underpinnings of these neurocognitive processes and behaviors.

In the first part of this thesis, I test whether there is a common neural substrate that underlies evaluation of the self from an internal (self-appraisal) and external (social feedback) point of view in typically developing young adults (**chapter 2**). To this end, typically developing young adults (n = 40) performed two tasks: (1) the self-concept fMRI task (van der Crujisen et al., 2018), in which participants rated the applicability of positive and negative traits in different domains (prosocial, physical) and (2) the Social Network Aggression Task (SNAT) (Achterberg et al., 2016) in which participants received positive, negative and neutral peer feedback and subsequently retaliated or regulated their behavior by sending a noise blast towards the peers. Using the same tasks allowed me to study to what extent the neural underpinnings of these internal and external self-evaluative functions overlap within individuals. Moreover, this study allowed me to replicate previous behavioral and neural findings obtained using the same tasks (Achterberg et al., 2016; van der Crujisen et al., 2018), allowing me to assess the robustness and reproducibility of these prior findings.

In the second part of this thesis, I focus on a twofold aim to better understand which functional and neurobiological mechanisms are associated with antisocial behavior and psychopathic traits in early adulthood. First, I aim to investigate potential differences in neurocognitive social information processing mechanisms between young adults with a desistant and persistent antisocial developmental trajectory, and without a history of antisocial behavior. Second, I examine how varying levels of psychopathic traits, and separate psychopathic trait sub-dimensions influence neurocognitive social information processing in early adulthood. For these two aims (**chapters 3-5**), I focus on three candidate neurocognitive mechanisms of interest: (1) self-concept appraisal (**chapter 2, 3**), (2) social feedback processing and subsequent aggression regulation (**chapter 2, 4**), and (3) social reinforcement learning (**chapter 5**). Some of these mechanisms refer to internal social cognitive input processes, like self-concept appraisal (**chapter 3**), and external social cognitive input process, like social feedback processing (**chapter 4**). Moreover, regarding social cognitive output processes, or the interaction between input and output processes, I consider how people learn action-outcome associations for self and others simultaneously (**chapter 5**) and how they (fail to) regulate aggressive responses to social feedback (**chapter 4**). For each of these processes, to integrate emotion and cognition, I will focus on positively, negatively and/or neutrally valenced information. Moreover, I will study the social information processing mechanisms in different domains and contexts. Ultimately, increased characterization of these different processes and behaviors will clarify whether potential social-cognitive, behavioral and neural deficits

related to the development of antisocial behavior are general or context- and valence specific, which is important to understand the exact mechanisms underlying antisocial behavior and identify possible avenues for intervention efforts (Nelson et al., 2008; Smeijers et al., 2020).

In **chapter 3**, I focus on the question whether early adults' different developmental trajectories of antisocial behavior and varying levels of psychopathic traits form similar or diverging neural responses to self-evaluation across domains and contexts. To answer this question, I examine the neural basis of self-concept appraisal in the RESIST cohort sample (van Domburgh, 2009; Cohn et al., 2015) who are currently in young adulthood, and the sample of typically developing young adults described in **chapter 2**. Participants with a persistent or desistant history of antisocial behavior ($n = 54$) and typically developing young adults ($n = 40$), all with varying levels of psychopathic traits (ranging from low to severe, Cohn et al., 2015), performed a self-concept fMRI task by rating which positive and negative traits in different domains (prosocial, physical) best characterized themselves. This study, reported in **chapter 3**, thus allowed me to examine potential domain- and valence specific differences in internal social cognition input processes associated with antisocial and psychopathic tendencies, on a behavioral and neural level.

In **chapter 4**, I set out to establish whether young adults with different developmental trajectories of antisocial behavior and varying levels of psychopathic traits process social feedback information in a different manner and differ in their (aggressive) behavioral and neural responses after receiving feedback. Given that neuroscience studies may provide a direct marker of salience, in this study, young adults with a persistent, desistant ($n = 54$) or no history of antisocial behavior ($n = 40$) performed the SNAT fMRI task described in **chapter 2**. Prior research showed that individuals give the loudest noise blasts following negative feedback, less following neutral feedback and least following positive feedback. The SNAT paradigm allowed me to test whether young adults with persistent antisocial behavior show a similar pattern, or tend to show more aggression, which might be either more pronounced after social rejection, or present regardless of the feedback valence – and associated with less activity in the dlPFC. Taken together, the study described in **chapter 4** allowed me to examine (1) potential valence-specific neural differences in external social cognition input processes and (2) potential valence-specific behavioral and neural differences in the interplay between input and output steps of the SIP model associated with antisocial and psychopathic tendencies.

In **chapter 5**, I report another fMRI study in young adults with varying levels of psychopathic traits ($n = 53$) who performed a social probabilistic learning task. In this study, I examined (1) how they learn which of their actions result in positive or negative incentives, for themselves and an unknown other simultaneously, and (2) the striatal neural underpinnings of this social reinforcement learning. Participants were required to learn in different contexts with different outcome contingencies. In particular, in some task conditions, outcomes would remain stable for participants themselves, but could

result in positive or negative outcomes for the other. Conversely, in other task conditions, outcomes would remain stable for the other, but could result in positive or negative outcomes for themselves. Thus, this study allowed me to examine context- and valence-specific effects in reinforcement learning, which supports the interplay between output steps of the SIP model associated with psychopathic tendencies.

The research outlined in this introduction is discussed in detail in the remaining chapters of this dissertation – both in the empirical chapters (**chapters 2-5**), and in a general discussion chapter (**chapter 6**). In this final chapter, I summarize the results of the empirical chapters 2-5, and discuss the main findings and their theoretical and practical implications, as well directions for future research.

REFERENCES

1. Achterberg, M., Duijvenvoorde, A. C. K. van, IJzendoorn, M. H. van, Bakermans-Kranenburg, M. J., & Crone, E. A. (2020). Longitudinal changes in DLPFC activation during childhood are related to decreased aggression following social rejection. *Proceedings of the National Academy of Sciences*, *117*(15), 8602–8610. <https://doi.org/10.1073/pnas.1915124117>
2. Achterberg, M., van Duijvenvoorde, A.C.K., van der Meulen, M., & Crone, E. A. (2018). Heritability of aggression following social evaluation in middle childhood: An fMRI study. *Human Brain Mapping*, *39*(7), 2828–2841. <https://doi.org/10.1002/hbm.24043>
3. 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. <https://doi.org/10.1093/scan/nsv154>
4. Achterberg, M., van Duijvenvoorde, A. C. K., van der Meulen, M., Euser, S., Bakermans-Kranenburg, M. J., & Crone, E. A. (2017). The neural and behavioral correlates of social evaluation in childhood. *Developmental Cognitive Neuroscience*, *24*, 107–117. <https://doi.org/10.1016/j.dcn.2017.02.007>
5. Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, *1191*(1), 42–61. <https://doi.org/10.1111/j.1749-6632.2010.05445.x>
6. Andershed, H. A., Kerr, M., Stattin, H., & Levander, S. (2002). *Psychopathic traits in non-referred youths: A new assessment tool* (pp. 131–158). Elsevier. <http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-6748>
7. Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*(5), 469. <https://doi.org/10.1037/0003-066X.55.5.469>
8. Arnett, J. J. (2007). Emerging Adulthood: What Is It, and What Is It Good For? *Child Development Perspectives*, *1*(2), 68–73. <https://doi.org/10.1111/j.1750-8606.2007.00016.x>
9. Arnett, J. J., Kloep, M., Hendry, L. B., & Tanner, J. L. (2011). *Debating emerging adulthood: Stage or process?* Oxford University Press.
10. Baskin-Sommers, A., & Newman, J. P. (2013). Differentiating the cognition-emotion interactions that characterize psychopathy versus externalizing. In *Handbook of cognition and emotion* (pp. 501–520).
11. Bersani, B. E., & Doherty, E. E. (2018). Desistance from Offending in the Twenty-First Century. *Annual Review of Criminology*, *1*(1), 311–334. <https://doi.org/10.1146/annurev-criminol-032317-092112>
12. Bertsch, K., Florange, J., & Herpertz, S. C. (2020). Understanding Brain Mechanisms of Reactive Aggression. *Current Psychiatry Reports*, *22*(12), 81. <https://doi.org/10.1007/s11920-020-01208-6>
13. Bevilacqua, L., Hale, D., Barker, E. D., & Viner, R. (2018). Conduct problems trajectories and psychosocial outcomes: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *27*(10), 1239–1260. <https://doi.org/10.1007/s00787-017-1053-4>
14. Bickart, K. C., Dickerson, B. C., & Feldman Barrett, L. (2014). The amygdala as a hub in brain networks that support social life. *Neuropsychologia*, *63*, 235–248. <https://doi.org/10.1016/j.neuropsychologia.2014.08.013>
15. Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience*, *14*(11), 786–799. <https://doi.org/10.1038/nrn3577>
16. Blankenstein, N. E., de Rooij, M., van Ginkel, J., Wilderjans, T. F., de Ruigh, E. L., Oldenhof, H. C., Zijlmans, J., Jambroes, T., Platje, E., & de Vries-Bouw, M. (2021). Neurobiological correlates of antisociality

- across adolescence and young adulthood: A multi-sample, multi-method study. *Psychological Medicine*, 1–16. <https://doi.org/10.1017/S0033291721003457>
17. Bowen, K. N., Roberts, J. J., & Kocian, E. J. (2016). Decision making of inmates: Testing social information processing concepts using vignettes. *Applied Psychology in Criminal Justice*, 12(1), 1–17.
 18. Brazil, I. A., van Dongen, J. D. M., Maes, J. H. R., Mars, R. B., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews*, 91, 259–277. <https://doi.org/10.1016/j.neubiorev.2016.10.010>
 19. Brennan, G. M., Crowley, M. J., Wu, J., Mayes, L. C., & Baskin-Sommers, A. R. (2018). Neural processing of social exclusion in individuals with psychopathic traits: Links to anger and aggression. *Psychiatry Research*, 268, 263–271. <https://doi.org/10.1016/j.psychres.2018.07.024>
 20. Brockett, A. T., Tennyson, S. S., deBettencourt, C. A., Gaye, F., & Roesch, M. R. (2020). Anterior cingulate cortex is necessary for adaptation of action plans. *Proceedings of the National Academy of Sciences*, 117(11), 6196–6204. <https://doi.org/10.1073/pnas.1919303117>
 21. Cacioppo, S., Frum, C., Asp, E., Weiss, R. M., Lewis, J. W., & Cacioppo, J. T. (2013). A quantitative meta-analysis of functional imaging studies of social rejection. *Scientific Reports*, 3(1), 1–3. <https://doi.org/10.1038/srep02027>
 22. Carlisi, C. O., Moffitt, T. E., Knodt, A. R., Harrington, H., Ireland, D., Melzer, T. R., Poulton, R., Ramrakha, S., Caspi, A., Hariri, A. R., & Viding, E. (2020). Associations between life-course-persistent antisocial behaviour and brain structure in a population-representative longitudinal birth cohort. *The Lancet Psychiatry*, 7(3), 245–253. [https://doi.org/10.1016/S2215-0366\(20\)30002-X](https://doi.org/10.1016/S2215-0366(20)30002-X)
 23. Carlisi, C. O., Moffitt, T. E., Knodt, A. R., Harrington, H., Langevin, S., Ireland, D., Melzer, T. R., Poulton, R., Ramrakha, S., Caspi, A., Hariri, A. R., & Viding, E. (2021). Association of subcortical gray-matter volumes with life-course-persistent antisocial behavior in a population-representative longitudinal birth cohort. *Development and Psychopathology*, 1–11. <https://doi.org/10.1017/S0954579421000377>
 24. 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. <https://doi.org/10.1080/17470919.2012.703623>
 25. Cauffman, E., & Steinberg, L. (2000). (Im) maturity of judgment in adolescence: Why adolescents may be less culpable than adults. *Behavioral Sciences & the Law*, 18(6), 741–760. <https://doi.org/10.1002/bsl.416>
 26. Chester, D. S., & DeWall, C. N. (2014). Prefrontal Recruitment During Social Rejection Predicts Greater Subsequent Self-Regulatory Imbalance and Impairment: Neural and Longitudinal Evidence. *NeuroImage*, 101, 485–493. <https://doi.org/10.1016/j.neuroimage.2014.07.054>
 27. Chester, D. S., Lynam, D. R., Milich, R., & DeWall, C. N. (2018). Neural mechanisms of the rejection–aggression link. *Social Cognitive and Affective Neuroscience*, 13(5), 501–512. <https://doi.org/10/gdmnjj>
 28. Choe, D. E., Shaw, D. S., & Forbes, E. E. (2015). Maladaptive social information processing in childhood predicts young men’s atypical amygdala reactivity to threat. *Journal of Child Psychology and Psychiatry*, 56(5), 549–557. <https://doi.org/10.1111/jcpp.12316>
 29. Cohn, M. D., Popma, A., van den Brink, W., Pape, L. E., Kindt, M., van Domburgh, L., Doreleijers, T. a. H., & Veltman, D. J. (2013). Fear conditioning, persistence of disruptive behavior and psychopathic traits: An fMRI study. *Translational Psychiatry*, 3(10), e319–e319. <https://doi.org/10.1038/tp.2013.89>
 30. Cohn, M. D., van Lith, K., Kindt, M., Pape, L. E., Doreleijers, T. A. H., van den Brink, W., Veltman, D. J., & Popma, A. (2016). Fear extinction, persistent disruptive behavior and psychopathic traits: FMRI in

- late adolescence. *Social Cognitive and Affective Neuroscience*, 11(7), 1027–1035. <https://doi.org/10.1093/scan/nsv067>
31. Cohn, M. D., Veltman, D. J., Pape, L. E., van Lith, K., Vermeiren, R. R. J. M., van den Brink, W., Doreleijers, T. A. H., & Popma, A. (2015). Incentive Processing in Persistent Disruptive Behavior and Psychopathic Traits: A Functional Magnetic Resonance Imaging Study in Adolescents. *Biological Psychiatry*, 78(9), 615–624. <https://doi.org/10.1016/j.biopsych.2014.08.017>
 32. Cohn, M. D., Viding, E., McCrory, E., Pape, L., van den Brink, W., Doreleijers, T. A. H., Veltman, D. J., & Popma, A. (2016). Regional grey matter volume and concentration in at-risk adolescents: Untangling associations with callous-unemotional traits and conduct disorder symptoms. *Psychiatry Research: Neuroimaging*, 254, 180–187. <https://doi.org/10.1016/j.psychresns.2016.07.003>
 33. Crew, L. A., Covington, H. E., & Hyman, J. M. (2021). Aggression: How the anterior cingulate cortex helps to ensure a fair fight. *Current Biology*, 31(11), R716–R718. <https://doi.org/10.1016/j.cub.2021.04.024>
 34. Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115(1), 74–101. <https://doi.org/10.1037/0033-2909.115.1.74>
 35. Crone, E. A., & Fuligni, A. J. (2020). Self and others in adolescence. *Annual Review of Psychology*, 71, 447–469. <https://doi.org/10/ggctq9>
 36. Crone, E. A., Green, K. H., van de Groep, I. H., & van der Crujisen, R. (2022). A Neurocognitive Model of Self-Concept Development in Adolescence. *Annual Review of Developmental Psychology*, 4. <https://doi.org/10.1146/annurev-devpsych-120920-023842>
 37. Cutler, J., Wittmann, M. K., Abdurahman, A., Hargitai, L. D., Drew, D., Husain, M., & Lockwood, P. L. (2021). Ageing is associated with disrupted reinforcement learning whilst learning to help others is preserved. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-24576-w>
 38. Dalgleish, T., Walsh, N. D., Mobbs, D., Schweizer, S., van Harmelen, A.-L., Dunn, B., Dunn, V., Goodyer, I., & Stretton, J. (2017). Social pain and social gain in the adolescent brain: A common neural circuitry underlying both positive and negative social evaluation. *Scientific Reports*, 7(1), 42010. <https://doi.org/10.1038/srep42010>
 39. Denny, B.T., Kober, H., Wager, T.D., & Ochsner, K.N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_00233
 40. DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *Journal of Personality and Social Psychology*, 96(1), 45. <https://doi.org/10.1037/a0013196>
 41. Dobbelaar, S., Achterberg, M., van Drunen, L., van IJzendoorn, M. H., & Crone, E. A. (2022). Development of social feedback processing and responses in childhood: An fMRI test-replication design in two age cohorts. *Social Cognitive and Affective Neuroscience*. <https://doi.org/10.1093/scan/nsac039>
 42. Dobbelaar, S., Van Duijvenvoorde, A. C., Achterberg, M., van der Meulen, M., & Crone, E. A. (2021). A bi-dimensional taxonomy of social responsivity in middle childhood: Prosociality and reactive aggression predict externalizing behavior over time. *Frontiers in Psychology*, 11, 586633. <https://doi.org/10.3389/fpsyg.2020.586633>

43. Dodge, K. A., Bai, Y., Godwin, J., Lansford, J. E., Bates, J. E., Pettit, G. S., & Jones, D. (2022). A defensive mindset: A pattern of social information processing that develops early and predicts life course outcomes. *Child Development, 93*(4), e357–e378. <https://doi.org/10.1111/cdev.13751>
44. Dodge, K. A., Coie, J. D., & Lynam, D. (2006). *Aggression and antisocial behavior in youth*.
45. Dodge, K. A., & Crick, N. R. (1990). Social Information-Processing Bases of Aggressive Behavior in Children. *Personality and Social Psychology Bulletin, 16*(1), 8–22. <https://doi.org/10.1177/0146167290161002>
46. Dodge, K. A., & Pettit, G. S. (2003). A Biopsychosocial Model of the Development of Chronic Conduct Problems in Adolescence. *Developmental Psychology, 39*(2), 349–371. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2755613/>
47. Doerfler, S. M., Tajmiriyahi, M., Ickes, W., & Jonason, P. K. (2021). The self-concepts of people with Dark Triad traits tend to be weaker, less clearly defined, and more state-related. *Personality and Individual Differences, 180*, 110977. <https://doi.org/10/gjzxrn>
48. Dugré, J. R., Dumais, A., Bitar, N., & Potvin, S. (2018). Loss anticipation and outcome during the *Monetary Incentive Delay Task*: A neuroimaging systematic review and meta-analysis. *PeerJ, 6*, e4749. <https://doi.org/10.7717/peerj.4749>
49. Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: A common neural alarm system for physical and social pain. *Trends in Cognitive Sciences, 8*(7), 294–300. <https://doi.org/10.1016/j.tics.2004.05.010>
50. Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science, 31*(7), 792–806. <https://doi.org/10.1177/0956797620916786>
51. Eme, R. (2020). Life course persistent antisocial behavior silver anniversary. *Aggression and Violent Behavior, 50*, 101344. <https://doi.org/10.1016/j.avb.2019.101344>
52. Finger, E. C., Marsh, A. A., Blair, K. S., Reid, M. E., Sims, C., Ng, P., Pine, D. S., & Blair, R. J. R. (2011). Disrupted reinforcement signaling in the orbitofrontal cortex and caudate in youths with conduct disorder or oppositional defiant disorder and a high level of psychopathic traits. *American Journal of Psychiatry, 168*(2), 152–162. <https://doi.org/10.1176/appi.ajp.2010.10010129>
53. Fishbach, A., & Finkelstein, S. R. (2012). How feedback influences persistence, disengagement, and change in goal pursuit. *Goal-Directed Behavior, 203–230*.
54. Flagan, T., & Beer, J. S. (2013). Three ways in which midline regions contribute to self-evaluation. *Frontiers in Human Neuroscience, 7*, 450. <https://doi.org/10.3389/fnhum.2013.00450>
55. Flechsenhar, A., Kanske, P., Krach, S., Korn, C., & Bertsch, K. (2022). The (un)learning of social functions and its significance for mental health. *Clinical Psychology Review, 102*204. <https://doi.org/10.1016/j.cpr.2022.102204>
56. Freitag, C. M., Konrad, K., Stadler, C., De Brito, S. A., Popma, A., Herpertz, S. C., Herpertz-Dahlmann, B., Neumann, I., Kieser, M., & Chiacchetti, A. G. (2018). Conduct disorder in adolescent females: Current state of research and study design of the FemNAT-CD consortium. *European Child & Adolescent Psychiatry, 27*(9), 1077–1093. <https://doi.org/10.1007/s00787-018-1172-6>
57. Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Annual research review: A developmental psychopathology approach to understanding callous-unemotional traits in children and adolescents with serious conduct problems. *Journal of Child Psychology and Psychiatry, 55*(6), 532–548. <https://doi.org/10.1111/jcpp.12152>

58. Frick, P. J., Robertson, E. L., & Clark, J. E. (2018). 6—Callous–unemotional traits. In M. M. Martel (Ed.), *Developmental Pathways to Disruptive, Impulse-Control and Conduct Disorders* (pp. 139–160). Academic Press. <https://doi.org/10.1016/B978-0-12-811323-3.00006-7>
59. Galán, C. A., Feldman, J. S., & McClaine, R. N. (2022). Using the social information processing model to understand gender differences in the manifestation and frequency of aggression. *Aggression and Violent Behavior*, 101766. <https://doi.org/10.1016/j.avb.2022.101766>
60. Garofalo, C., Neumann, C. S., & Velotti, P. (2018). Difficulties in emotion regulation and psychopathic traits in violent offenders. *Journal of Criminal Justice*, 57, 116–125. <https://doi.org/10.1016/j.jcrimjus.2018.05.013>
61. Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, 101(21), 8174–8179. <https://doi.org/10.1073/pnas.0402680101>
62. Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, 86(3), 141–155. <https://doi.org/10.1016/j.pneurobio.2008.09.004>
63. Guyer, A. E., Lau, J. Y., McClure-Tone, E. B., Parrish, J., Shiffrin, N. D., Reynolds, R. C., Chen, G., Blair, R. J. R., Leibenluft, E., & Fox, N. A. (2008). Amygdala and ventrolateral prefrontal cortex function during anticipated peer evaluation in pediatric social anxiety. *Archives of General Psychiatry*, 65(11), 1303–1312. <https://doi.org/10.1001/archpsyc.65.11.1303>
64. Harter, S. (2012). *The construction of the self: Developmental and sociocultural foundations, 2nd ed* (pp. xxiii, 440). The Guilford Press.
65. Herting, M. M., Johnson, C., Mills, K. L., Vijayakumar, N., Dennison, M., Liu, C., Goddings, A.-L., Dahl, R. E., Sowell, E. R., Whittle, S., Allen, N. B., & Tamnes, C. K. (2018). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *NeuroImage*, 172, 194–205. <https://doi.org/10.1016/j.neuroimage.2018.01.020>
66. Huesmann, L. R. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In *Human aggression* (pp. 73–109). Elsevier.
67. Jennings, W. G., & Reingle, J. M. (2012). On the number and shape of developmental/life-course violence, aggression, and delinquency trajectories: A state-of-the-art review. *Journal of Criminal Justice*, 40(6), 472–489. <https://doi.org/10.1016/j.jcrimjus.2012.07.001>
68. Klein Tunte, S., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults—A systematic review. *Aggression and Violent Behavior*, 46, 66–81. <https://doi.org/10.1016/j.avb.2019.01.009>
69. Kragel, P. A., Han, X., Kraynak, T. E., Gianaros, P. J., & Wager, T. D. (2021). Functional MRI Can Be Highly Reliable, but It Depends on What You Measure: A Commentary on Elliott et al. (2020). *Psychological Science*, 32(4), 622–626. <https://doi.org/10.1177/0956797621989730>
70. Kujawa, A., Arfer, K. B., Klein, D. N., & Proudft, G. H. (2014). Electrocortical reactivity to social feedback in youth: A pilot study of the Island Getaway task. *Developmental Cognitive Neuroscience*, 10, 140–147. <https://doi.org/10.1016/j.dcn.2014.08.008>
71. Lee, Y., & Kim, J. (2020). Psychopathic Traits Among Serious Juvenile Offenders: Developmental Pathways, Multidimensionality, and Stability. *Crime & Delinquency*, 0011128720926120. <https://doi.org/10.1177/0011128720926120>

72. Lemerise, E. A., & Arsenio, W. F. (2000). An Integrated Model of Emotion Processes and Cognition in Social Information Processing. *Child Development, 71*(1), 107–118. <https://doi.org/10.1111/1467-8624.00124>
73. Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews, 99*, 311–328. <https://doi.org/10.1016/j.neubiorev.2018.12.021>
74. Lilienfeld, S. O. (2018). The multidimensional nature of psychopathy: Five recommendations for research. *Journal of Psychopathology and Behavioral Assessment, 40*(1), 79–85. <https://doi.org/10.1007/s10862-018-9657-7>
75. Lockwood, P. L. (2016). The anatomy of empathy: Vicarious experience and disorders of social cognition. *Behavioural Brain Research, 311*, 255–266. <https://doi.org/10.1016/j.bbr.2016.05.048>
76. Mills, K. L., Goddings, A.-L., Herting, M. M., Meuwese, R., Blakemore, S.-J., Crone, E. A., Dahl, R. E., Guroğlu, B., Raznahan, A., & Sowell, E. R. (2016). Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *Neuroimage, 141*, 273–281. <https://doi.org/10.1016/j.neuroimage.2016.07.044>
77. Moffitt, T. E. (1993). *Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy*. 28.
78. Moffitt, T. E. (2018). Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour, 2*, 177. <http://s2443.www.ncbi.nlm.nih.gov/pmc/articles/PMC6157602/>
79. Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology, 14*(1), 179–207. <https://doi.org/10.1017/s0954579402001104>
80. Monahan, K. C., Steinberg, L., Cauffman, E., & Mulvey, E. P. (2013). Psychosocial (im)maturity from adolescence to early adulthood: Distinguishing between adolescence-limited and persisting antisocial behavior. *Development and Psychopathology, 25*(4pt1), 1093–1105. <https://doi.org/10.1017/S0954579413000394>
81. Moskowitz, G. B. (2012). *The representation and regulation of goals*.
82. Murray, L., Lopez-Duran, N. L., Mitchell, C., Monk, C. S., & Hyde, L. W. (2022). Antisocial behavior is associated with reduced frontoparietal activity to loss in a population-based sample of adolescents. *Psychological Medicine, 1*–9. <https://doi.org/10.1017/S0033291722000307>
83. Nelson, L. J. (2021). The Theory of Emerging Adulthood 20 Years Later: A Look at Where It Has Taken Us, What We Know Now, and Where We Need to Go. *Emerging Adulthood, 9*(3), 179–188. <https://doi.org/10.1177/2167696820950884>
84. Nentjes, L., Garofalo, C., & Kosson, D. S. (2022). 4 - Emotional functioning in psychopathy: A critical review and integration with general emotion theories. In P. B. Marques, M. Paulino, & L. Alho (Eds.), *Psychopathy and Criminal Behavior* (pp. 75–103). Academic Press. <https://doi.org/10.1016/B978-0-12-811419-3.00006-6>
85. Nguyen, H., & Loughran, T. A. (2018). On the Measurement and Identification of Turning Points in Criminology. *Annual Review of Criminology, 1*(1), 335–358. <https://doi.org/10.1146/annurev-criminol-032317-091949>
86. Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *NeuroImage, 31*(1), 440–457. <https://doi.org/10.1016/j.neuroimage.2005.12.002>

87. Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, 40, 100733. <https://doi.org/10.1016/j.dcn.2019.100733>
88. Oba, T., Katahira, K., & Ohira, H. (2019). The Effect of Reduced Learning Ability on Avoidance in Psychopathy: A Computational Approach. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02432>
89. OConnell, K., Walsh, M., Padgett, B., Connell, S., & Marsh, A. (2021). *Modeling variation in empathic sensitivity using go/no-go social reinforcement learning*. PsyArXiv. <https://doi.org/10.31234/osf.io/drcq6>
90. Odgers, C. L., Caspi, A., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Moffitt, T. E. (2007). Prediction of differential adult health burden by conduct problem subtypes in males. *Archives of General Psychiatry*, 64(4), 476–484. <https://doi.org/10.1001/archpsyc.64.4.476>
91. Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, 20(2), 673–716. <https://doi.org/10.1017/S0954579408000333>
92. Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, 39(8), 3398–3418. <https://doi.org/10.1002/hbm.24184>
93. Ostrov, J. M., & Houston, R. J. (2008). The utility of forms and functions of aggression in emerging adulthood: Association with personality disorder symptomatology. *Journal of Youth and Adolescence*, 37(9), 1147–1158. <https://doi.org/10.1007/s10964-008-9289-4>
94. Pape, L. E., Cohn, M. D., Caan, M. W. A., van Wingen, G., van den Brink, W., Veltman, D. J., & Popma, A. (2015). Psychopathic traits in adolescents are associated with higher structural connectivity. *Psychiatry Research: Neuroimaging*, 233(3), 474–480. <https://doi.org/10.1016/j.pscychresns.2015.07.023>
95. Pauli, R., & Lockwood, P. (2022). *The Computational Psychiatry of Antisocial Behaviour and Psychopathy*. PsyArXiv. <https://doi.org/10.31234/osf.io/mqbvu>
96. Perini, I., Gustafsson, P. A., Hamilton, J. P., Kämpe, R., Zetterqvist, M., & Heilig, M. (2018). The salience of self, not social pain, is encoded by dorsal anterior cingulate and insula. *Scientific Reports*, 8(1), 6165. <https://doi.org/10.1038/s41598-018-24658-8>
97. Piquero, A. R. (2008). Taking stock of developmental trajectories of criminal activity over the life course. In *The long view of crime: A synthesis of longitudinal research* (pp. 23–78). Springer.
98. Rappaport, B. I., & Barch, D. M. (2020). Brain responses to social feedback in internalizing disorders: A comprehensive review. *Neuroscience & Biobehavioral Reviews*, 118, 784–808. <https://doi.org/10.1016/j.neubiorev.2020.09.012>
99. Reidy, D. E., Shelley-Tremblay, J. F., & Lilienfeld, S. O. (2011). Psychopathy, reactive aggression, and precarious proclamations: A review of behavioral, cognitive, and biological research. *Aggression and Violent Behavior*, 16(6), 512–524. <https://doi.org/10.1016/j.avb.2011.06.002>
100. Rodman, A. M., Powers, K. E., & Somerville, L. H. (2017). Development of self-protective biases in response to social evaluative feedback. *Proceedings of the National Academy of Sciences*, 114(50), 13158–13163. <https://doi.org/10/gcqt6k>

101. Romeo, R., Knapp, M., & Scott, S. (2006). Economic cost of severe antisocial behaviour in children— And who pays it. *The British Journal of Psychiatry: The Journal of Mental Science*, *188*, 547–553. <https://doi.org/10.1192/bjp.bp.104.007625>
102. Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health*, *2*(3), 223–228. [https://doi.org/10.1016/S2352-4642\(18\)30022-1](https://doi.org/10.1016/S2352-4642(18)30022-1)
103. Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *32*(4), 811–830. <https://doi.org/10.1016/j.neubiorev.2007.12.002>
104. Shaw, D. S., & Gross, H. E. (2008). What we have learned about early childhood and the development of delinquency. In *The long view of crime: A synthesis of longitudinal research* (pp. 79–127). Springer.
105. Silk, J. S., Stroud, L. R., Siegle, G. J., Dahl, R. E., Lee, K. H., & Nelson, E. E. (2012). Peer acceptance and rejection through the eyes of youth: Pupillary, eyetracking and ecological data from the Chatroom Interact task. *Social Cognitive and Affective Neuroscience*, *7*(1), 93–105. <https://doi.org/10.1093/scan/nsr044>
106. Smeijers, D., Benbouriche, M., & Garofalo, C. (2020). The association between emotion, social information processing, and aggressive behavior: A systematic review. *European Psychologist*, *25*, 81–91. <https://doi.org/10.1027/1016-9040/a000395>
107. Smeijers, D., Bulten, E. B. H., & Brazil, I. A. (2019). The Computations of hostile biases (CHB) model: Grounding hostility biases in a unified cognitive framework. *Clinical Psychology Review*, *73*, 101775. <https://doi.org/10.1016/j.cpr.2019.101775>
108. Somerville, L. H., Heatherton, T. F., & Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature Neuroscience*, *9*(8), 1007–1008. <https://doi.org/10.1038/nn1728>
109. Stickle, T. R., Kirkpatrick, N. M., & Brush, L. N. (2009). Callous-Unemotional Traits and Social Information Processing: Multiple Risk-Factor Models for Understanding Aggressive Behavior in Antisocial Youth. *Law and Human Behavior*, *33*(6), 515–529. <https://doi.org/10.1007/s10979-008-9171-7>
110. Sussman, S., & Arnett, J. J. (2014). Emerging adulthood: Developmental period facilitative of the addictions. *Evaluation & the Health Professions*, *37*(2), 147–155. <https://doi.org/10.1177/0163278714521812>
111. Taber-Thomas, B., & Pérez-Edgar, K. (2015, November 1). *Emerging Adulthood Brain Development*. The Oxford Handbook of Emerging Adulthood. <https://doi.org/10.1093/oxfordhb/9780199795574.013.15>
112. Tamnes, C. K., Herting, M. M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., Crone, E. A., & Mills, K. L. (2017). Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(12), 3402–3412. <https://doi.org/10.1523/JNEUROSCI.3302-16.2017>
113. van der Crujisen, R., Peters, S., & Crone, E. A. (2017). Neural correlates of evaluating self and close-other in physical, academic and prosocial domains. *Brain and Cognition*, *118*, 45–53. <https://doi.org/10.1016/j.bandc.2017.07.008>
114. van der Crujisen, R., Peters, S., van der Aar, L.P.E & Crone, E.A. (2018). The neural signature of self-concept development in adolescence: The role of domain and valence distinctions. *Developmental Cognitive Neuroscience*. <https://doi.org/10.1016/j.dcn.2017.11.005>
115. van Domburgh, L. (2009). *Very Young Offenders: Characteristics of children and their environment in relation to (re)-offending*. LAP Lambert Academic Publishing.

116. van Heukelum, S., Tulva, K., Geers, F. E., van Dulm, S., Ruisch, I. H., Mill, J., Viana, J. F., Beckmann, C. F., Buitelaar, J. K., Poelmans, G., Glennon, J. C., Vogt, B. A., Havenith, M. N., & França, A. S. C. (2021). A central role for anterior cingulate cortex in the control of pathological aggression. *Current Biology*, *31*(11), 2321–2333.e5. <https://doi.org/10.1016/j.cub.2021.03.062>
117. Veenstra, R., Lindenberg, S., Verhulst, F. C., & Ormel, J. (2009). Childhood-Limited Versus Persistent Antisocial Behavior: Why Do Some Recover and Others Do Not? The TRAILS Study. *The Journal of Early Adolescence*, *29*(5), 718–742. <https://doi.org/10.1177/0272431608325501>
118. Viding, E., & McCrory, E. (2019). Towards understanding atypical social affiliation in psychopathy. *The Lancet Psychiatry*, *6*(5), 437–444. [https://doi.org/10.1016/S2215-0366\(19\)30049-5](https://doi.org/10.1016/S2215-0366(19)30049-5)
119. Vijayakumar, N., Cheng, T. W., & Pfeifer, J. H. (2017). Neural correlates of social exclusion across ages: A coordinate-based meta-analysis of functional MRI studies. *NeuroImage*, *153*, 359–368. <https://doi.org/10.1016/j.neuroimage.2017.02.050>
120. Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: A developmental perspective. *International Journal of Behavioral Development*, *30*(1), 12–19. <https://doi.org/10.1177/0165025406059968>
121. Westhoff, B., Blankenstein, N. E., Schreuders, E., Crone, E. A., & van Duijvenvoorde, A. C. K. (2021). Increased Ventromedial Prefrontal Cortex Activity in Adolescence Benefits Prosocial Reinforcement Learning. *Developmental Cognitive Neuroscience*, *52*, 101018. <https://doi.org/10.1016/j.dcn.2021.101018>
122. Williams, K. D., & Jarvis, B. (2006). Cyberball: A program for use in research on interpersonal ostracism and acceptance. *Behavior Research Methods*, *38*(1), 174–180. <https://doi.org/10.3758/BF03192765>
123. Zhang, D., Ao, X., Zheng, Z., Shen, J., Zhang, Y., & Gu, R. (2022). Modulating social feedback processing by deep TMS targeting the medial prefrontal cortex: Behavioral and electrophysiological manifestations. *NeuroImage*, *250*, 118967. <https://doi.org/10.1016/j.neuroimage.2022.118967>



Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others

This chapter is published as: van de Groep, I. H., Bos, M. G. N., Jansen, L. M. C., Achterberg, M., Popma, A., & Crone, E. A. (2021). Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others. *Neuropsychologia*, 161, 108000. <https://doi.org/10/gn267m>

ABSTRACT

Prior research has implicated the medial and lateral prefrontal cortex (PFC) in processing evaluations from the perspective of self (self-traits) and evaluations from others (peer feedback), suggesting that these areas form a neural substrate that serves an intertwined function in monitoring self in relation to others. To test this possibility, we examined neural activation overlap in medial and lateral PFC after processing self- and other-informed evaluations. Young adults (age range 18-30-yrs, $n = 40$) performed two fMRI tasks. The self-concept task involved rating whether positive and negative traits described themselves. The Social Network Aggression Task involved processing positive, neutral or negative feedback from others, with the possibility to retaliate by blasting a loud noise following feedback. The results show that rating positive self traits and receiving positive peer feedback was associated with increased activity in an overlapping region in medial PFC. There were no significant correlations on a behavioral level and medial PFC activity for self-versus-other evaluations. The study further replicated the finding from previous research showing that higher activity in dorsolateral PFC (dlPFC) when receiving negative social feedback was associated with reduced noise blast aggression. Finally, during retaliatory responses after receiving positive feedback, participants showed increased activity in the dlPFC. Together these findings suggest that medial PFC is more strongly involved in protecting positive self-views from both internal (self traits) and external (peer feedback) points of view, whereas dlPFC is more strongly involved in regulating retaliatory responses following social rejection, and actively inhibiting aggressive behavior after receiving positive peer feedback.

Keywords: self-evaluation; peer feedback; medial prefrontal cortex; dorsolateral prefrontal cortex; aggression

INTRODUCTION

Self-evaluation is a critical skill to monitor one's own thoughts, traits and actions relative to others. As such, it encompasses both the internal reflection on one's self, across multiple domains such as social, academic and physical appearance (Harter, 2012), as well as monitoring responses of others that can impact evaluation of self, including one's self-esteem (Rodman, Powers, & Somerville, 2017). The medial prefrontal cortex (PFC) has been associated with both evaluations of one's internal self-concept (i.e., internal self-evaluation, Denny, Kober, Wager, & Ochsner, 2012) as well as evaluations of self by others (external self-evaluation) (Flagan & Beer, 2013; Somerville, Heatherton, & Kelley, 2006), suggesting that it may have an intertwined function in monitoring self in relation to others (Crone & Fuligni, 2020). The first goal of the current study is to test this intertwined function by combining internal self-evaluation and external self-evaluation in one study, which allows us to determine to what extent the neural substrate of these functions overlaps within individuals.

To understand how people successfully adapt their social behavior, it is not only important to understand how people process evaluations of themselves, but also how they respond to such evaluations. For instance, people may exhibit compensatory behaviors that can help them to maintain favorable self-views, such as devaluation of feedback sources, or showing retaliatory aggressive behavior (Chester, Lynam, Milich, & DeWall, 2018). While many studies have shown evidence for the association between rejection and aggression and the neural correlates that precede such retaliatory behavior (Bertch, Florange & Herpertz, 2020), less is known about the brain areas involved in aggressive behavior following different types of social feedback (Riva, Romero Lauro, DeWall, Chester & Bushman, 2015). Therefore, the second goal of the current study was to explore the neural processes being engaged during aggression.

1.1 Internal self-evaluation

Prior studies on self-evaluation or self-concept have mainly focused on self-descriptions, such as describing oneself across multiple domains (Flagan & Beer, 2013; van der Crujisen, Peters, & Crone, 2017). Meta-analyses have demonstrated a robust role of the medial PFC, with stronger self-related activation for the more anterior parts of the medial PFC (de la Vega, Chang, Banich, Wager, & Yarkoni, 2016) and a spatial gradient suggesting that ventral medial PFC more strongly represents the "direct self" and dorsal medial PFC the "reflected self" (i.e., the perceived opinions of others about the self, Denny et al., 2012). Even though most studies demonstrated that this activation is observed for trait descriptions in general (Pfeifer & Peake, 2012; Veroude, Jolles, Croiset, & Krabbendam, 2014), there is evidence that anterior medial PFC activation is stronger for traits that are evaluated positively (van der Crujisen et al., 2017), or more applicable to self (D'Argembeau, 2013). Possibly, the anterior medial PFC is most strongly involved for positive self-traits, given that these are more

often attributed to self (D'Argembeau, 2013) and fits with our preferred or ideal self (Harter, 2012). These findings align with models that suggest that medial PFC is strongly involved in mentalizing about self from the perspective of self and others (Somerville et al., 2013), but it remains to be determined whether the same neural regions are involved in social cognition and self-processing (Lieberman, Straccia, Meyer, Du, & Tan, 2019).

1.2 External Self-evaluation

Studies that examined evaluations in response to feedback of others have reported that activity in medial PFC is strongly dependent on prior expectations of social evaluations. These studies made use of paradigms in which participants were evaluated on the basis of appearance by others, which could result in positive or negative feedback (Yoon, Somerville, & Kim, 2018). Initial studies showed that feedback that conflicts with prior expectations results in activation in the anterior cingulate cortex, part of the medial prefrontal cortex (Somerville et al., 2006), whereas feedback that aligns with expectations was associated with stronger activation in the subgenual medial PFC and ventral striatum (Gunther Moor, van Leijenhorst, Rombouts, Crone, & Van der Molen, 2010). Subsequent studies revealed that medial PFC may be particularly important to protecting self-views (Hughes & Beer, 2013), showing more activity in relation to trial-to-trial changes in self-worth following feedback (Yoon et al., 2018). As such, medial PFC may serve the mechanistic role of integrating the experience of positively biased self-processing and mentalizing about comparison of views of others about self (Korn, Prehn, Park, Walter, & Heekeren, 2012). However, as of yet, prior studies have been inconclusive about the exact mPFC location where neural activity is elicited following social peer feedback, with some studies pointing at the rostrolateral prefrontal cortex (Hughes & Beer, 2013), and others at the anterior cingulate cortex (Somerville et al., 2006) or both (Korn et al., 2012). Moreover, it remains unclear whether corresponding activity is stronger for positive (see Gunther Moor et al., 2010; Korn et al., 2012) or negative (see Hughes and Beer, 2013) social feedback of self (Lieberman et al., 2019).

1.3. Retaliation following external evaluations

A mechanism that aids in preserving positive self-views is the ability to down-grade the messenger of potentially threatening negative social feedback (Chester, Lynam, Milich, & DeWall, 2018). A recent study gave participants the opportunity to retaliate following social feedback, in a paradigm referred to as the Social Network Aggression Task (SNAT) (Achterberg, van Duijvenvoorde, Bakermans-Kranenburg, & Crone, 2016). In this paradigm, participants were presented with positive, neutral or negative feedback based on their previous acquired personal profile. After receiving the feedback, they were given the opportunity to respond with a noise blast. Both positive and negative feedback resulted in stronger activity in the anterior medial PFC in adults (Achterberg et al., 2016). In children, negative feedback specifically resulted in stronger activity in the anterior medial PFC

(Achterberg et al., 2017), and in the rostral medial PFC (Achterberg, van Duijvenvoorde, van der Meulen, Bakermans-Kranenburg, & Crone, 2018), in regions that overlap with meta-analyses of studies on self-evaluations (Denny et al., 2012). Thus, whereas prior studies suggest that in adults medial PFC plays a role in remaining positive self-views (Hughes & Beer, 2013), studies in adults showed that this same region was similarly sensitive to positive and negative social feedback (Achterberg et al., 2016), and studies in children showed that this region was more sensitive to negative social feedback (Achterberg, van Duijvenvoorde, van IJzendoorn, Bakermans-Kranenburg, & Crone, 2020). In addition, these latter studies have shown that the dorsolateral PFC (dlPFC) is involved in individual differences in aggression regulation. In particular, those participants who showed stronger activity during evaluation in the dlPFC, a region often associated with control of impulses, were less aggressive following negative feedback (Achterberg et al., 2016; 2020). Nevertheless, a need for a replication in different samples remains to examine how generalizable the results are. In addition, it remains unclear whether the dlPFC, alongside possible other brain areas, is also activated *during* retaliatory behavior, and whether feedback valence influences the strength and direction of such activity. Together, these questions will enable us to further unravel the role of the medial PFC versus dorsal PFC in relation to positive and negative evaluations.

1.4 The current study

The first goal of the current study was to examine the relation between self-evaluations in response to internal representations to self (internal self-evaluation) with self-evaluations of external representations of self (external self-evaluation). For this purpose, we examined internal self-evaluation, where we aimed to (1) test medial PFC activity in response to self-representation in general, and in response to positive versus negative traits specifically, consistent with prior research showing that positive evaluations result in strongest activation in medial PFC (van der Crujisen et al., 2017). For external self-evaluation, we aimed to (2) replicate prior research using the SNAT paradigm testing for neural responses to positive, neutral and negative social feedback (Achterberg et al., 2016), in a new participants sample, and (3) replicate the SNAT negative social feedback - retaliation relation with the dorsolateral PFC (Achterberg et al., 2016). Next, we (4) tested (a) how the neural responses of internal- and external self-evaluation overlap and are correlated with each other and (b) whether internal and external evaluations and their neural responses (in the mPFC and dlPFC) correlate with two external self-report measures (empathy and reactive / proactive aggression) that have been associated with (individual differences in) both types of evaluations and subsequent behavioral responses (Davis, 1983). For our second goal, (i.e., to examine what neural processes are being engaged during aggression) we (5) exploratively test (a) whether different types of feedback elicit different patterns of neural activation *during* retaliatory responses and (b) whether stronger activity in the dlPFC during retaliatory responses is associated with less aggression (i.e., shorter noise blast duration) following negative feedback.

2. METHOD

2.1 Participants

The study sample consisted of 40 young adult participants between the ages of 18 and 30 (23 females, 17 males; M age = 22.68 years, SD age = 3.09). Participants identified themselves as having a Dutch (90%), Surinamese (5%), different western (2.5%) and different non-western (2.5%) ethnicities. We recruited participants within a diverse range of educational levels (39% vocational, 31.7% vocation/college, 9.8% university, 19.5% other). We included both right-handed ($N = 35$), as well as left-handed participants ($N = 5$) (see supplement page 4-5). An additional six participants who expressed interest in participating failed to meet the selection criteria during a telephonic screening ($n = 2$: current medication use, $n = 1$: current psychiatric disorder, $n = 2$: recent tattoo/piercing, $n = 1$: insufficient Dutch level) and were not included in the study. All included participants had normal-to-corrected vision. Included participants completed two subscales of the Wechsler Adult Intelligence Scale-IV (Similarities and Block Design) to estimate their intelligence quotient (IQ) scores. Estimated IQ scores ranged from 85 to 132.55 ($M = 107.35$, $SD = 11.52$).

Participants were excluded from fMRI analyses in case of excessive head motion or incomplete data. Specifically, for the internal self-evaluation fMRI analysis, one participant was excluded for excessive motion ($mm > 3$), and one for corrupted fMRI data, resulting in a sample of 38 participants. For the fMRI analyses of external self-evaluations, 5 participants were excluded, 3 for movement during fMRI ($mm > 3$), one participant for corrupted fMRI data, and one participant for failure to complete the SNAT task. Hence, the final analyzed fMRI sample for the SNAT task consisted of 35 participants. Analyses on behavioral results were conducted for those participants who completed the tasks ($N = 39$ for internal self-evaluation; $N = 39$ for external self-evaluation,). For the analyses on overlapping neural correlates, the sample was $N = 34$. Descriptive statistics for each sub-sample are described in supplementary table S1.

All participants signed informed consent prior to participation. The study was approved by the local medical ethical committee board. After completing the experiment, participants were debriefed about the aim of the study and received a financial reimbursement for their participation (€75).

2.2 Materials

2.2.1 Self-Concept task

To investigate internal self-evaluation (i.e., self-concept or trait evaluations), we used an adapted, brief version of the self-concept fMRI task (van der Crujisen, Peters, van der Aar, & Crone, 2018) (see Figure 1A). During this task, participants were asked to evaluate whether trait statements from two domains (prosocial and physical) applied to them (self-condition, 40 trials) on a four-point scale, or to categorize trait statements into four categories

(prosocial, physical, academic and I don't know) (control-condition, 12 trials). For each condition, trait sentences could either have a positive or negative valence. Valence and domains were distributed evenly among trials.

Each condition (self-condition, control-condition) was completed in a separate run. Within each run, a trial started with a fixation cross (400 ms), followed by the display with (1) the trait description and (2) response options (4600 ms, see figure 1A). Upon response, the chosen response was displayed in yellow for the remaining display time. If a participant did not respond in time, a message 'too late' was displayed for 1000 ms. The order of trials and jitter timing were optimized for our design using Optseq2 (Dale, 1999), with jittered timing intervals varying between 0 and 4400ms.

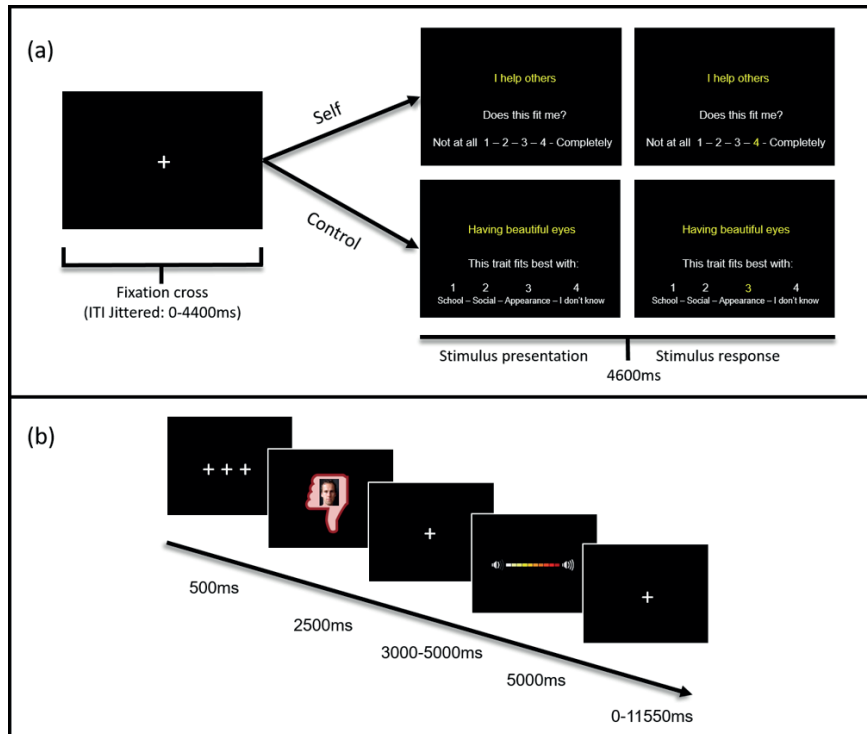


Figure 1. Task designs for the Self-concept and Social Network Aggression Task (SNAT). **(A)** During the Self-concept task, participants were asked to evaluate whether trait statements from two domains (prosocial and physical) applied to them (self-condition) on a four-point scale, or to categorize trait statements into four categories (prosocial, physical, academic and I don't know) (control-condition). **(B)** During the SNAT, participants received social feedback (positive, neutral, negative) from unknown same-aged peers. Following the social feedback, participants were asked to send hypothetical noise blasts to the same-aged peers. Noise blast duration was visualized by a volume bar.

2.2.2 Social Network Aggression Task

To investigate external self-evaluation (peer feedback) and retaliation responses, we used the Social Network Aggression Task (Achterberg et al., 2016). During this task, participants received social feedback (positive, negative, neutral) from unknown same-aged peers. Participants were told that the social feedback was given in response to a personal profile that each participant filled out prior to the experiment. Social feedback was signaled by different icons (green thumbs up for positive feedback, grey circle for neutral feedback, red thumbs down for negative feedback; see Figure 1B), which were displayed together with neutral pictures of the same-aged peers.

Following the social feedback, participants were asked to imagine that they could send noise blasts to the same-aged peers. They could determine the loudness of the noise blast with a button press, where a longer button press indicated a long noise blast duration (i.e., a loud noise). Noise blast duration was visualized by a volume bar (see Figure 1B). Prior studies have shown that imagining sending a noise blast reduces deception and is effective in eliciting aggressive responses (Achterberg et al., 2016; Konijn, Bijvank, & Bushman, 2007). The actual noise sound was only presented to participants prior to, but not during the experiment, to ensure that the button presses would not feel like punishment to the participants.

The SNAT consisted of three blocks of 20 trials (60 in total). We opted to use three blocks, instead of the division into two blocks that has been previously used by Achterberg and colleagues (2016), to facilitate participants' concentration during the task. The social feedback conditions (20 trials for each condition) were semi-randomized across these blocks, with the restriction that trials of the same condition could not be shown more than three times in a row. At the beginning of each trial, a fixation cross was presented (500ms), followed by social feedback (2500ms), another fixation screen (jittered between 3000-5000ms), and the volume bar screen (5000ms). Upon response, the volume bar fills up, with a new colored block appearing every 350ms, until the button is released, or the maximum volume is reached (at 3500ms). Subsequently, the volume bar screen remained visible in that state for the remainder of the 5000ms. Before subsequent new trials, another fixation cross was presented (jittered between 0 and 11550ms). The order of trials and jitter timing were optimized for our design using Optseq2 (Dale, 1999).

2.2.3 Cognitive and Affective Empathy Questionnaire (QCAE)

The QCAE (Reniers, Corcoran, Drake, Shryane & Völlm, 2010) is an adult self-report questionnaire that measures two aspects of trait empathy: cognitive and affective empathy. The 65 items (29 cognitive, 36 affective) were rated on a 4-point Likert scale, ranging from strongly agree to strongly disagree. The QCAE was filled in prior/after the MRI session. Both sub-scales were reliable, with Cronbach's alpha values of respectively 0.86 and 0.93. Participants had an average score of 36.8 for cognitive empathy (SD = 7.4, range 22-55), and 26.52 for affective empathy (SD = 5.41, range 17-36).

2.2.4 Reactive and Proactive aggression Questionnaire (RPQ)

The RPQ (Raine et al., 2006) is a 23-item self-report questionnaire that can be used to assess two types of aggression: reactive (11 items) and proactive aggression (12 items). Participants were asked to rate the items on a 3-point Likert scale (never, sometimes, often). The RPQ was filled in prior/after the MRI session. Both sub-scales were reliable, with Cronbach's alpha values of 0.88 and 0.96, respectively. Participants had an average score of 5.78 for reactive aggression ($SD = 3.49$, range 0-12), and 1.62 for proactive aggression ($SD = 2.12$, range 0-9).

2.3 Procedure

Prior to participation, participants were informed about the study by telephone and a digital information letter. After signing the informed consent, participants filled out several questionnaires prior to the scanning session (including the RPQ and QCAE). During the scanning session, participants first received instructions about the tasks and performed practice versions of the fMRI tasks. Since the current study was part of a larger project, several additional measures were taken during the MRI session. Hence, the session consisted of a structural MRI scan; functional scans for a reward task, the self-concept task and for the SNAT; a resting state scan and a diffusion tensor imaging scan. Subsequently, participants filled out several additional questionnaires.

2.4 Neuroimaging Methods

2.4.1 Neuroimaging Methods: MRI Data Acquisition

MRI data were acquired using a 3-T MRI scanner (Philips Achieva TX) with a standard whole-head coil. For functional MRI scans, we collected T2*-weighted gradient echo-planar images (repetition time = 2.2 sec, echo time = 30 ms, flip angle = 8°, sequential acquisition: 38 slices, voxel size = 2.75 × 2.75 × 2.75 mm, 80 × 80 matrix, field of view = 220 × 220 × 115 mm). Prior to the first functional scan of each run, five dummy scans were acquired. For the SNAT task, functional scans were acquired during three runs, which consisted of 150 dynamic scans each. For the self-concept task, the functional scans consisted of two runs with 120 and 40 volumes, respectively. Stimuli were presented on a screen that participants could see through a mirror attached to the head coil. Participants' head movements were restricted by using foam inserts at both sides of the head, to minimize free space in the head coil during acquisition. In addition to the fMRI sequences, we collected structural images for anatomical reference for a duration of 4 min 12 sec. (high-resolution 3-D T1; repetition time = 7.9 ms, echo time = 3.5 ms, flip angle = 8°, 3-D matrix size for 3-D acquisitions: 228 × 177 × 155 slices, axial slice orientation, voxel size = 1.1 × 1.1 × 1.1 mm, field of view = 250 × 196 × 170 mm).

2.4.2 Neuroimaging Methods: Preprocessing

Data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology) to allow for a direct comparison with previously published articles (Achterberg et al., 2016; van der Crujisen et al., 2018). For preprocessing, we first performed realignment (using rigid body transformation) and slice-time correction (using the middle slice as reference), followed by spatial normalization to T1 templates, and spatial smoothing with a 6-mm FWHM isotropic Gaussian kernel. Subsequently, all volumes were resampled to voxels of 3 mm³. Our templates were based on the MNI305 stereotaxic space (Cocosco, Kollokian, Kwan, & Evans, 1997). Functional images were visually checked before preprocessing and following each pre-processing step to ensure quality control.

2.4.3 Neuroimaging Methods: First level analyses

To analyze individual participants' data, we used the general linear model in SPM8. For the self-concept task, the fMRI time series were modelled as a series of zero duration events convolved with the HRF, with 'Physical-Positive', 'Physical-Negative', 'Prosocial-Positive' and 'Prosocial-Negative' being used as regressors for the self-evaluation part of the task (i.e., when participants indicate whether trait statements are applicable to themselves). Trials in the control condition were modeled as event-related events similarly as the trials in the self-condition, but without being divided into separate contrasts by valence or domain.

For the SNAT task, the fMRI time series were modeled as a series of two events convolved with the hemodynamic response function (HRF). First, we modelled the onset of social feedback with a zero duration and with separate regressors for the positive, negative and neutral peer feedback and second, we modelled the start of the noise blast for the length of the noise blast duration (i.e. length of button press) and with separate regressors for noise blast after positive, negative and neutral feedback.

For both tasks, trials on which participants failed to respond were modeled separately as a covariate of no interest and were excluded from analyses (0.625 % of trials for self-concept; 1.49 % of trials for SNAT). Moreover, six motion regressors were added as nuisance regressors, to correct for possible motion induced error not solved by realignment, and a high-pass filter was applied (120s cutoff). The least-square parameter estimates of the height of the best-fitting canonical hemodynamic response function for each condition were used in pairwise contrasts for both tasks. The pairwise comparisons resulted in participant-specific contrast images, which were subsequently submitted to second-level group analyses.

2.4.4 Neuroimaging Methods: Second level analyses

To explore neural responses to self-representation across the whole brain in the internal self-evaluation (self-concept) task, we tested four contrasts using t-tests: "self vs. control", "positive self vs. control", "negative self vs. control" and "positive self vs. negative self" (and

the reversed contrast). The goal of these analyses was to reveal regions that were more engaged during self-evaluations, or when considering positive or negative traits, respectively. Task-related responses were considered significant when they exceeded a false discovery rate (FDR) cluster correction of $p < .05$, with an initial uncorrected threshold of $p < .001$ (Woo et al., 2014). This threshold was chosen for comparability with previous work (Van der Crujisen et al, 2018).

To examine the neural responses to social feedback on a whole-brain level in the SNAT task, we performed a full factorial analysis of variance (ANOVA) with three levels (positive, negative, neutral feedback). We calculated and tested the contrast “positive vs. negative valence” (and the reversed contrast) “positive vs. neutral valence” (and the reversed contrast), negative vs. neutral valence (and the reversed contrast) to investigate brain regions that were specifically activated for social rejection or social acceptance. In addition, we calculated the conjunction “positive vs. neutral + negative vs. neutral valence” to examine brain regions that were specifically activated in response to valence. Next, to examine the association between brain activity and behavior in reaction to negative social feedback, we conducted a whole brain regression analysis at the moment negative social feedback is received (“negative vs. positive feedback”, “negative vs. neutral feedback”), using the difference in noise blast duration after negative and positive, or negative and neutral feedback, respectively as a regressor (see also Achterberg et al., 2020). All results were false discovery rate (FDR) cluster corrected at $p < .05$, with a primary voxel-wise threshold of $p < .005$ (Woo et al., 2014). This threshold was chosen for comparability with previous work (Achterberg et al., 2016).

In addition, for the SNAT, we explored whether receiving different types of feedback also results in different patterns of neural activation *during* the noise blast (i.e., the second event), using a full factorial ANOVA with three levels (positive, negative, neutral feedback). Similar to the analyses for the social evaluation (i.e., the first event), we calculated and tested the contrast “positive vs. negative valence”, “positive vs. neutral valence”, negative vs. neutral valence and the reverse contrasts. For these comparisons, task-related responses were considered significant when they exceeded a false discovery rate (FDR) cluster correction of $p < .05$, with an initial uncorrected threshold of $p < .001$ (Woo et al., 2014).

Coordinates for local maxima are reported in MNI space. Unthresholded statistical maps of all reported whole-brain analyses are available on Neurovault (Gorgolewski et al., 2015); see <https://neurovault.org/collections/FOUSRROB>.

2.4.5 Neuroimaging Methods: Region-of-Interest analyses

In order to examine whether neural activity findings for self-evaluations (van der Crujisen et al., 2017; van der Crujisen et al., 2018) and social feedback activity (i.e., external evaluations, Achterberg et al., 2016) could be replicated, we performed region-of-interest (ROI) analyses using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002). For internal self-evaluations, we performed our analyses on a predefined anatomical ROI

based on the meta-analysis (Denny et al., 2012) of the medial PFC (coordinates: $x = -6, y = 50, z = 4$). For external evaluations, we selected five ROIs: the left insula ($x = -36, y = 23, z = -2$), right insula ($x = 33, y = 20, z = -11$), ACC ($x = 0, y = 38, z = 16$) (Achterberg et al., 2016), medial PFC ($x = -9, y = 59, z = 25$) (Achterberg et al., 2018), and right dlPFC ($x = 48, y = 17, z = 37$) (Achterberg et al., 2018), based on a-priori hypotheses. For the exploratory analyses on the neural processes during noise blasts, we extracted the left dlPFC ROI from the functional whole brain analysis ($x = -43, y = 26, z = 24.6$).

Finally, to test whether we could observe associations between the overlapping medial PFC region from both task analyses, we correlated these values with each other, and with the behavioral ratings of both tasks.

2.5 Statistical analyses

Behavioral and ROI data were analyzed using R (Version 3.5.1, R Core team, 2018). Assumptions were checked for all analyses. If the assumption of sphericity was violated in behavioral and ROI repeated measures analyses, Greenhouse-Geisser (GG) corrections were applied. Based on visual inspection of boxplots, combined with the *identify_outliers* function (rstatix v0.4.0), two univariate outliers were detected for the variable noise blast, when grouped by feedback type. Hence, we winsorized these scores (Tabachnick & Fidell, 2013), which were both observed for positive feedback. Results did not change before and after winsorizing. Here, we report the winsorized results.

3. RESULTS

3.1 Behavioral results

3.1.1. Internal self-evaluation

Participants' trait applicability ratings were submitted to a 2 (domain: prosocial vs. physical) \times 2 (valence: positive vs. negative) repeated-measures ANOVA. As can be seen in Figure 2A and supplementary table S2, there was a main effect of valence, $F(1, 38) = 412.94, p < .001, \eta_p^2 = .92$, showing that participants rated their trait applicability higher on the positive items than the negative items. In addition, the analysis yielded a main effect of domain, $F(1,38) = 16.32, p < .001, \eta_p^2 = .30$, indicating that participants rated themselves higher on prosocial traits than on physical traits, independent of whether these were positive or negative traits. These effects were not qualified by a domain \times valence interaction effect, $F(1,38) = 3.28, p = .08$.

3.1.2. External self-evaluation

To examine the effects of social feedback on noise blast duration, we performed a 3-way (valence: positive vs. neutral vs. negative) repeated-measures ANOVA. As can be seen in Figure 2B and supplementary table S3, This analysis showed that participants significantly differentiated between valences, $F(2, 76) = 41.15, p < .001, \eta_p^2 = .52$ (GG-corrected), such

that noise blasts were longest following negative feedback, shorter for neutral feedback, and shortest for positive feedback (all post hoc comparisons (bonferroni-corrected), p 's < .001).

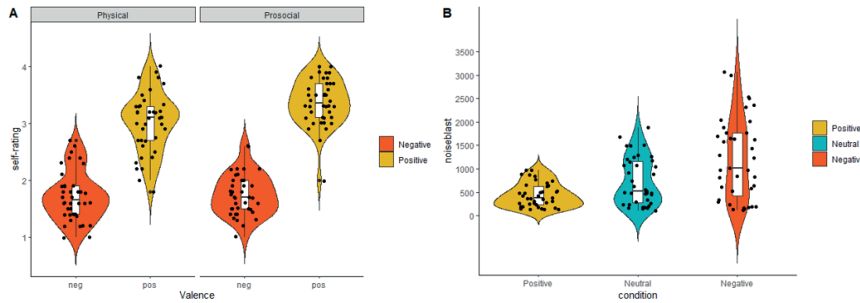


Figure 2. (A) Average applicability scores in the self-concept task. In general, positive traits were rated as more applicable than negative traits, and prosocial traits as more applicable than physical traits. (B) Average noise blast duration following social feedback in the SNAT. Noise blast duration was longest following negative feedback, shorter for neutral feedback and shortest for positive feedback.

3.1.3. Correlations among tasks and with self-report measures

Table 1 shows correlations among the task measures, and correlations with self-reported reactive and proactive aggression, and cognitive and affective empathy. We did not find any significant correlations between internal and external self-evaluation.

When comparing behavioral measures with self-report measures, we observed that the difference score between positive and negative self-evaluations was negatively correlated with cognitive empathy. There were no other correlations among task measures and self-report measures. Correlations among the self-report measures showed that cognitive empathy was negatively correlated with proactive aggression. In addition, we confirmed a positive correlation between proactive and reactive aggression.

3.2 Neuroimaging results

3.2.1 Internal self-evaluation

Whole brain - The whole brain analyses focused on the same three t-contrasts to test if activation was observed in other regions. First, the Self > Control contrast revealed no significant activations. Second, the Positive Self > Control contrast also revealed no significant activations. However, the reversed contrast (Control > Positive Self) did result in significant activation in the lateral PFC (see Table 2). Third, the Self-Positive > Self-Negative contrast showed significant activation in several regions including the medial PFC (see Figure 3A and Table 2). The reversed contrast, Self-Negative > Self-Positive did not result in significant activation, nor did the contrasts Negative Self > Control and Control > Negative Self. The unthresholded statistical map of these contrasts are available through the

NeuroVault repository under <https://neurovault.org/collections/FOUSRROB>, images 394913-394916 and 510293- 510295

ROI- To test whether neural activity for self-evaluations (van der Crujisen et al., 2017; van der Crujisen et al., 2018) could be replicated, we performed repeated measures ANOVAs for the a priori defined medial PFC ROIs based on the meta-analysis by Denny et al. (2012). The first repeated-measures ANOVA for task (self versus control) revealed no significant differences in medial PFC activity, $F(1, 37) = 2.35, p = .13$. In contrast, the second ANOVA for positive condition (positive self versus control) yielded significantly higher activity for positive self traits compared to control evaluations, $F(1, 37) = 9.01, p = .005, \eta_p^2 = .20$. Finally, the ANOVA for valence (positive traits versus negative traits within the task) revealed the medial PFC was significantly more active for the positive self traits compared to the negative self traits $F(1, 37) = 20.29, p < .001, \eta_p^2 = .35$ (see Figure 3B and supplementary table S4-5).

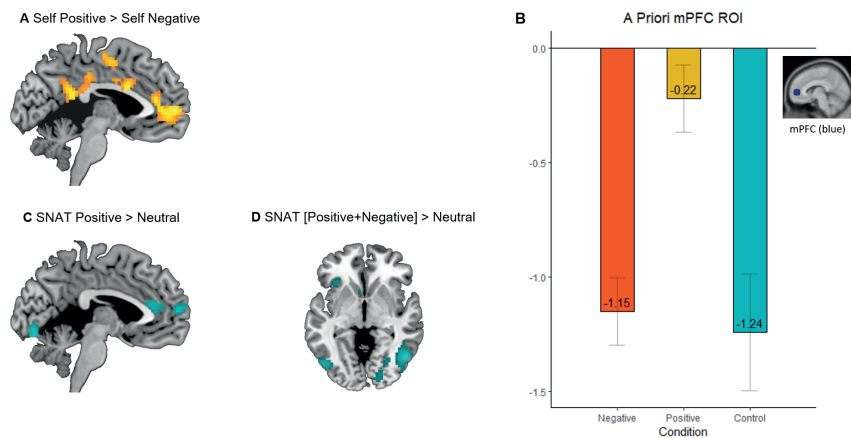


Figure 3. (A) Whole brain t-test for positive vs. negative trait statements in the self-concept task. (B) Task condition effects (for self-concept) in a pre-defined mPFC ROI, showing more activation (i.e., less deactivation) for positive self traits compared to negative traits and control statements. (C-D) Whole brain full factorial ANOVA conducted at the group level for the contrasts positive vs. negative feedback (C) and the conjunction positive vs. neutral and negative vs. neutral (D).

Table 1
Correlations with confidence intervals between task-based behavioral difference scores, questionnaires and task-based ROI conjunctions.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Positive self-evaluation															
2. Negative self-evaluation	-.25														
	[-.55, .11]														
3. Noise blast duration	-.05	.19													
	[-.39, .31]	[.17, .51]													
Negative feedback															
4. Noise blast duration	-.04	.25	.88**												
	[-.38, .31]	[.11, .55]	[.77, .94]												
Positive feedback															
5. Noise blast duration	.00	.15	.70**	.81**											
	[-.34, .35]	[-.21, .47]	[.46, .84]	[.65, .91]											
positive feedback															
6. Mean Noise blast duration	-.04	.21	.97**	.96**	.83**										
	[-.38, .31]	[.15, .52]	[.94, .99]	[.92, .98]	[.67, .91]										
7. Δ Noise blast duration (pos > neg)	.06	-.18	-.97**	-.78**	-.49**	-.88**									
	[-.30, .40]	[-.50, .18]	[-.98, -.93]	[-.89, -.60]	[-.72, -.17]	[-.94, -.77]									
8. Δ Self-evaluations (pos > neg)	.80**	-.78**	-.15	-.18	-.09	-.16	.15								
	[.63, .90]	[-.89, -.59]	[-.47, .21]	[-.50, .18]	[-.43, .27]	[-.48, .20]	[-.21, .47]								
9. Δ Noise blast duration (neg > pos)	-.06	-.04	.56**	.53**	.43*	.55**	-.54**	-.01							
	[-.41, .30]	[-.39, .32]	[.25, .76]	[.21, .74]	[.09, .68]	[.25, .76]	[-.75, -.22]	[-.37, .34]							

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
10. Δ dlPFC activity (neg > pos feedback)	.13 [-23, .47]	-.09 [-43, .27]	-.37* [-64, -.01]	-.40* [-66, -.05]	-.49** [-72, -.16]	-.42* [-67, -.07]	.28 [-.08, .58]	.14 [-22, .47]	-.30 [-59, .07]							
11. mPFC ROI self-concept (pos > neg)	.10 [-25, .44]	-.45** [-69, -.12]	-.31 [-60, .04]	-.21 [-52, .15]	-.16 [-48, .20]	-.27 [-57, .08]	.32 [-.03, .60]	.35 [-.00, .62]	-.05 [-39, .31]	-.28 [-59, .09]						
12. mPFC ROI SNAT (pos > neut)	-.11 [-44, .25]	.17 [-19, .49]	-.07 [-41, .29]	-.06 [-40, .29]	.04 [-31, .39]	-.05 [-39, .30]	.10 [-26, .43]	-.17 [-49, .19]	-.20 [-51, .17]	.10 [-26, .44]	-.17 [-49, .19]					
13. Reactive Aggression	-.24 [-55, .12]	.02 [-33, .37]	.12 [-24, .45]	.19 [-17, .50]	.16 [-20, .48]	.15 [-21, .48]	-.09 [-42, .27]	-.17 [-49, .19]	.13 [-24, .46]	-.32 [-60, .04]	.04 [-31, .38]	.00 [-35, .35]				
14. Proactive aggression	.17 [-19, .49]	-.19 [-51, .17]	.08 [-28, .42]	.05 [-30, .39]	.15 [-21, .47]	.09 [-27, .42]	-.04 [-39, .31]	.23 [-13, .54]	.16 [-21, .49]	-.41* [-67, -.07]	.25 [-11, .55]	-.27 [-56, .09]	.52** [.21, .74]			
15. Cognitive Empathy	-.60** [-78, -.31]	.16 [-20, .48]	.01 [-34, .36]	.03 [-33, .37]	-.05 [-39, .31]	.01 [-34, .35]	-.03 [-38, .32]	-.48** [-71, -.16]	-.13 [-46, .23]	.09 [-27, .43]	-.33 [-61, .02]	.13 [-23, .46]	-.06 [-40, .29]	-.38* [-64, -.04]		
16. Affective Empathy	.13 [-23, .46]	-.01 [-36, .34]	-.16 [-48, .20]	-.14 [-47, .22]	.03 [-33, .37]	-.13 [-46, .23]	.21 [-15, .52]	.09 [-26, .43]	-.14 [-47, .23]	-.22 [-53, .15]	.17 [-19, .49]	-.08 [-42, .27]	-.15 [-48, .21]	.11 [-25, .44]	.20 [-16, .51]	

Note: Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). * indicates $p < .05$. ** indicates $p < .01$.

Table 2 MNI coordinates of local maxima activated the contrasts control > positive self and positive > negative for the Self-concept task. Results were calculated using a primary voxel-wise threshold of $p < .001$ (uncorrected), with a cluster corrected threshold of $p < .05$ FDR corrected. The reversed contrast negative > positive did not result in significant effects.

Area of activation	MNI Coordinates			Test statistic <i>T</i>	Cluster Size
	x	y	z		
Control > Positive Self					
Left Lingual	-9	82	-14	6.14	143
Left Frontal Inferior Triangularis	-48	35	22	6.10	118
Positive > Negative					
Right Lingual	15	-76	-14	6.62	693
Left Precentral	-42	-25	61	6.33	964
Right Middle Temporal	42	-52	10	5.71	109
Left Frontal Superior Medial (i.e., mPFC)	-3	56	4	5.31	452
Right Rolandic Operculum	48	2	10	4.82	80
Left Anterior Cingulum	-3	8	25	4.82	163
Left Insula	-42	-4	-2	4.77	207
Left Middle Occipital	-45	-76	10	4.49	97
Right Supramarginal	60	-19	22	4.44	120
Left Postcentral	-60	-22	25	4.31	98
Left Supplementary Motor Area	-3	-7	52	4.08	64

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed. See <https://neurovault.org/collections/FOUSPROB> for a full, unthresholded overview of activation.

Table 3

MNI coordinates of local maxima activated the contrasts positive > negative, positive > neutral and positive + negative > neutral for the Social Network Aggression Task during social evaluation. Results were calculated using a primary voxel-wise threshold of $p < .005$ (uncorrected), with a cluster corrected threshold of $p < .05$ FDR corrected (cf. Achterberg et al., 2016). The reversed contrasts did not result in significant effects.

Area of activation	MNI Coordinates			Test statistic <i>T</i>	Cluster Size
	x	y	z		
Positive > Negative feedback					
Left Calcarine	3	-82	-8	6.32	571
Positive > Neutral feedback					
Right Fusiform	24	-73	-11	4.86	873
Right Frontal Superior Medial (i.e., mPFC)	9	62	4	4.13	218
Negative > Neutral feedback					
Left Middle Occipital	-45	82	4	5.93	425
Right Inferior Temporal	51	-64	5	5.04	798
Positive + Negative > Neutral feedback					
Left Middle Occipital	-45	82	7	6.04	418
Right Inferior Temporal	51	-64	-5	4.95	852
Left Putamen	-12	11	-11	9.95	142

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed. See <https://neurovault.org/collections/FOUSPROB> for a full, unthresholded overview of activation.

Table 4 MNI coordinates of local maxima activated the contrasts positive > negative, and positive > neutral for the second event of the Social Network Aggression Task (i.e., during the noise blast). Results were calculated using a primary voxel-wise threshold of $p < .001$ (uncorrected), with a cluster corrected threshold of $p < .05$ FDR corrected. The reversed contrasts did not result in significant effects.

Area of activation	MNI Coordinates			Test statistic <i>T</i>	Cluster Size
	x	y	z		
Positive > Negative feedback					
Left Parietal Inferior	-27	-67	46	5.53	1397
Left Middle Frontal Gyrus	-54	11	31	5.22	715
Right Middle Occipital Gyrus	24	-73	19	4.79	920
Right Middle Cingulum	-6	11	37	4.71	273
Right Cerebellum	21	-46	-20	4.65	297
Right Frontal Inferior Triangularis	48	32	22	4.58	269
Right Supramarginal Gyrus	51	-46	52	4.04	154
Positive > Neutral feedback					
Left Frontal Inferior Triangularis	-48	14	31	5.52	612

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed. See <https://neurovault.org/collections/FOUSPROB> for a full, unthresholded overview of activation.

3.2.2 External self-evaluation

Whole brain - Whole brain analyses were performed within a single ANOVA for social feedback (positive, neutral, negative) to test whether there was additional activation outside of the regions that were examined using ROI analyses. First, the Positive > Neutral contrast showed significant activation in the medial PFC (see Figure 3C). The reversed contrast, Neutral > Positive did not result in any significant activation. Second, the contrast Negative > Neutral resulted in significant activity in left middle occipital and right inferior temporal. Third, the contrast Positive > Negative only showed activation in the left calcarine (primary visual cortex). The reversed contrast Negative > Positive revealed no significant activation (see Table 3). Fourth, the all-valence (positive + negative) > neutral contrast revealed significant activity in the left insula (see Figure 3D and Table 3). The reversed contrast did not result in any significant activation. The unthresholded statistical maps of these contrasts are available through the NeuroVault repository under <https://neurovault.org/collections/FOUSRROB/>, images 394917-394920.

Next, whole brain regression analyses were performed with the behavioral difference scores negative > positive feedback and negative > neutral feedback, to test for relations between neural activity and noise blast length. None of the contrasts resulted in significant activation. The unthresholded statistical maps of these contrasts are available through the NeuroVault repository under <https://neurovault.org/collections/FOUSRROB/>, images 394921-394924.

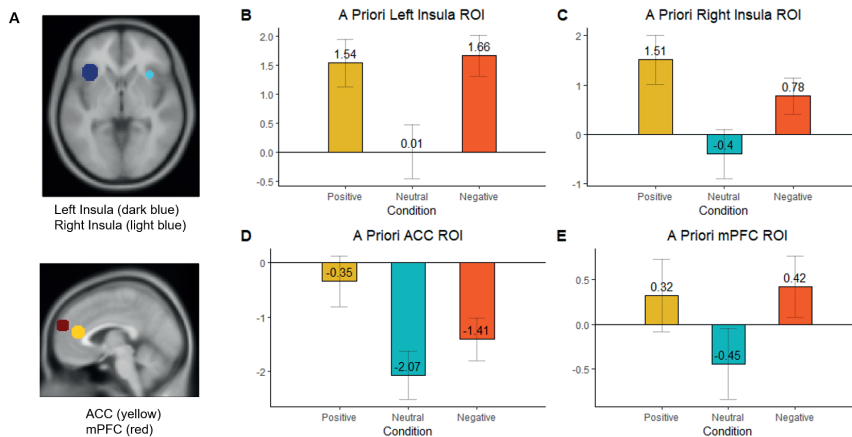


Figure 4. Task condition effects (for social feedback) in (A) four pre-defined ROIs. In general, activation was lowest for neutral feedback and higher for positive and negative feedback. Post hoc tests revealed that (B) the left insula showed more activation following negative feedback compared to neutral feedback, (C + D) the right insula and ACC showed more activation (or less de-activation) following positive feedback compared to neutral feedback, (E) there were no significant differences between conditions in mPFC activity.

ROI - To test if we could replicate the effects of social feedback, we performed repeated measures ANOVAs for the four a priori defined ROIs based on the Achterberg et al. (2016); left insula, right insula, ACC and medial PFC (see Figure 4A and supplementary Tables S7-10). The analyses resulted in main effects of feedback in left insula $F(1.83, 67.53) = 4.95$, $p = .012$, $\eta_p^2 = .118$, right insula $F(1.76, 65.06) = 4.38$, $p = .02$, $\eta_p^2 = .106$, and ACC, $F(1.93, 71.44) = 3.97$, $p = .024$, $\eta_p^2 = .097$ (all GG-corrected).

As can be seen in Figure 4B-D, for all regions, activation was lowest for neutral feedback and higher for positive and negative feedback. Post-hoc tests (bonferroni-corrected) revealed significant increased activation following positive feedback compared to neutral feedback in the ACC, $p = .004$. Post-hoc tests yielded a significantly higher activation following negative feedback compared to activation following neutral feedback in the left insula, $p = .024$. There was significantly higher activity following positive feedback vs. neutral feedback in the right insula, $p = 0.05$. The other differences between conditions in the ACC and left / right insula were not significant, all other p 's $> .076$. The repeated measures ANOVA for the medial PFC showed no significant differences between conditions, $F(2, 68) = 1.56$, $p = .22$ (see Figure 4E).

To test the relations with behavioral noise blasts, we performed correlation analyses for the a priori defined ROIs based on the Achterberg et al. (2016); the right DLPFC (Figure 5). The parameter estimates difference scores for negative > positive feedback and negative > neutral feedback were correlated with the noise blast difference scores. The correlates for the negative > positive feedback resulted in a significant negative correlation, $r = -.45$, $p = .004$, showing that smaller differences in noise blast duration between negative and positive feedback were associated with stronger DLPFC activity for negative relative to positive feedback (Figure 5). The correlation for the negative > neutral difference scores was not significant.

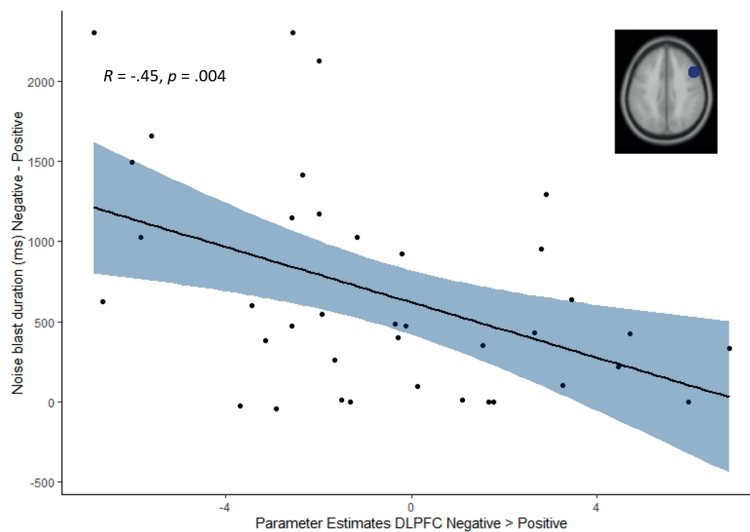


Figure 5. Negative brain-behavior association between difference scores in dIPFC activity (negative > positive feedback) and noise blast duration (negative > positive feedback).

3.3 Commonalities

3.3.1 Conjunction ROI correlations

Next, we tested whether there were associations between the overlap medial PFC region when overlapping the contrasts Positive Self > Negative Self (i.e., internal self-evaluation; SELF task) and Positive feedback > Neutral feedback (i.e., external self-evaluation; SNAT). The correlations between these mPFC activation contrasts and the behavioral ratings of each task are presented in Table 1. As can be seen in the Table, the correlation between neural activities across tasks was not significant (also see Figure 6 for a visual presentation of the mPFC overlap).

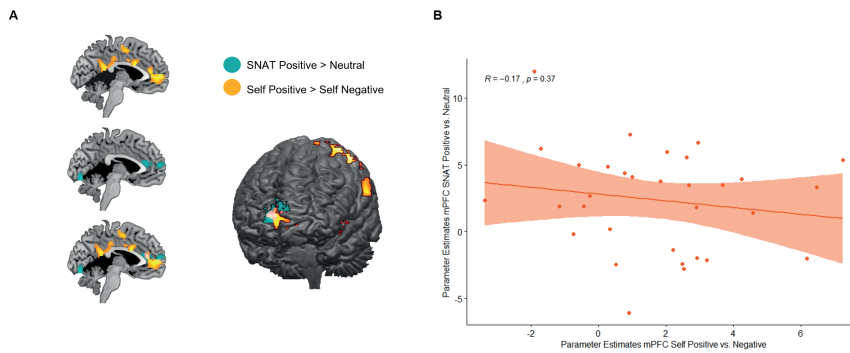


Figure 6. (A) Whole brain effects of the contrasts SNAT Positive > Neutral and Positive Self > Negative Self and their overlap and (B) Correlations between parameter estimates for the mPFC contrasts Positive Self > Negative Self and Positive feedback > Neutral feedback.

3.4 Neural activity during aggressive responses

Next, we explored whether the different feedback types (positive, negative, neutral) elicited different patterns of neural activity during the noise blast (see Table 4 for an overview of the results). First, the contrast Positive > Negative feedback resulted in significant activity in several areas, including the lateral PFC. Second, the contrast Positive > Neutral also showed activity in the lateral PFC (see Figure 7). The contrast Negative > Neutral, and all the reversed contrasts did not yield any significant results. The unthresholded statistical maps of these contrasts are available through the NeuroVault repository under <https://neurovault.org/collections/FOUSRROB/>, images 510295-510301. Finally, we exploratively tested whether difference scores in dlPFC parameter estimates during the noise blast for negative > positive feedback and negative > neutral feedback were correlated with the noise blast duration differences scores. The analyses revealed no significant correlations.

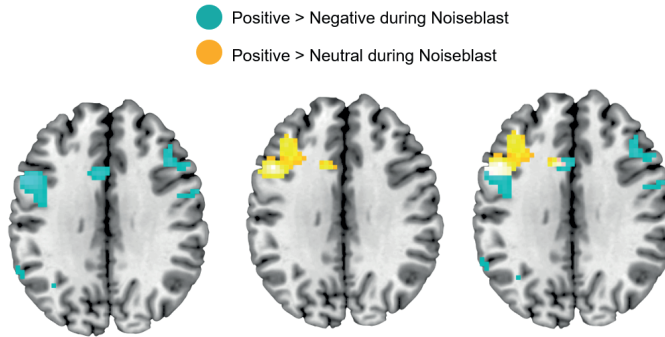


Figure 7. Whole brain effects of the contrasts SNAT Positive > Neutral and Positive > Neutral Self and their overlap during the noise blast response.

4. DISCUSSION

The main goal of this study was to examine the common and distinct neural correlates of internal self-evaluations and external self-evaluations (evaluations of the self by others when receiving social feedback), with a specific focus on the mPFC. The analyses resulted in five main findings. Our first three findings indicate that, in line with our expectations, we could replicate previous results on the distinct neural correlates of internal and external self-evaluations. In particular, (1) neural activity for self-evaluations (van der Crujisen et al., 2017; van der Crujisen et al., 2018) and (2) social feedback activity (Achterberg et al., 2016) could be replicated using region of interest analyses. Third, negative social feedback was associated with higher noise blasts, whereas stronger dorsolateral prefrontal cortex was associated with shorter noise blasts following negative social feedback, replicating prior research with the same paradigm (Achterberg et al., 2016; Achterberg et al., 2020). Fourth, with regard to a possible common neural substrate of internal and external self-evaluations, the results showed overlapping group-level activity in anterior medial PFC for positive internal self-evaluation and for receiving positive social feedback by others (external self-evaluation) (Lieberman et al., 2019). Our final exploratory aim was to examine brain activity during noise blasts, specifically related to the control of aggression after positive, neutral and negative feedback. Our findings show more lateral PFC during noise blasts to positive feedback compared to other feedback, possibly suggesting increased cognitive control to inhibit aggressive responses.

This study used two existing paradigms to examine neural activity related to internal and external self-evaluations. Internal self-evaluation was examined by testing neural activity in response to trait evaluations (Pfeifer et al., 2013; van der Crujisen et al., 2017; Veroude et al., 2014). Contrary to expectations, on a whole brain level, self-evaluations (i.e.,

positive and negative traits across domains combined) compared to control trials in which participants rated the categories of trait words, did not result in the expected medial PFC activity compared to a prior study in adolescents that used the same control condition (van der Crujisen et al., 2018). However, we did observe this activity at the ROI level for positive self-traits relative to the control condition (Denny et al., 2012; van der Crujisen et al., 2018). Furthermore, this study replicated less deactivation in medial PFC at the ROI and whole brain level for positive relative to negative self-evaluations (van der Crujisen et al., 2017). A similar anterior medial PFC region is often implicated for self-processing specifically (Koski, McHaney, Rigney, & Beer, 2020).

External self-evaluation (in response to feedback by others) was examined using the Social Network Aggression Task (Achterberg et al., 2016). Our primary aim was to test whether results previously observed in adults could be replicated using ROI analyses. Indeed, the ROI results demonstrated higher activity in anterior insula and ACC for positive and negative feedback relative to neutral trials. These findings are in line with the hypothesis that these regions are part of a salience network for high valence stimuli (Achterberg et al., 2016). However, this study showed subtle differences in the extent to which left insula, right insula and ACC were responsive to positively and negatively valenced feedback. Notably, the more rostral part of the medial PFC did not differentiate between feedback types, but this region was previously only found in children (Achterberg et al., 2018), and not in adults (Achterberg et al. 2016). These results fit with prior meta-analyses suggesting gradients of social feedback sensitivity in the medial PFC (Lieberman et al., 2019).

The next goal of this study was to test whether we could replicate the previously observed relation between dorsolateral PFC activity and retaliation following negative social feedback (Achterberg et al., 2016). Prior research demonstrated that whereas medial PFC is involved in processing self, lateral PFC is involved in working memory and control (Gilbert et al., 2006). Using the same ROI as reported in a study using the same paradigm in adults (Achterberg et al., 2016), we replicated the brain-behavior correlations following negative social feedback in the dorsolateral PFC. Behavioral analyses confirmed the expected behavioral differences in noise blast duration showing longest noise blast durations following negative social feedback and shortest noise blast durations following positive social feedback. Furthermore, stronger activity in the dorsolateral PFC following negative feedback (relative to positive feedback) was associated with shorter noise blasts after negative feedback (relative to positive feedback), thereby showing that this effect can be observed in a different sample. Similar results were previously also observed in children from a diverse social-economic background (Achterberg et al., 2020). One possibility is that retaliation may help to preserve positive self-views and therefore has an important function in protecting self (Rodman et al., 2017). In future studies, it would be interesting to test whether the shift in neural and behavioral response to a stronger focus

on positive versus negative social feedback depends on whether this is a possibility to retaliate and protect the self.

Another goal of this study was to examine whether internal and external evaluations would result in overlapping neural activity based on prior meta-analyses that have implicated medial PFC in processing of self from multiple perspectives (Lieberman et al., 2019). We observed that specifically for the positively valenced evaluations in the internal- and external self-evaluation task, there was overlap in activity in an anterior/rostral medial PFC region (Figure 6). This region corresponds most strongly with a region that was in a previous meta-analysis associated with 'self', whereas 'social' was more strongly associated with a more dorsal medial PFC region (Lieberman et al., 2019). These results might indicate that receiving positive social feedback is associated with valuing self, which was previously associated with preserving self-views following social feedback (Yoon et al., 2018). It is important to note that the overlap was observed when visualizing the whole brain results but not when correlating the parameter estimates for both types of evaluation, suggesting that the extent of medial PFC activation is not correlated across tasks. Recently, it was demonstrated that activity observed with fMRI may help to elucidate replicable mechanistic insights at the group level, but has low test-retest reliability at the individual level (Elliott et al., 2020). Therefore, more research is necessary to elucidate whether overlapping group-level activity is related to a third unmeasured variable that affects both tasks, or whether neural activation is variable across time within individuals.

In addition, we investigated whether internal and external evaluations and their neural correlates were associated with each other and with self-report measures (empathy and reactive / proactive aggression). While we did not find any significant associations between (the neural correlates of) internal and external self-evaluations, we did observe a negative association between cognitive empathy with positive and negative self-evaluations, which fits with earlier findings that self-evaluations rely on social perspective-taking skills and are informed by the opinion of others (van der Cruisen et al., 2019). Moreover, we confirmed the well-established association between proactive and reactive aggression (Card & Little, 2006). In addition, we found that cognitive empathy was negatively associated with proactive aggression, and not with reactive aggression. These findings confirm that perspective taking may be an interesting individual difference to consider with regard to observed differences in instrumental aggression (Eisenberg, Eggum & Di Giunta, 2010; Euler, Steinlin & Stadler, 2017), rather than in reactive aggression, which is thought to result from impaired emotion regulation and cognitive control (Euler et al., 2017). Nonetheless, it remains important for future research to replicate these findings in larger samples with more variation in scores (e.g. higher levels of aggression in antisocial populations), preferably while manipulating these constructs experimentally (rather than them being measured as traits) and aiming to limit shared method variance (Euler et al., 2017).

Finally, we explored whether different types of social evaluation (positive, negative, neutral) were also associated with different patterns of neural activation during aggressive responses. Our results indicate that positive feedback elicits stronger activity in the dlPFC prefrontal cortex during noise blasts, compared to both negative and neutral feedback. These findings point towards similarity in the role of the DLPFC as a region that is important for regulating aggressive responses, either directly following receipt of negative feedback and during forced noise blast responses to positive and neutral feedback. Future studies should examine the time course of DLPFC during different stages of the task in more detail and use TMS to further test potential causal roles of DLPFC in aggression regulation (Riva et al., 2015; Perach-Barzilay et al., 2013; Zheng, Li, Ye & Luo, 2020; Hoppenbrouwers et al., 2013). However, due to the explorative nature of these analyses, replication is warranted to confirm our conclusions.

Strengths, limitations and future directions

This study has a number of strengths, including the strong focus on replicability of earlier findings and the use of experimental designs to disentangle the role of valence in (the distinct and overlapping neural correlates of) internal and external self-evaluations. Examining these constructs on multiple levels of understanding (e.g. behavioral, neural) allowed us to show, for the first time, that there is indeed evidence for an overlapping neural substrate in the mPFC for internal and external self-evaluations in emerging adulthood.

However, this study also had several limitations. For instance, with the objective to directly compare tasks, the tasks themselves were relatively short and therefore based on a limited number of trials, preventing us from examining time-related effects within the task (Yoon et al., 2018). The limited number of trials is particularly evident in the control condition of the self-concept task, which may have resulted in a lack of power to detect effects between the self- and control-conditions. However, as our primary focus was to compare positive and negative trials within the self-condition, this limitation does not affect our main analyses and conclusions. For future research, we recommend larger and more diverse groups with more trials per task, and multiple measurements to test for convergent validation and reliability across time. Inclusion of more, both positive and negative trials in a control condition also offers the opportunity to test for interaction effects between valence and condition on a neural level, which may enhance our understanding of the processes at hand.

A second limitation is that the different conditions of the self-concept task may involve related, but slightly different processes (i.e., categorization of traits in the control-condition vs. rating of traits in the self-condition), which may have introduced a confound. Our aim was to select a control condition that involves thinking about the traits, but not applying them to self or others (Van der Crujisen et al., 2018). Notably, control conditions that involve rating whether traits apply to (close and distant) others have their own limitations.

In particular, rating close others (such as friends and family members) results in much overlap with self-evaluations (e.g., Denny et al., 2012), possibly because the self is a social construct and it is not possible to completely distinguish self from close others (Van der Crujisen et al., 2019). Having multiple control task conditions including more distant to more similar others may help to reveal self-related activation in more detail (Feng, Yan, Huang, Han & Ma, 2018).

Finally, a third limitation is that we cannot completely rule out the possibility that differences in hemispheric localization influenced our neural findings. In the current study, we took a population-based approach and therefore included both right and left handed individuals, with the proportion of left-handed participants approximately corresponding to the observed frequency in the general population (Willems et al., 2014). The majority of results remained the same when left-handed participants were excluded, with the exception of the external evaluation condition effects on left and right insula activity, which may have been due to either differences in hemispheric localization or reduced statistical power associated with the sample size reduction. Future studies may consider either including larger sub-samples of left -and right-handed participants to allow for direct comparisons of lateralization effects, or use individual level statistical analyses methods that are less sensitive to lateralization effects than group level analyses (Willems et al., 2014).

Conclusion

Taken together, this study confirmed an overlapping role of medial PFC in internal self-evaluation and external self-evaluations (social feedback processing). As such, the study contributes to the literature on self-evaluation that has reported multiple sub region activities in medial PFC for direct and social self (Denny et al., 2012; Lieberman et al., 2019). We replicate previously reported activities using ROI analyses, both in general task contrasts as well and brain-behavior relations. In addition, we show that receiving positive feedback results in increased dlPFC activity during retaliatory responses, shedding light at the neural processes that are engaged *during* aggressive responses in social contexts. These findings add to the need to have replications within functional MRI neuroscience (Elliott et al., 2020). The next step is to see how these findings have implications for studies on individual differences such as for individuals with low self-esteem (Somerville, Kelley, & Heatherton, 2010) or antisocial behaviors that may differentially affect evaluation of self and others (Moffitt et al., 2011).

REFERENCES

1. Achterberg, M., van Duijvenvoorde, A. C., Bakermans-Kranenburg, M. J., & Crone, E. A. (2016). Control your anger! The neural basis of aggression regulation in response to negative social feedback. *Soc Cogn Affect Neurosci*, *11*(5), 712-720. doi:10.1093/scan/nsv154
2. Achterberg, M., van Duijvenvoorde, A. C. K., van der Meulen, M., Bakermans-Kranenburg, M. J., & Crone, E. A. (2018). Heritability of aggression following social evaluation in middle childhood: An fMRI study. *Hum Brain Mapp*, *39*(7), 2828-2841. doi:10.1002/hbm.24043
3. Achterberg, M., van Duijvenvoorde, A. C. K., van der Meulen, M., Euser, S., Bakermans-Kranenburg, M. J., & Crone, E. A. (2017). The neural and behavioral correlates of social evaluation in childhood. *Dev Cogn Neurosci*, *24*, 107-117. doi:10.1016/j.dcn.2017.02.007
4. Achterberg, M., van Duijvenvoorde, A. C. K., van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Crone, E. A. (2020). Longitudinal changes in DLPFC activation during childhood are related to decreased aggression following social rejection. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.1915124117
5. Bertsch, K., Florange, J., & Herpertz, S. C. (2020). Understanding brain mechanisms of reactive aggression. *Current psychiatry reports*, *22*(12), 1-16. doi:10.1007/s11920-020-01208-6
6. Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. *Neuroimage*, *16*(2), 497.
7. Card, N. A., & Little, T. D. (2006). Proactive and reactive aggression in childhood and adolescence: A meta-analysis of differential relations with psychosocial adjustment. *International Journal of Behavioral Development*, *30*(5), 466-480. <https://doi.org/10.1177/0165025406071904>
8. Chester, D. S., Lynam, D. R., Milich, R., & DeWall, C. N. (2018). Neural mechanisms of the rejection-aggression link. *Soc Cogn Affect Neurosci*, *13*(5), 501-512. doi:10.1093/scan/nsy025
9. Card, N. A., & Little, T. D. (2006). Proactive and reactive aggression in childhood and adolescence: A meta-analysis of differential relations with psychosocial adjustment. *International Journal of Behavioral Development*, *30*(5), 466-480. <https://doi.org/10.1177/0165025406071904>
10. Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B., & Evans, A. C. (1997). BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage*, *5*, 425.
11. Cumming, G. (2013). The New Statistics: Why and How. *Psychological Science*. <https://doi.org/10.1177/0956797613504966>
12. Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113-126. <https://doi.org/10.1037/0022-3514.44.1.113>
13. Eisenberg, N., Eggum, N. D., & Giunta, L. D. (2010). Empathy-Related Responding: Associations with Prosocial Behavior, Aggression, and Intergroup Relations. *Social Issues and Policy Review*, *4*(1), 143-180. <https://doi.org/10.1111/j.1751-2409.2010.01020.x>
14. Feng, C., Yan, X., Huang, W., Han, S., & Ma, Y. (2018). Neural representations of the multidimensional self in the cortical midline structures. *NeuroImage*, *183*, 291-299. <https://doi.org/10.1016/j.neuroimage.2018.08.018>
15. Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: A web-based

- repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00008>
16. Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*, 6th Edn. Northridge, CA: California State University.[Google Scholar].
 17. Crone, E. A., & Fuligni, A. J. (2020). Self and Others in Adolescence. *Annu Rev Psychol*, 71, 447-469. doi:10.1146/annurev-psych-010419-050937
 18. Dale, A.M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8, 109 – 14. doi: 10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.3.CO;2-N
 19. D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: the valuation hypothesis. *Front Hum Neurosci*, 7, 372. doi:10.3389/fnhum.2013.00372
 20. Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
 21. de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D., & Yarkoni, T. (2016). Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *J Neurosci*, 36(24), 6553-6562. doi:10.1523/JNEUROSCI.4402-15.2016
 22. Denny, B.T., Kober, H., Wager, T.D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J Cogn Neurosci*, 24(8), 1742-1752. doi:10.1162/jocn_a_00233
 23. Eisenberg, N., Eggum, N. D., & Giunta, L. D. (2010). Empathy-Related Responding: Associations with Prosocial Behavior, Aggression, and Intergroup Relations. *Social Issues and Policy Review*, 4(1), 143–180. <https://doi.org/10.1111/j.1751-2409.2010.01020.x>
 24. Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., . . . Hariri, A. R. (2020). What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. *Psychol Sci*, 956797620916786. doi:10.1177/0956797620916786
 25. Flagan, T., & Beer, J. S. (2013). Three ways in which midline regions contribute to self-evaluation. *Front Hum Neurosci*, 7, 450. doi:10.3389/fnhum.2013.00450
 26. Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *J Cogn Neurosci*, 18(6), 932-948. doi:10.1162/jocn.2006.18.6.932
 27. Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00008>
 28. Gunther Moor, B., van Leijenhorst, L., Rombouts, S. A., Crone, E. A., & Van der Molen, M. W. (2010). Do you like me? Neural correlates of social evaluation and developmental trajectories. *Soc Neurosci*, 5(5-6), 461-482. doi:10.1080/17470910903526155
 29. Harter, S. (2012). *The construction of the self: Second Edition: Developmental and Sociocultural Foundations*. New York: The Guilford Press.
 30. Hoppenbrouwers, S. S., De Jesus, D. R., Stirpe, T., Fitzgerald, P. B., Voineskos, A. N., Schutter, D. J. L. G., & Daskalakis, Z. J. (2013). Inhibitory deficits in the dorsolateral prefrontal cortex in psychopathic offenders. *Cortex*, 49(5), 1377–1385. <https://doi.org/10.1016/j.cortex.2012.06.003>
 31. Hughes, B. L., & Beer, J. S. (2013). Protecting the self: the effect of social-evaluative threat on neural representations of self. *J Cogn Neurosci*, 25(4), 613-622. doi:10.1162/jocn_a_00343

32. Konijn, E. A., Bijvank, M. N., & Bushman, B. J. (2007). I wish I were a warrior: the role of wishful identification in the effects of violent video games on aggression in adolescent boys. *Dev Psychol*, 43(4), 1038-1044. doi:10.1037/0012-1649.43.4.1038
33. Korn, C. W., Prehn, K., Park, S. Q., Walter, H., & Heekeren, H. R. (2012). Positively biased processing of self-relevant social feedback. *J Neurosci*, 32(47), 16832-16844. doi:10.1523/JNEUROSCI.3016-12.2012
34. Koski, J. E., McHaney, J. R., Rigney, A. E., & Beer, J. S. (2020). Reconsidering longstanding assumptions about the role of medial prefrontal cortex (MPFC) in social evaluation. *Neuroimage*, 214, 116752. doi:10.1016/j.neuroimage.2020.116752
35. Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neurosci Biobehav Rev*, 99, 311-328. doi:10.1016/j.neubiorev.2018.12.021
36. Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proc Natl Acad Sci U S A*, 108(7), 2693-2698. doi:10.1073/pnas.1010076108
37. 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. <https://doi.org/10.1080/17470919.2012.720602>
38. Pfeifer, J. H., Kahn, L. E., Merchant, J. S., Peake, S. J., Veroude, K., Masten, C. L., . . . Dapretto, M. (2013). Longitudinal change in the neural bases of adolescent social self-evaluations: effects of age and pubertal development. *J Neurosci*, 33(17), 7415-7419. doi:10.1523/JNEUROSCI.4074-12.2013
39. Pfeifer, J. H., & Peake, S. J. (2012). Self-development: integrating cognitive, socioemotional, and neuroimaging perspectives. *Dev Cogn Neurosci*, 2(1), 55-69. doi:10.1016/j.dcn.2011.07.012
40. 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: Official Journal of the International Society for Research on Aggression*, 32(2), 159-171. doi: 10.1002/ab.20115
41. Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of personality assessment*, 93(1), 84-95. doi: 10.1080/00223891.2010.528484
42. Riva, P., Romero Lauro, L. J., DeWall, C. N., Chester, D. S., & Bushman, B. J. (2015). Reducing aggressive responses to social exclusion using transcranial direct current stimulation. *Social cognitive and affective neuroscience*, 10(3), 352-356. <https://doi.org/10.1093/scan/nsu053>
43. Rodman, A. M., Powers, K. E., & Somerville, L. H. (2017). Development of self-protective biases in response to social evaluative feedback. *Proc Natl Acad Sci U S A*, 114(50), 13158-13163. doi:10.1073/pnas.1712398114
44. Somerville, L. H., Heatherton, T. F., & Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat Neurosci*, 9(8), 1007-1008. doi:10.1038/nn1728
45. Somerville, L. H., Jones, R. M., Ruberry, E. J., Dyke, J. P., Glover, G., & Casey, B. J. (2013). The medial prefrontal cortex and the emergence of self-conscious emotion in adolescence. *Psychol Sci*, 24(8), 1554-1562. doi:10.1177/0956797613475633

46. Somerville, L. H., Kelley, W. M., & Heatherton, T. F. (2010). Self-esteem modulates medial prefrontal cortical responses to evaluative social feedback. *Cereb Cortex*, *20*(12), 3005-3013. doi:10.1093/cercor/bhq049
47. Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*, 6th Edn. Northridge, CA: California State University.
48. van der Cruijssen, R., Peters, S., & Crone, E. A. (2017). Neural correlates of evaluating self and close-other in physical, academic and prosocial domains. *Brain Cogn*, *118*, 45-53. doi:10.1016/j.bandc.2017.07.008
49. van der Cruijssen, R., Peters, S., van der Aar, L. P. E., & Crone, E. A. (2018). The neural signature of self-concept development in adolescence: The role of domain and valence distinctions. *Dev Cogn Neurosci*, *30*, 1-12. doi:10.1016/j.dcn.2017.11.005
50. Van der Cruijssen, R., Peters, S., Zoetendaal, K. P. M., Pfeifer, J. H., & Crone, E. A. (2019). Direct and reflected self-concept show increasing similarity across adolescence: A functional neuroimaging study. *Neuropsychologia*, *129*, 407-417. <https://doi.org/10.1016/j.neuropsychologia.2019.05.001>
51. Veroude, K., Jolles, J., Croiset, G., & Krabbendam, L. (2014). Sex differences in the neural bases of social appraisals. *Soc Cogn Affect Neurosci*, *9*(4), 513-519. doi:10.1093/scan/nst015
52. Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience*, *15*(3), 193-201. doi:10.1038/nrn3679
53. Woo, C. W., Koban, L., Kross, E., Lindquist, M. A., Banich, M. T., Ruzic, L., . . . Wager, T. D. (2014). Separate neural representations for physical pain and social rejection. *Nat Commun*, *5*, 5380. doi:10.1038/ncomms6380
54. Yoon, L., Somerville, L. H., & Kim, H. (2018). Development of MPFC function mediates shifts in self-protective behavior provoked by social feedback. *Nat Commun*, *9*(1), 3086. doi:10.1038/s41467-018-05553-2
55. Zheng, W., Li, Y., Ye, H., & Luo, J. (2021). Effect of Modulating DLPFC Activity on Antisocial and Prosocial Behavior: Evidence From a tDCS Study. *Frontiers in Psychology*, *11*, 579792. <https://doi.org/10.3389/fpsyg.2020.579792>

Acknowledgements

The authors declare no conflict of interest. We would like to thank all participants for their participation, and Desana Kocevka, Merel Spaander, Cevdet Acaroy, Sjors Bootsman, Esmee Haverkort, Angelique de Groot and Eline Vink for their help with data collection.

SUPPLEMENTARY MATERIALS

Table S1

Sample and sub-sample descriptive statistics

Measure	Group			
	Behavioral analyses (n = 39)	fMRI internal evaluations (n = 38)	fMRI external evaluations (n = 35)	fMRI overlap (n = 34)
Gender [n males/females]	16/23	16/22	15/20	15/19
Age [M (SD)]	22.7 (3.07)	22.8 (3.11)	22.94 (3.08)	23.06 (3.04)
Education [n]				
Vocational	16	15	14	13
College	18	17	17	17
University	4	4	4	4
Other	1	1	0	0
IQ [M (SD)] ¹	107.47 (11.64)	107.67 (11.65)	107.89 (12.11)	
Right-handed [n right/left]		34/4	31/4	30/4
Ethnicity [n]				
Dutch	35	34	32	31
Surinamese	2	2	1	1
Different Western	1	1	1	1
Different non-Western	1	1	1	1

Note. IQ, estimated IQ based on two subscales of the Wechsler Adult Intelligence Scale-IV (Similarities and Block Design).

Table S2

ANOVA results for the influence of domain and valence on mean internal self-evaluation ratings

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
Valence	1	38	412.94	<.0001	.92
Domain	1	38	16.32	.0003	.30
Valence x Domain	1	38	3.28	.08	.08

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S3

ANOVA results for the influence of condition on mean noise blast durations

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
Intercept	1	38	94.44	< .001	.71
Condition	2	72	41.15	< .001	.52

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S4

ANOVA results for the influence of task condition (self vs. control) on mPFC activity (internal self-evaluations)

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1	37	33.33	<.0001	.47
Condition	1	37	2.35	.13	.06

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S5

ANOVA results the influence of task condition (positive self vs. control) on mPFC activity (internal self-evaluations)

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1	37	15.72	< .001	.30
Condition	1	37	9.01	.005	.20

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S6

ANOVA results for the influence of valence on mPFC activity (internal self-evaluations)

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1	37	18.57	< .001	0.33
Valence	1	37	20.29	<.0001	0.35

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S7

ANOVA results for the effect of condition on ACC activity (external evaluations)

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	34.00		21.12	.000	.363
Condition	1.93	71.44	0.97	3.97	.024	.097

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Table S8

ANOVA results for the effect of condition on left Insula activity (external evaluations)

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	34.00		10.81	.002	.226
Condition	1.83	67.53	0.91	4.95	.012	.118

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Table S9

ANOVA results for the effect of condition on right Insula activity (external evaluations)

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	34.00		4.37	.044	.106
Condition	1.76	65.06	0.88	4.38	.020	.106

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Table S10

ANOVA results for the effect of condition on mPFC activity (external evaluations)

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	34.00		0.10	.750	.00
condition	1.93	65.61	0.96	1.56	.219	.02

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Appendix A. The influence of handedness of neural results.

To test whether the reported neural effects (sections 3.2.) were influenced by handedness of the participants, we repeated the analyses when left-handed participants ($n = 5$) were excluded. Left-handedness did not seem to influence the significance or direction of the majority of the effects, with the exception of the bilateral insula during external evaluations.

Table S11

ANOVA results for the influence of task condition (self vs. control) on mPFC activity (internal self-evaluations)

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1.00	32.00	23.29	< .0001	.421
Condition	1.00	32.00	1.163	.289	.035

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S12

ANOVA results the influence of task condition (positive self vs. control) on mPFC activity (internal self-evaluations)

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1.00	32.00	10.39	.003	.245
Condition	1.00	32.00	5.86	.021	.155

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S13

ANOVA results for the influence of valence on mPFC activity (internal self-evaluations)

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1.00	32.00	15.31	< .001	.323
Valence	1.00	32.00	16.84	< .001	.345

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. η^2_p indicates partial eta-squared.

Table S14

ANOVA results for the effect of condition on ACC activity (external evaluations)

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	29.00		25.23	.000	.465
Condition	1.94	56.29	0.97	4.40	.018	.132

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Table S15*ANOVA results for the effect of condition on left Insula activity (external evaluations)*

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	29.00		13.61	< .001	.319
Condition	1.84	53.35	0.92	2.70	.076	.085

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Table S16*ANOVA results for the effect of condition on right Insula activity (external evaluations)*

Predictor	df_{Num}	df_{Den}	F	p	η^2_p
(Intercept)	1.00	29.00	6.26	.018	.177
Condition	2.00	58.00	2.81	.068	.088

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Table S17*ANOVA results for the effect of condition on mPFC activity (external evaluations)*

Predictor	df_{Num}	df_{Den}	Epsilon	F	p	η^2_p
(Intercept)	1.00	29.00		0.00	.696	.00
condition	1.88	54.66	0.94	0.59	.550	.02

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. Epsilon indicates Greenhouse-Geisser multiplier for degrees of freedom, p -values and degrees of freedom in the table incorporate this correction. η^2_p indicates partial eta-squared.

Appendix B. Traits sentences used in the self-concept task

Trait sentences used in the self-evaluation condition of the self-concept task:

Physical			
Original Dutch		Free English Translation	
Positive	Negative	Positive	Negative
Ik zie er aantrekkelijk uit	Ik ben dik	I look attractive	I am fat
Ik ben mooi	Ik heb overgewicht	I am beautiful	I am overweight
Ik ben knap	Ik ben lelijk	I am handsome	I am ugly
Ik zie er goed uit	Ik zie er suf uit	I look good	I look silly
Ik heb een goede kledingstijl	Ik zweet veel	I have a great clothing style	I sweat a lot
Ik heb een goed lichaam	Ik ben te zwaar	I have a great body	I am too heavy
Ik heb een mooi figuur	Ik ben onaantrekkelijk	I have a great figure	I am unattractive
Ik mag blij zijn met mijn lichaam	Ik zie er onverzorgd uit	I can be happy with my body	I look unkempt
Ik zie er stralend uit	Ik heb lelijke tanden	I look radiant	I have ugly teeth
Ik heb een mooie lach	Ik heb een slechte huid	I have a nice smile	I have bad skin

Prosocial			
Original Dutch		Free English Translation	
Positive	Negative	Positive	Negative
Ik troost anderen	Ik zorg alleen voor mezelf	I comfort others	I only take care of myself
Ik doe graag iets voor een ander	Ik houd alleen rekening met mezelf	I like to do something for someone else	I only consider myself
Ik houd rekening met anderen	Ik laat anderen hun problemen zelf oplossen	I take others into account	I let others solve their problems themselves
Ik leef met anderen mee	Ik houd alles voor mezelf	I empathize with others	I keep everything to myself
Ik geef om anderen	Ik help anderen alleen als Ik er iets voor terug krijg	I care about others	I only help others if I get something in return
Ik voel met anderen mee	Ik kies altijd voor mezelf	I sympathize with others	I always choose myself over others

Prosocial			
Original Dutch		Free English Translation	
Positive	Negative	Positive	Negative
Ik help anderen	Ik denk vooral aan mezelf	I help others	I mostly think about myself
Ik zorg graag voor anderen	Ik negeer andermans problemen	I like to take care of others	I ignore other people's problems
Ik kom voor anderen op	Ik leen mijn spullen niet graag uit	I stand up for others	I don't like to lend my stuff
Ik deel graag met anderen	Ik help nooit een vreemde	I like to share with others	I never help a stranger

Trait sentences used in the control condition of the self-concept task:

Physical			
Original Dutch		Free English Translation	
Positive	Negative	Positive	Negative
Mooie ogen hebben	Mollig zijn	Having beautiful eyes	Being chubby
Een gezond gewicht hebben	Er moe uitzien	Have a healthy weight	Looking tired
Tevreden zijn met je uiterlijk	Puistjes hebben	Be happy with your appearance	Having pimples

Prosocial			
Original Dutch		Free English Translation	
Positive	Negative	Positive	Negative
Anderen vergeven	Gemeen zijn tegen anderen	Forgiving others	Be mean to others
Spullen delen met anderen	Anderen buiten sluiten	Share stuff with others	Exclude others
Ruzies oplossen	Ruzie maken	Solve fights	Fight/argue



Through the looking glass: the neural basis of self-concept in young adults with antisocial trajectories

This manuscript is in revision, and has been preprinted as: van de Groep, I. H., Bos, M., Nauta-Jansen, L., Popma, A., & Crone, E. (2022). Through the looking glass: the neural basis of self-concept in young adults with antisocial trajectories. <https://doi.org/10.31234/osf.io/b7f6k>

ABSTRACT

Self-concept is shaped by social experiences, but it is not yet well understood how the neural and behavioral development of self-concept is influenced by a history of antisocial behavior. In this pre-registered study, we examined neural responses to self-evaluations in young adults who engaged with antisocial behavior in childhood and either desisted or persisted in antisocial behavior. A self-concept task was performed by 94 young adults (age range 18–30-yrs). During the task, participants with a persistent or desistant antisocial trajectory ($N = 54$), and typically developing young adults ($N = 40$) rated whether positive and negative traits in different domains (prosocial, physical) described themselves. We examined both the effects of a history of antisocial behavior as well as current heterogeneity in psychopathic traits on self-concept appraisal and its neural underpinnings. Participants endorsed more positive than negative trait statements across domains, which did not differ between antisocial-history groups. However, current psychopathic traits were negatively associated with prosocial self-concept and mPFC activity during self-evaluation. Together, these findings suggest that antisocial tendencies might indeed be reflected in self-concept development of young adults, specifically in the prosocial domain.

Keywords: Self-Concept; Antisocial behavior; Psychopathy; fMRI

1. INTRODUCTION

Self-concept, defined as the knowledge and evaluation of ourselves, is shaped by both cognitive development and social experiences (van de Schoot & Wong, 2012). Forming a positive and stable self-concept is an important developmental milestone (Crone et al., 2022), which enhances our general well-being and protects us from mental health problems (Marsh et al., 2001), while persistent negative and unstable evaluations often increase the risk of problem behavior and difficulties in socially adaptive functioning (Ybrandt, 2008). Together, self-appraisal (i.e., how positively or negatively we evaluate our traits) and self-concept clarity (i.e., whether self-knowledge is clearly defined, internally consistent and stable over time) comprise the cognitive parts of the self (Crone et al., 2022; Campbell et al., 1996).

Recent theories indicate that these cognitive aspects of the self are important for understanding different developmental trajectories of antisocial behavior. For instance, identity and self-concept development are thought to play an important role in desistance from crime and antisocial behavior (see e.g., Rocque et al., 2016). More specifically, desistance from antisocial behavior is likely preceded or accompanied by identity changes to a more positive and prosocial self (Paternoster et al., 2016; Paternoster & Bushway, 2009; Rocque et al., 2016). While some of these changes already start in adolescence, they may be particularly prominent during early adulthood, a period characterized by continuing identity development, changes in social contexts and exploration of social roles (Arnett, 2007; Rocque et al., 2016). In particular, during adolescence, the self-concept becomes increasingly complex and multi-faceted, a process supported by ongoing brain maturation and the development of higher-order cognitive capacities, such as self-reflection and perspective-taking (van der Crujisen et al., 2017; 2018; Crone et al., 2022). Identifying whether and in whom these cognitive changes arise is important, given that self-concept influences how people see themselves and others in social interactions, and accordingly how (well) they behave in social contexts (Crone et al., 2022). Therefore, we aim to examine and characterize how a history of antisocial behavior (i.e., young adults who persisted in or desisted from antisocial behavior throughout development, relative to typically developing controls) is associated with the neural correlates of positive and negative self-concept evaluations in early adulthood. However, given that the development of antisocial behavior is marked by substantial heterogeneity, we additionally use an individual difference approach using psychopathic traits to further understand the association between self-concept appraisal and antisocial experiences.

The neural basis of self-concept

Neuroimaging methods have demonstrated to be a valuable approach to study self-evaluations by overcoming several difficulties (i.e., that self-concept is not directly observable and that self-report measures are subject to response bias) and providing the

added value of identifying neurobiological mechanisms underlying behavior. The neural basis of self-concept is often studied using trait judgment paradigms, by asking participants whether and to what extent certain trait statements are descriptive of themselves. Studies using such paradigms have repeatedly shown increased activity in the ventral and rostral medial prefrontal cortex (Denny et al., 2012; Lieberman et al., 2019; Murray et al., 2012) during self-evaluations, compared to other-evaluations or baseline conditions. While there is indeed ample support for general neural mechanisms underlying self-representation, research increasingly supports the view that self-concept is a multi-dimensional construct that also involves distinct, specific knowledge structures. In particular, self-concept evaluations and their neural underpinnings may depend on the domain (e.g., physical, prosocial) and valence (i.e., positive, negative) of self-knowledge (e.g. van der Cruijssen et al., 2018; van der Cruijssen et al., 2017). Positive self-concepts, or more applicable traits, are associated with stronger activation in the anterior medial prefrontal cortex (mPFC) compared to negative traits (D'Argembeau, 2013; van der Cruijssen et al., 2017).

Possibly, if people with persistent antisocial tendencies are prone to more negative self-evaluations, they show attenuated mPFC activity during self-evaluations (Herpertz et al., 2018; van der Aar et al., 2019). Indeed, youth diagnosed with conduct disorder show reduced mPFC activity in the rostral anterior mPFC during fMRI tasks that require self-reflective thoughts (Dalwani et al., 2014). This attenuated activity was accompanied by aberrant Default Mode Network (DMN) connectivity between the aMPFC and other DMN sub-regions important for self-related processing, such as the Posterior Cingulate (PCC) (Broulidakis et al., 2016; Dalwani et al., 2014). Moreover, prior studies indicate that mPFC is often structurally or functionally impaired in antisocial populations (Alegria et al., 2016; Fairchild et al., 2019; Boccardi et al., 2011; de Oliveira-Souza et al., 2008; Ermer et al., 2012; Raine, 2008). Taken together, it is important to examine whether individuals with persistent antisocial behavior (1) evaluate their traits more negatively, and (2) show less prefrontal activity during positive self-appraisals.

Self-concept develops in different social contexts. While self-traits can be evaluated globally, they often depend on respective domains, such as prosocial traits or physical characteristics (van der Cruijssen, 2017; 2018). In previous studies, we observed that evaluating self-traits in the physical domain (e.g. looking attractive, a positive physical trait) resulted in increased activity in the lateral prefrontal cortex (e.g. dlPFC), whereas evaluations of prosocial traits (e.g. caring for others) were associated with increased medial frontal activity (van der Cruijssen et al., 2017, 2018). Considering domain-specificity in the relationship between antisocial behavior and self-concept may further enhance our understanding of self-concept in individuals showing persistent versus desisting antisocial behavior (Ostrowsky, 2010; Paternoster et al., 2016). Particularly self-concept in the domain of prosocial behavior (such as giving, helping others) is of interest, given that individuals who desist from antisocial behavior are thought to construe a more prosocial self-image

prior to and during the desistance process (Paternoster et al., 2016; Rocque et al., 2016) than individuals who persist in antisocial behavior.

Psychopathic traits

While differentiation between different antisocial developmental trajectories improves our understanding of general antisocial behavioral patterns over time, additionally using individual difference approaches helps to explain heterogeneity in antisocial behavior (Garvey et al., 2016). Antisocial behavior is often accompanied by high levels of psychopathic personality traits, which contribute to the emotional, interpersonal, and behavioral difficulties associated with maladaptive social behavior (Zalk & Zalk, 2015). Although the overall construct of psychopathy may be associated with a negative self-concept in adults (Gudjonsson & Roberts, 1983), and with structural and functional impairments of the mPFC (Ermer et al., 2012; Fanti et al., 2018; Johanson et al., 2020; Koenigs, 2012), recent studies suggest that while the different but interrelated dimensions of psychopathy (i.e., Grandiose-Manipulative, Callous-Unemotional, and Impulsive-Irresponsible traits) tend to co-occur in individuals, they are associated with different behavioral outcomes and physiological and neurobiological underpinnings, both in adolescence and early adulthood. In youth, Callous-Unemotional traits have been associated with decreased mPFC activity during self-referential processing (Bontemps et al., 2022). In contrast, Grandiose-Manipulative traits have been associated with increased activity in medial frontal regions, as well as the right dlPFC in a recent EEG study (Bontemps et al., 2022). Impulsive-Irresponsible traits have also been associated with dysfunctional functioning in the dlPFC during self-processing (Bontemps et al., 2022), although this deficit may be more left-lateralized (Bontemps et al., 2022; Hoppenbrouwers et al., 2013). However, another study did not find differences between Impulsive-Irresponsible tendencies and medial and lateral frontal regions during self-processing (Deming et al., 2018).

The current study

Taken together, we used two complementary approaches to study self-concept and antisocial tendencies. First, we used a group-based approach, to identify possible behavioral and neural differences in self-concept between individuals with different developmental trajectories of antisocial behavior (persistent, desistant, control). On a behavioral level, we hypothesized that individuals showing persistent antisocial behavior would endorse fewer positive traits and more negative traits (compared to desisters and the control group) (Hypothesis 1a, Paternoster et al., 2009, 2016; van der Crujisen et al., 2017, 2018). Moreover, we expected that this difference would be more pronounced in the prosocial domain than in the physical domain (Hypothesis 1b, Paternoster et al., 2009; 2016). On a neural level, we hypothesized (Hypothesis 2a) that individuals who persisted in antisocial behavior would show less neural activity in the mPFC compared to desisters

and the control group in both domains, but possibly more evidently in the prosocial domain (Herpertz, Bertsch & Jeung, 2018; van der Aar et al., 2019). Regarding the domain-specificity of the effects, we expected to replicate that the contrast physical > prosocial traits would result in stronger activity in the dlPFC, whereas the contrast prosocial traits > physical traits was expected to result in increased mPFC activity, across all participants (Hypothesis 2b, van der Cruisen, 2017; 2018). Moreover, we expected that individuals who persisted in antisocial behavior would show less neural activity in the mPFC during prosocial trait evaluations, compared to physical trait evaluations.

Second, we used an individual difference approach to investigate the role of psychopathic traits in self-evaluation. We hypothesized that different psychopathic traits dimensions would be differentially associated with self-evaluations (Hypothesis 3a, see Appendix A). Third, we explored (Hypothesis 3b) whether (1) individuals with high Callous-Unemotional traits showed decreased mPFC activity (Marsh et al., 2008) and (2) whether individuals with high Grandiose-Manipulative traits showed increased or decreased mPFC and dlPFC activity when making positive self-evaluations (vs. negative and control evaluations). Moreover, we explored the associations between Impulsive-Irresponsible traits and neural activity during self-appraisal.

METHODS

2.1 Participants

In the current, pre-registered study (<https://osf.io/6fgbs/>), we included two subsamples, comprised of (1) young adults from a childhood arrestee cohort (i.e. who were arrested by the police before the age of 12), who participated in the current and previous waves of this longitudinal study (childhood arrestee sample: $N = 54$, see van de Groep et al., 2022), and (2) young adults without a history of antisocial behavior, who completed the same measures and MRI protocol (control sample: $N = 40$; see van de Groep et al., 2021; 2022; see Appendix A for the procedure). Participants from the childhood arrestee sample (all arrested before the age of 12 years (initial sample $N = 364$)) were followed in their early and late adolescence, into young adulthood (see Figure S5A for an overview of the five different time points; T1 (2003–2006, mean age 10.9 (SD = 1.4)), T2 (2004–2008, mean age 11.4 (SD = 1.5)), T3 (2005–2008, mean age 13.1 (SD = 1.5)), T4 (2010–2012, mean age 17.6 (SD = 1.4)), T5 (2019–2021, mean age 25.5 (SD = 1.7))), and were classified as showing persistent or desistant antisocial developmental trajectories ($N = 54$, see van de Groep et al., 2022, see also section 2.2.3).

Participants were screened for fMRI contra-indications, had normal (or corrected-to-normal) vision, and spoke Dutch fluently. They were excluded from fMRI analyses in case they did not perform or complete the task, if the MRI data was corrupted or if participants showed excessive head motion (> 3mm), resulting in final sample of 90 participants (Control sample $N = 38$, Childhood arrestee cohort $N = 52$). Analyses on

behavioral results were conducted for all participants (Control sample N = 40, Childhood arrestee cohort N = 54) (for descriptive data, see Table 1).

All subjects gave written informed consent, were debriefed about the study aim after the experiment, and received a financial reimbursement for their participation. The study protocol was approved by the VU University Medical Center Medical Ethical Committee (registration number 2009.268 - NL28844.029.09), with local approval from the Leiden Institute for Brain and Cognition.

2.2. Materials

2.2.1 Self-concept task

To investigate self-concept, participants performed an adapted version of the self-concept fMRI task (van der Cruisen, et al, 2018; van de Groep et al., 2021) (see Figure 1A). Note that analyses on the behavioral and neural correlates of self-concept in the control group (N = 40) were reported previously (van de Groep et al., 2021). During the task, participants were asked to (1) evaluate whether trait statements applied to themselves (self-condition, 40 trials) on a four-point scale, or to (2) categorize trait statements into four categories (prosocial, physical, academic and I don't know) (control-condition, 12 trials). For both conditions, trait sentences could have a positive or negative valence, and be from the physical appearance or prosocial domains, with an equal distribution of valence and domains amongst trials.

Both task conditions (self-condition, control-condition) were completed in separate runs and counterbalanced across participants. The order of trials and jitter timing were optimized for our design using Optseq2 (Dale, 1999), with jittered timing intervals varying between 0 and 4400 ms. Every trial started with a fixation cross (400 ms), followed by the display with (1) the trait description and (2) response options (4600 ms, see Figure 1A). Upon response, the chosen response was displayed in yellow for the remaining stimulus time. If a participant did not respond in the required timeframe, a message 'too late' was displayed for 1000 ms.

2.2.2 Youth Psychopathic Traits Inventory

Psychopathic traits were measured using the Youth Psychopathy Inventory (Andershed et al., 2002). This 50-item self-report questionnaire distinguishes three trait dimensions: Grandiose-Manipulative, Callous-Unemotional, and Impulsive-Irresponsible traits. Answers are scored on a 4-point Likert scale (1 = does not apply at all, to 4 = applies very well). For both subsamples (i.e., childhood arrestee cohort and control sample), the reliability of the total YPI score, GM and II subscales was good to excellent, and reliability of Callous-Unemotional traits was poor (see Table 1 and van de Groep et al., 2022).

2.2.3 Antisocial behavior

For the Childhood arrestee cohort, we determined whether individuals met the criteria of persistent antisocial behavior (see also van de Groep et al., 2022), which was true if they showed an early onset (i.e., were convicted for an index crime before the age of 12) and received a recent diagnosis of disruptive behavior disorder (DBD) or Antisocial Personality Disorder (ASPD) (at wave 4/5 of the longitudinal study; also see Table S3). DBD diagnoses at the previous waves were determined using the National Institute of Mental Health DISC-IV (Shaffer et al., 2000), while antisocial personality disorder at the current wave was determined by using the MINI (Lecrubier et al., 1997). Of the 54 participants who completed the self-concept task, 12 were classified as persister, and 42 as desister (see Table 1 and Table S3). One participant could not be classified due to incomplete MINI administration, and was therefore excluded from all sub-group analyses. As can be seen in Table 1, there were significant differences in psychopathic traits (total and sub-scale) scores between persisters, desisters, and controls, with post hoc tests revealing persisters scoring significantly higher on all traits compared to the other two groups, all t 's < -2.597 , all p 's $< .029$.

2.3 Behavioral Analyses

Behavioral and ROI data were analyzed using R (Version 4.0.1, R Core team, 2020). Assumptions were checked for all analyses. We identified several univariate outliers ($> 3SD$) in the psychopathic traits, for total and sub-scores (i.e., Total psychopathic traits: 1, Callous-Unemotional traits: 1, Grandiose-Manipulative traits: 3). However, given that these extreme values were valid scores, and removing them did not change any of the reported results, we retained them in the analyses we report below.

2.4 Neuroimaging methods

2.4.1 Neuroimaging Methods: MRI Data Acquisition.

For acquiring (functional) MRI data, we used a 3T Philips scanner (Philips Achieva TX, Erlangen, Germany) with a standard eight-channel whole-head coil. The self-concept task was projected on a screen, and viewed through a mirror on the head coil. Head movement was restricted by placing foam inserts inside the coil. Functional scans were acquired during two runs of 120 (self-condition) and 40 (control-condition) dynamics, using T2* echo-planar imaging (EPI). The volumes covered the entire brain (repetition time (TR)=2.2s; echo time (TE)=30ms; sequential acquisition, 38 slices; voxel size $2.75 \times 2.75 \times 2.75$ mm; field of view (FOV)= $220 \times 220 \times 115$ mm). Before the first functional scan of each run, five dummy scans were acquired. Prior to the self-concept task, we collected a high-resolution 3D T1 scan for anatomical reference (TR =7.6 ms, TE =3.5 ms, 140 slices, voxel size $1.1 \times 1.1 \times 1.1$ mm, mm, FOV = $250 \times 196 \times 170$ mm).

2.4.2 Neuroimaging Methods: Preprocessing.

Data were preprocessed and analyzed using SPM12 (Wellcome Department of Cognitive Neurology, London, United Kingdom). Images were corrected for slice timing acquisition and rigid body motion. We spatially normalized functional volumes to T1 templates, and performed spatial smoothing using a 6 mm full width at half maximum (FWHM) isotropic Gaussian kernel. Subsequently, all volumes were resampled to voxels of 3 mm³. Templates were based on the MNI305 stereotaxic space (Cocosco et al., 1997). To ensure quality control, functional images were visually checked before preprocessing, and following each pre-processing step.

2.4.3 Neuroimaging Methods: First level analyses.

To perform statistical analyses on individual subjects' fMRI data, we used the general linear model (GLM) in SPM12. We modeled the fMRI time series as a series of zero duration events a canonical hemodynamic response function (HRF), using 'Physical-Positive', 'Physical-Negative', 'Prosocial-Positive' and 'Prosocial-Negative' as regressors for the self-evaluation part of the task. For the control condition, we only modeled one event-related event: 'Control' (i.e., not divided into separate contrasts by valence or domain). Trials with no response were modeled separately as a regressor of no interest and were excluded from analyses (0.78% of trials). Six motion parameters were included as nuisance regressors. The pairwise comparisons resulted in participant-specific contrast images, which we subsequently submitted to second-level group analyses.

2.4.4 Neuroimaging Methods: Second level analyses.

To explore whole-brain neural responses to self-representation, we performed two analyses. First, to reveal regions that were specific for self-evaluations, we compared self-condition trials (collapsed across domains and valences) to control trials using a one sample t-test for the contrast Self > Control, and the reversed contrast. Second, to examine valence- and domain-specific neural activity, we performed a whole brain 2 (Valence: Positive vs. Negative) x 2 (Domain: Physical vs. Prosocial) ANOVA. All results were corrected using a primary voxel-wise threshold of $p < .001$, and coordinates for local maxima are reported in MNI space. Unthresholded statistical maps of all reported whole-brain analyses are available on Neurovault (Gorgolewski et al., 2015); see <https://neurovault.org/collections/DNPFQNK/>.

2.4.5 Neuroimaging Methods: Region-of-Interest analyses (ROIs).

The a-priori regions of interest (ROI) in which we test our main hypotheses were defined anatomically and based on previous research: medial PFC (coordinates: $x = -6, y = 50, z = 4$; based on a meta-analysis by Denny et al., 2012, cf. van de Group et al., 2021), dorsolateral PFC (coordinates left dlPFC: $x = -48, y = 35, z = 16$; coordinates right dlPFC: $x = 48, y = 35, z = 16$, based on van der Crujisen et al., 2017). All ROIs were created by extracting 10 mm

spheres around the specified coordinates. For all ROIs, we applied Bonferroni correction for correlated variables with a threshold of $\alpha = 0.011$ (Perneger, 1998).

3. RESULTS

3.1 Behavioral Results

3.1.1 Group-based differences in domain and valence-specific self-evaluations

A mixed-measures ANOVA (Valence (positive vs. negative) x Domain (prosocial vs. physical), and Group (Persister vs. Desister vs. Control) on self-ratings revealed a main effect of Valence ($F(1, 90) = 362.41, p < .001, \eta^2 = 0.801$) and Domain ($F(1, 90) = 42.12, p < .001, \eta^2 = 0.319$). Participants rated positive items and prosocial traits as more applicable (see Figure 1B). There was no significant interaction of Group x Valence ($F(2, 90) = 0.30, p = .743, \eta^2 = 0.007$, (disconfirming hypothesis 1a) nor of Valence x Group x Domain ($F(2, 90) = 2.73, p = .071, \eta^2 = 0.057$) (disconfirming hypothesis 1b). Note that adding Age, Sex, IQ, and Education as covariates did not change the results (see Supplementary Materials Appendix B).

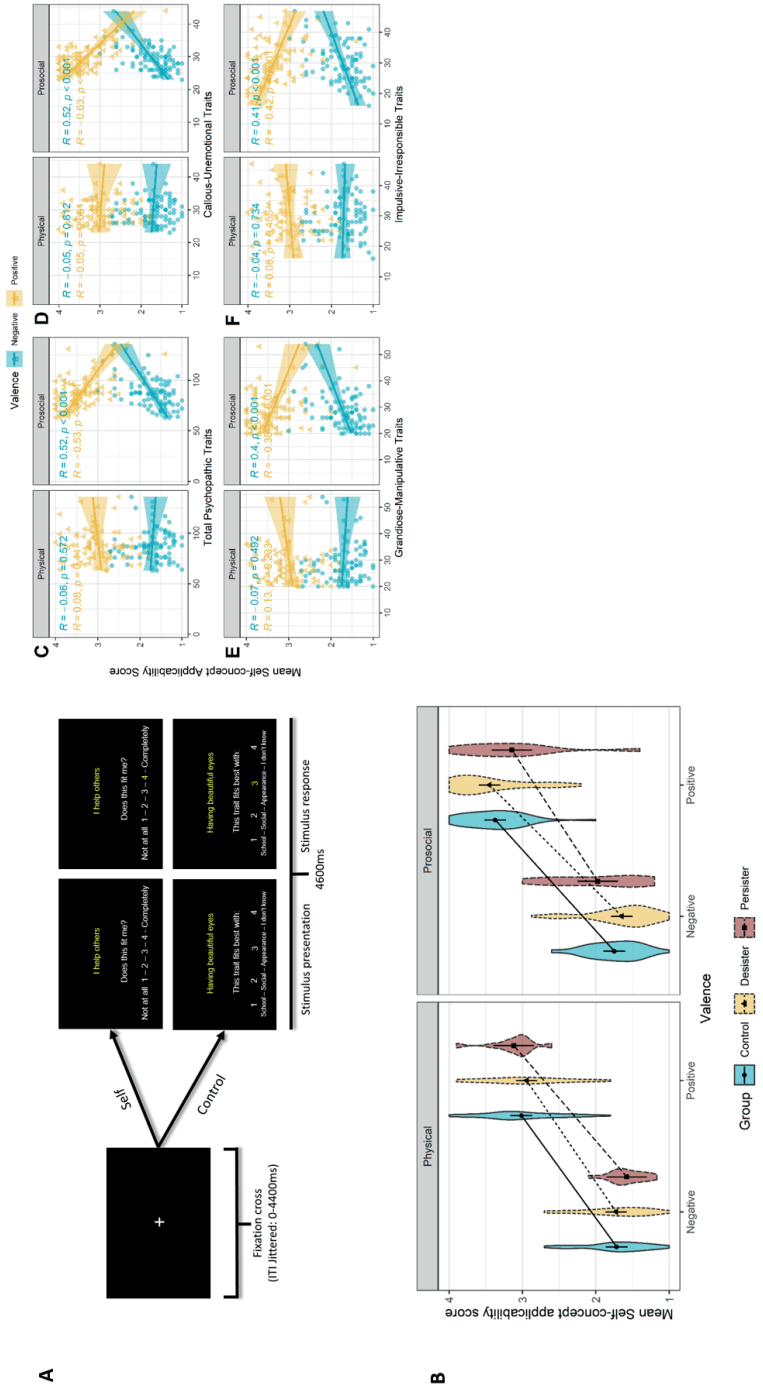


Figure 1 (A) Schematic depiction of the Self-concept Task. During the task, participants evaluate the applicability of positive and negative trait statements in two domains (prosocial and physical) on a four-point scale (self-condition), or categorize trait statements into four categories (prosocial, physical, academic and I don't know) (control-condition). (B) In general, positive traits were rated as more applicable than negative traits, and prosocial traits as more applicable than physical traits. (C) Correlations between positive and negative self-concept endorsement ratings and total psychopathic trait scores in the physical (left panel) and prosocial domain (right panel). (D-F) Correlations between positive and negative endorsement ratings self-concept and (sub dimensions of) psychopathic trait scores (panel D: Callous-Unemotional Traits, panel E: Grandiose-Manipulative Traits, panel F: Impulsive-Irresponsible Traits), in the prosocial domain and physical domain.

3.1.2 Individual differences in domain and valence-specific self-evaluations

Next, we tested the pre-registered hypothesis (3a) that total score and sub-dimensions of the YPI would be (differentially) associated with the endorsement of positive and negative traits. The associations for the separate trait dimensions were all in the same direction, showing a positive relationship between the psychopathic trait scores for Callous-Unemotional, Grandiose-Manipulative and Impulsive-Irresponsible traits, and the applicability of negatively valenced self-traits; and a negative relationship with the applicability of positively valenced self-traits (see Figure 1D-F, and supplementary material Appendix B). Additionally, higher total psychopathic traits scores were associated with increased endorsement of negative trait statements, and decreased endorsement of positive trait statements, but only for evaluations in the prosocial (negative: $r(91) = .51, p < .001$, positive: $r(91) = -.54, p < .001$) and not the physical domain (negative: $r(91) = -.06, p = .054$, positive: $r(91) = .09, p = .396$) (Figure 1C). In line with these observations, a mixed measures ANOVA on average applicability scores revealed a significant interaction between Domain, Valence and YPI total scores ($F(1, 91) = 24.58, p < .001, \eta_p^2 = 0.21$).

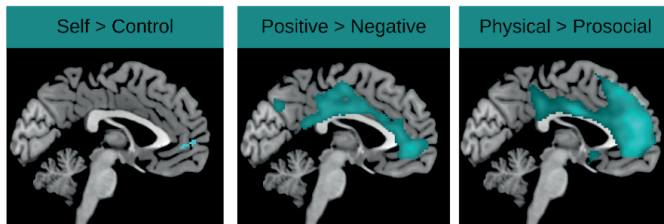


Figure 2. Whole brain t-contrasts conducted at the group level for the contrasts Self > Control, Positive > Negative and Physical > Prosocial.

3.2 fMRI results

3.2.1 Self-evaluative, domain- and valence-specific neural activation

To examine which neural activation was specific for self-evaluations, we examined the following contrasts within a whole brain t-test (see Table S4, Figure 2 and Appendix A for an overview of the results).

3.2.2 Group differences in task-, valence- and domain-specific neural activation in the mPFC

To test whether individuals who persisted in antisocial behavior showed less neural activity in the mPFC (compared to desister and control groups) (hypothesis 2a), we conducted two mixed measure ANOVAs using Group as between-subjects factor, and Condition (for the contrast positive self > control) and Valence (positive vs. negative) as within-subjects factors, respectively. There was no interaction effect between Group and Condition, $F(2, 89) = 0.09, p = .918$ (see Figure S1) (disconfirming hypothesis 2a), nor

between Group and Valence, $F(2, 89) = 1.81, p = .170, \eta_p^2 = .041$ for the contrast positive > negative valence, suggesting no group differences in valence-related activity patterns of mPFC activity (see Figure S2).

Next, we examined whether evaluating prosocial traits resulted in increased mPFC activity relative to physical traits, and whether these patterns differed between groups (hypothesis 2b-2) While there was indeed a main effect of Domain, the direction of this effect was opposite to our hypothesis, showing that evaluating physical statements evoked increased activity in the mPFC, $F(1, 89) = 142.64, p < .001, \eta_p^2 = 0.616$ (see Figure S2C). Also contradictory to our hypotheses, there was no Group x Domain interaction (hypothesis 2b-2 and 2b-3) ($F(1, 89) = 0.70, p = .51$). Hence, we found no evidence (hypothesis 2b-2), for less neural activity in the mPFC in individuals who persisted in antisocial behavior when evaluating prosocial statements, compared to when they evaluated physical trait statements, $t(86) = 2.69, p = .088$. Note that adding Age, Sex, IQ, and Education as covariates did not change the results (see Supplementary Materials Appendix B).

3.2.3 Group differences in domain-specific neural activation in the dlPFC

Our next aim was to investigate whether we could replicate prior findings showing that evaluating physical traits (vs. prosocial traits) results in stronger activity in the dlPFC, whereas evaluating prosocial traits (vs. physical traits) results in increased mPFC activity, across all participants (hypothesis 2b-1). First, we conducted two mixed measure ANOVAs using the activity extracted from the left and right dlPFC ROI as a dependent variables and Valence, Group and Domain as independent variables. In line with hypothesis 2b, evaluating physical traits resulted in increased activity in the left dlPFC, $F(1, 89) = 114.20, p < .001, \eta_p^2 = 0.562$ (see Figure S2), and right dlPFC, $F(1, 89) = 12.65, p < .001, \eta_p^2 = 0.128$, compared to prosocial traits (see Figure S2). These analyses did not reveal main effects of Group or interaction effects between Group and Conditions or Valence, all p 's > 0.178. Note that adding Age, Sex, IQ, and Education as covariates did not render any of these effects significant.

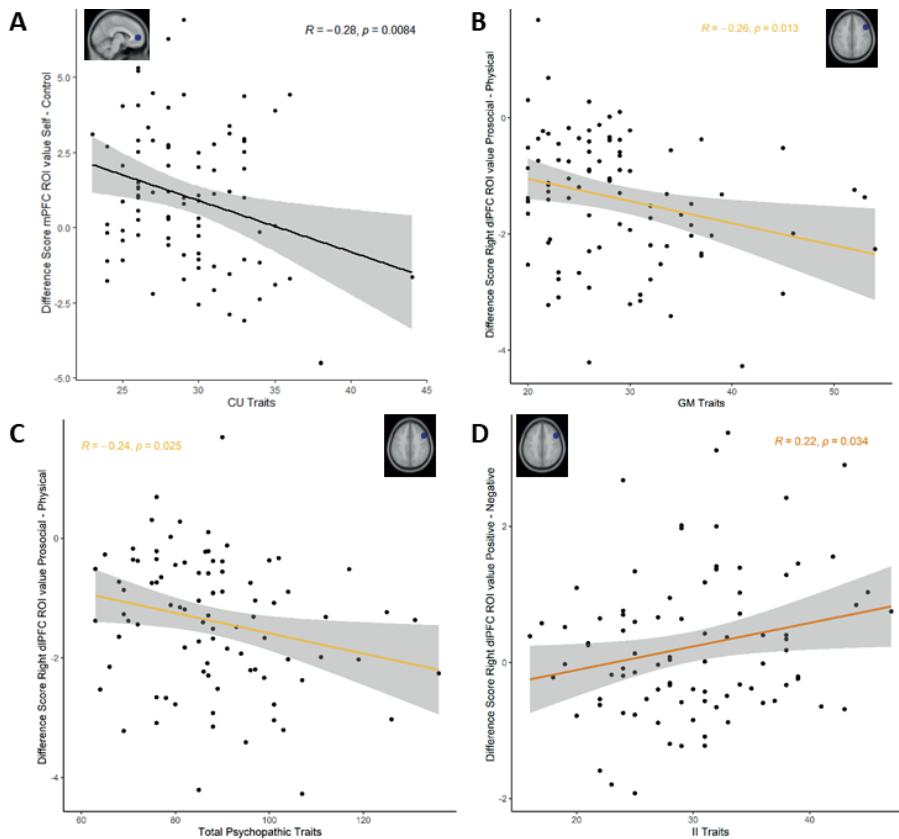


Figure 3. (A-D) Correlations between difference scores in mPFC and right dlPFC activity and (sub-dimensions of) psychopathic traits. **(A)** The mPFC activity difference score between the contrast Self > Control and Callous-Unemotional traits was negatively correlated. **(B, C)** The dlPFC activity difference score between the contrast Prosocial > Physical and GM and total psychopathic traits, respectively, showed a negative association. **(D)** The dlPFC activity difference score between the contrast Positive > Negative and Impulsive-Irresponsible traits was positively associated.

3.2.4 Associations between the neural basis of self-concept and psychopathic traits

3.2.4.1 mPFC

Our next aim was to test whether specific sub-dimensions of the YPI would be differentially associated with mPFC activity (hypothesis 3b), using the difference scores in ROI parameter estimates between (1) self > control, (2) positive self > control, (3) positive self > negative self and (4) prosocial > physical. As can be seen in Figure 2B, Callous-Unemotional traits were negatively associated with the difference score in ROI parameter estimates between

self > control, $r(90) = -.28, p = .0084$, suggesting that higher Callous-Unemotional traits were associated with decreased mPFC activity during the self-condition, compared to the control condition. In line with this observation, a repeated measures ANCOVA with Condition (Self. vs Control) and Callous-Unemotional traits as covariate, revealed a significant interaction effect between Condition and Callous-Unemotional traits, $F(1, 86) = 7.88, p = .006, \eta_p^2 = 0.046$. None of the remaining associations with YPI sub-scales were significant.

3.2.4.2 dlPFC

Next, we explored whether specific sub-dimensions of the YPI would be differentially associated with left and right dlPFC activity (hypothesis 3b), using the difference scores in ROI parameter estimates between (1) self > control, (2) positive self > control, (3) positive self > negative self and (4) prosocial > physical. For the left dlPFC, none of the associations were significant.

For the right dlPFC, Grandiose-Manipulative traits were negatively associated with the difference score in ROI parameter estimates between prosocial > physical, $r(90) = -.26, p = .013$ (see Figure 2C), suggesting that higher Grandiose-Manipulative traits were associated with decreased dlPFC activity when making evaluations in the prosocial domain, compared to the physical domain, which was confirmed by a ANCOVA, showing an interaction effect between Grandiose-Manipulative traits and Domain, $F(1,86) = 5.91, p = .017, \eta_p^2 = .009$. Likewise, total psychopathic trait scores were negatively associated with the difference score in ROI parameter estimates between prosocial > physical, $r(90) = -.24, p = .025$ (see Figure 2D). In line with this observation, an ANCOVA revealed an interaction effect between Total psychopathic traits and Domain, $F(1,86) = 4.53, p = .036, \eta_p^2 = .036$. Finally, Impulsive-Irresponsible traits were positively associated with the difference score in ROI parameter estimates between positive > negative, $r(90) = .22, p = .034$ (see Figure 2E), showing that higher Impulsive-Irresponsible traits were associated with increased dlPFC activity when evaluating positive, rather than negative trait statements. However, a follow-up ANCOVA showed no significant interaction effect between Impulsive-Irresponsible traits and Valence, $F(1,86) = 1.64, p = .204$. The dlPFC results did not survive Bonferroni correction ($p < .011$). In addition, none of the remaining associations with YPI sub- or total scores were significant.

4. DISCUSSION

In the current study, we examined whether young adults with diverging developmental trajectories of antisocial behavior differed in the endorsement of positive and negative self-evaluations, and corresponding neural responses. To this end, participants completed a self-evaluation task, rating trait statements that varied in valence (positive vs. negative) and domain (prosocial vs. physical). To account for heterogeneity in antisocial behavior,

we additionally examined the role of different sub-dimensions of psychopathic traits in self-evaluation and associated neural correlates.

Five findings stand out. First, we replicated findings from earlier studies showing that (1) people find positive self-traits more applicable than negative traits, and prosocial traits more applicable than physical traits (van de Groep et al., 2021; van der Cruijssen et al., 2017; 2018); (2) that evaluating self-traits in general, and positive traits specifically, results in increased activity in cortical midline structures, such as the mPFC (vs. negative traits) and (3) that evaluations in the physical (vs. prosocial) domain result in increased bilateral dlPFC activity (Denny et al., 2012; Moran et al., 2010; van der Cruijssen et al., 2017; 2018; Lieberman et al., 2019). Unexpectedly, however, our results revealed no significant differences in self-evaluations between groups characterized by persistent, desistant or no prior antisocial behavior, nor in the neural underpinnings of such behavior. However, we found that higher levels of total psychopathic trait scores were associated with increased endorsement of negative self-evaluations, and decreased endorsement of positive self-evaluations in the prosocial, but not in the physical domain. Finally, we found preliminary evidence that (sub-dimensions of) psychopathic traits might be differentially associated with mPFC and right dlPFC activity during self-evaluations. More specifically, Callous-Unemotional traits were negatively associated with self-related mPFC activity, and total psychopathic traits were positively associated with domain-specific dlPFC activity, which was mainly driven by Grandiose manipulative traits.

4.1 Psychopathic traits and prosocial self-concept positivity

Developing a positive view of the self is an important developmental milestone, which promotes mental wellbeing and effective social relationships (Crone et al., 2022; Ybrant, 2008; Jankowski & Bağ, 2020). In our study, participants generally showed positive self-evaluations (Denny et al., 2012; van der Cruijssen et al., 2017; 2018; Lieberman et al., 2019). However, contrary to our expectation, we did not observe differences in behavioral or neural responses between the persisting, desisting and control subgroups. This was surprising, given that repeated negative social interactions – which are often observed in individuals with persistent antisocial behavior (Paternoster et al., 2009; 2016) - are thought to shape one's self-concept as more negative (Harter, 2012). However, we found that higher levels of psychopathic traits were associated with a more negative, and less positive self-concept for prosocial, but not physical trait statements. Total level and subscale levels of psychopathic traits were highest in participants with a persistent developmental trajectory (total and subscales; see Figure S4). Our finding that people high (vs. low) on psychopathic traits show less valence-specific differentiation in the applicability of prosocial traits fits with the idea that self-concept may be affected in some, but not all domains in people with externalizing tendencies (Kita & Inoue, 2017). Taken together, while antisocial and criminal behavior might indeed be reflected in how young adults evaluate their prosocial traits (Paternoster et al., 2009; 2016), individual differences in

psychopathic traits better capture the complex association between antisocial histories and self-concept evaluation than group comparisons (Northoff & Tsumati, 2019).

Why do individuals high on psychopathic traits have a more negative prosocial self-concept? A positive and accurate self-concept requires a balance between internal representations of the self, and external input (e.g., social feedback from peers), and flexibility in the ability to focus attention towards the self facilitates self-regulation in social situations (Jankowski & Bąk, 2020). Previous research suggests that too much or too little self-focus is common in various internalizing and externalizing problems characterized by a negative self-concept (Northoff & Tsumati, 2019; Philippi & Koenigs, 2014; Zhao et al., 2013), such as depression (Davey & Harrison, 2022) and ADHD (Kita & Inoue, 2017). In depression and anxiety disorders, a negative self-concept is likely the result of too much self-focus and ruminative thoughts (Northoff & Tsumati, 2019; Philippi & Koenigs, 2014). Psychopathy, on the other hand, has been associated with an extreme external focus and limited self-focus and self-reflection (Doerfler et al., 2021; Philippi & Koenigs, 2014). Possibly, individuals high on psychopathic traits spend little time reflecting on their prosocial traits, which may prevent normative developmental processes that bias self-knowledge towards positivity (Jankowski & Bąk, 2020). Moreover, it should be noted that while a negative and unstable sense of self is generally thought to be maladaptive, in the case of psychopathy, it may arise and be adaptive in quickly changing social environments to facilitate goals associated with a fast life strategy (Doerfler et al., 2021). Hence, future studies should further explore how self-concept valence relates to antisocial tendencies in different social contexts.

4.2 The neural correlates of self-concept appraisal and psychopathic trait subdimensions

Although different psychopathy dimensions are inter-related, they often result in different outcomes, and may differentially contribute to the etiology and maintenance of antisocial behavior (McCuish et al., 2014; Salihovic et al., 2014; Zalk & Zalk, 2015; Miglin et al., 2021). Hence, we expected that different psychopathic trait dimensions (Callous-Unemotional traits, Grandiose-Manipulative traits and Impulsive-Irresponsible traits) would be differently associated with self-concept valence. While some associations we found were in the hypothesized direction, all three sub-dimensions showed similar behavioral patterns (i.e., showing positive associations with the endorsement of negative, but not positive self-traits), which was echoed in the total psychopathic traits scores.

However, the different sub-dimensions were related to differences in neural functioning during self-appraisal. More specifically, higher Callous-Unemotional traits were associated with decreased self-related mPFC activity (regardless of domain), which may reflect diminished self-relevance or personal value (D'Argembeau et al., 2013, van der Crujssen et al., in preparation). This finding fits with an earlier Electroencephalogram (EEG) study on self-referential processing, showing attenuated mPFC activity in individuals high

on Callous-Unemotional traits (Bontemps, 2022). Possibly, individuals with elevated Callous-Unemotional traits have difficulty constraining their abstract self-referential schemas toward personally significant information (i.e., have less self-focused thoughts, Zamani et al., 2022, Northoff & Tumati, 2019) – which results in a more negative self-concept. Alternatively, decreased personal relevance in psychopathy might arise from difficulties in identifying, describing and retrieving feelings about the self (Sifneos, 1973). Indeed, alexithymia – a condition characterized by difficulties in the experience, verbalization, identification and regulation of emotions (Larsen et al., 2013), which is closely related to psychopathy (Dawel et al., 2012) and specifically Callous-unemotional traits, (Cecil et al., 2018; Dadds et al., 2018; Huffman & Oshri, 2022) – is possibly also associated with decreased mPFC activity during self-evaluations. Interestingly, previous studies show that alexithymia symptoms are more pronounced in youth high on Callous-Unemotional traits with comorbid internalizing symptoms (Cecil et al., 2018), while people with mere internalizing problems tend to over-constrain, rather than under-constrain, their attention towards self-thoughts (Davey & Harrison, 2022, Northoff & Tumati, 2019). In our study, we also observed comorbid internalizing problems during development, particularly in the group where Callous-Unemotional traits were the highest (i.e., the persister group, see Table S2). As such, future research is warranted to further explore different etiological mechanisms that give rise to a negative self-concept, and its association with different types of (comorbid) psychopathology and maladjusted social behavior (Lee & Stone, 2012; Schettini et al., 2021; Northoff & Tumati, 2019).

We also found that higher psychopathic traits, and particularly Grandiose-Manipulative traits, were associated with increased domain-specific right dlPFC activity in the physical domain (vs. prosocial domain). Increased dlPFC activity has often been found during trait evaluations of physical appearance (van der Crujisen et al., 2017; Moran et al., 2010), and might reflect preferential selection, retrieval and/or maintenance of physical (rather than prosocial) information during self-appraisals (Balconi, 2013; Curtis & D’Esposito, 2003; Meshi et al., 2016; van der Crujisen et al., 2017). Moreover, physical traits may require and recruit more image-based visualizations or retrieval, which has been attributed to the right dlPFC (Balconi, 2013). Alternatively, in individuals high on psychopathic traits, physical traits may be more readily available than prosocial traits, and thus require more self-focused inhibition to constrain their thoughts (Lemogne et al., 2009). These domain-related difficulties might be the result of disruptions in higher-order emotional processes, such as empathy and Theory of Mind (Brüne & Brüne-Cohrs, 2006; Bird & Viding, 2014). Future studies should examine these possibilities in more detail.

4.3 Limitations and future directions

The current study has several limitations. First, while we managed to include a relatively large proportion of individuals who showed persistent antisocial behavior (i.e., 20% of the childhood arrestee cohort participants who completed the final time point), the size of

this group was nevertheless small (see Table 1), which may have limited our statistical power to detect significant differences between groups. Moreover, low power also decreases the likelihood that positive findings are true positives (Button et al., 2013), which should be considered when interpreting the results of the current study. Second, reliability of the Callous-Unemotional traits subscale was low, mainly in the control group, which fits with earlier reports that these traits are difficult to capture reliably (Salekin et al., 2018, see also van de Groep et al., 2022). Hence, results involving these traits should be interpreted with this limitation in mind. Third, contrary to our expectation, associations for the separate trait dimensions were all in the same direction, which hints at the possibility that they are not differentially associated with self-evaluations. Fourth, we found an association between self-appraisals and psychopathic traits, but no group differences, which can be partly explained by the small persister group size and observed descriptive differences between groups (e.g., in IQ). However, this discrepancy might also be related to closer conceptual and methodological (i.e., common-method variance) overlap between self-concept appraisals and self-report of psychopathic traits, relative to a diagnostic assessment of antisocial personality disorder or criminal offenses. Finally, our analyses on the associations between psychopathic traits and dlPFC activity specifically were not sufficiently powered to survive corrections for multiple testing. Hence, future research is needed to determine whether our findings would replicate in larger samples.

4.4. Conclusion

In conclusion, this study shows that individual differences in psychopathic traits, rather than group-differences in the presence of prior or continuous antisocial behavior, are important in determining how positively people appraise their prosocial self-concept. Moreover, we show preliminary evidence that individual differences in psychopathic traits are accompanied by different levels of neural activity in the mPFC and dlPFC during self-appraisal, which hints at the possibility that separable neural mechanisms underlie how people with psychopathic tendencies appraise their cognitive self-knowledge and constrain information during this process. Hence, our findings provide important starting points to understand why and how self-concept and identity play a role in desistance from crime and antisocial behavior.

Table 1 Sample description, group comparisons and reliability estimates

Measure	Group			Statistical comparison
	Desister (n = 42)	Persister (n = 12)	Control (n = 40)	
Sex [n males/females]	36/6	11/1	17/23	
Age [M (SD)]	26.20 (1.63)	26.62 (1.13)	22.7 (3.09)	$F(2,91) = 27.77, p < .001^a$
Education [n]				$\chi^2 = 13.128, p = 0.04$
Vocational	25	6	17	
College	9	2	18	
University	4	1	4	
Other	4	3	1	
Ethnicity				$\chi^2 = 7.18, p = 0.71$
Dutch	37	10	36	
Surinamese	0	1	2	
Turkish	1	0	0	
Moroccan	1	0	0	
Different Western	1	1	1	
Different Non-Western	2	0	1	
IQ [M (SD)] ¹	103.47 (14.13)	100.87 (11.08)	107.35 (11.52)	$F(2,91) = 4.12, p = .009$
YPI Total psychopathic traits [M (SD)] ²	84.91 (16.37)	106.67 (19.51)	87.78 (9.68)	$F(2,91) = 10.89, p < .001^b$
YPI Callous-Unemotional Traits [M (SD)]	28.77 (3.64)	33.5 (5.03)	29.20 (3.05)	$F(2,91) = 8.28, p < .001^b$
YPI Grandiose-Manipulative Traits [M (SD)]	27.53 (7.47)	37.33 (11.16)	28.38 (4.96)	$F(2,91) = 9.22, p < .001^b$
YPI Impulsive-Irresponsible Traits [M (SD)]	28.59 (7.53)	35.83 (7.19)	30.12 (5.66)	$F(2,91) = 5.37, p = .006^b$

Cronbach's Alpha	Group	
	Childhood Arrestee Cohort	Control Group
YPI Total	0.92	0.81
YPI Callous-Unemotional Traits	0.52	0.39 ³
YPI Grandiose-Manipulative Traits	0.91	0.78
YPI Impulsive-Irresponsible Traits	0.86	0.77

Note: IQ, estimated IQ based on two subscales of the Wechsler Adult Intelligence Scale-IV (Similarities and Block Design), YPI = Youth Psychopathic traits Inventory.

^a Significant differences between controls and desisters, and controls and persisters

^b Significant differences between persisters and desisters, and persisters and controls

¹ Note that for three participants who completed the fMRI session, the IQ tests at T5 were not completed. Therefore, we estimated these scores using multiple imputation based on the other variables reported in this table, as well as prior IQ scores (T4).

² Note that for two participants who completed the fMRI session, the YPI was not completed. Therefore, we estimated these scores using multiple imputation, based on the other variables reported in this table, as well as prior IQ scores (T4).

³ Note that we did use and report a measure with a reliability lower than .5 (for CU traits in the control group) contrary to our pre-registered plans, given the widespread use of this measure in psychopathy research. Similar to earlier reports, we note the limitation of this approach in the discussion and acknowledge that these traits are difficult to capture reliably (although we did manage to do so for the childhood arrestee cohort and total YPI scores).

REFERENCES

1. Alegria, A. A., Radua, J., & Rubia, K. (2016). Meta-Analysis of fMRI Studies of Disruptive Behavior Disorders. *American Journal of Psychiatry*, 173(11), 1119–1130. <https://doi.org/10.1176/appi.ajp.2016.15081089>
2. Andershed, H. A., Kerr, M., Stattin, H., & Levander, S. (2002). *Psychopathic traits in non-referred youths: A new assessment tool* (pp. 131–158). Elsevier. <http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-6748>
3. Balconi, M. (2013). Dorsolateral prefrontal cortex, working memory and episodic memory processes: Insight through transcranial magnetic stimulation techniques. *Neuroscience Bulletin*, 29(3), 381–389. <https://doi.org/10.1007/s12264-013-1309-z>
4. Barry, T. D., Thompson, A., Barry, C. T., Lochman, J. E., Adler, K., & Hill, K. (2007). The importance of narcissism in predicting proactive and reactive aggression in moderately to highly aggressive children. *Aggressive Behavior*, 33(3), 185–197. <https://doi.org/10.1002/ab.20198>
5. Boccardi, M., Frisoni, G. B., Hare, R. D., Cavedo, E., Najt, P., Pievani, M., Rasser, P. E., Laakso, M. P., Aronen, H. J., Repo-Tiihonen, E., Vaurio, O., Thompson, P. M., & Tiihonen, J. (2011). Cortex and amygdala morphology in psychopathy. *Psychiatry Research: Neuroimaging*, 193(2), 85–92. <https://doi.org/10.1016/j.pscychresns.2010.12.013>
6. Bontemps, A. P., Batky, B. D., Houser, R. A., & Salekin, R. T. (2022). Psychopathic Traits, Conduct Problems, and the Examination of Self-Referential Processing Using EEG in Incarcerated Adolescents. *Journal of Psychopathology and Behavioral Assessment*. <https://doi.org/10/gn63k9>
7. Broulidakis, M. J., Fairchild, G., Sully, K., Blumensath, T., Darekar, A., & Sonuga-Barke, E. J. S. (2016). Reduced Default Mode Connectivity in Adolescents With Conduct Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(9), 800–808.e1. <https://doi.org/10.1016/j.jaac.2016.05.021>
8. Button, K. S., Ioannidis, J., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
9. Caldwell, B. M., Anderson, N. E., Harenski, K. A., Sitney, M. H., Caldwell, M. F., Van Rybroek, G. J., & Kiehl, K. A. (2019). The structural brain correlates of callous-unemotional traits in incarcerated male adolescents. *NeuroImage: Clinical*, 22, 101703. <https://doi.org/10.1016/j.nicl.2019.101703>
10. Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavalley, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, 70(1), 141–156. <https://doi.org/10.1037/0022-3514.70.1.141>
11. Cecil, C. A. M., McCrory, E. J., Barker, E. D., Guiney, J., & Viding, E. (2018). Characterising youth with callous–unemotional traits and concurrent anxiety: Evidence for a high-risk clinical group. *European Child & Adolescent Psychiatry*, 27(7), 885–898. <https://doi.org/10.1007/s00787-017-1086-8>
12. Crone, E.A., Green, K.G., van de Groep, I.H., van der Cruijssen, R. (2022). A Neurocognitive Model of Self-Concept Development in Adolescence. *Annual Review of Developmental Psychology*. <https://doi.org/10.1146/annurev-devpsych-120920-023842>
13. Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415–423. [https://doi.org/10.1016/S1364-6613\(03\)00197-9](https://doi.org/10.1016/S1364-6613(03)00197-9)
14. Dadds, M., Kimonis, E., Schollar-Root, O., Moul, C., & Hawes, D. (2018). Are impairments in emotion recognition a core feature of callous–unemotional traits? Testing the primary versus secondary

- variants model in children. *Development and Psychopathology*, 30(1), 67-77. doi:10.1017/S0954579417000475
15. Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8(2-3), 109-114. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:2<3<109::AID-HBM7>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0193(1999)8:2<3<109::AID-HBM7>3.0.CO;2-W)
 16. Dalwani, M. S., Tregellas, J. R., Andrews-Hanna, J. R., Mikulich-Gilbertson, S. K., Raymond, K. M., Banich, M. T., Crowley, T. J., & Sakai, J. T. (2014). Default mode network activity in male adolescents with conduct and substance use disorder. *Drug and Alcohol Dependence*, 134, 242-250. <https://doi.org/10.1016/j.drugalcdep.2013.10.009>
 17. D'Argembeau, A. (2013). On the Role of the Ventromedial Prefrontal Cortex in Self-Processing: The Valuation Hypothesis. *Frontiers in Human Neuroscience*, 7. <https://www.frontiersin.org/article/10.3389/fnhum.2013.00372>
 18. Davey, C. G., & Harrison, B. J. (2022). The self on its axis: A framework for understanding depression. *Translational Psychiatry*, 12(1), 1-9. <https://doi.org/10/gn6p2h>
 19. Dawel, A., O'Kearney, R., McKone, E., & Palermo, R. (2012). Not just fear and sadness: meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. *Neuroscience and biobehavioral reviews*, 36(10), 2288-2304. <https://doi-org.ezproxy.leidenuniv.nl/10.1016/j.neubiorev.2012.08.006>
 20. De Brito, S. A., Mechelli, A., Wilke, M., Laurens, K. R., Jones, A. P., Barker, G. J., Hodgins, S., & Viding, E. (2009). Size matters: Increased grey matter in boys with conduct problems and callous-unemotional traits. *Brain*, 132(4), 843-852. <https://doi.org/10.1093/brain/awp011>
 21. de Oliveira-Souza, R., Hare, R. D., Bramati, I. E., Garrido, G. J., Azevedo Ignácio, F., Tovar-Moll, F., & Moll, J. (2008). Psychopathy as a disorder of the moral brain: Fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry. *NeuroImage*, 40(3), 1202-1213. <https://doi.org/10.1016/j.neuroimage.2007.12.054>
 22. Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A Meta-analysis of Functional Neuroimaging Studies of Self- and Other Judgments Reveals a Spatial Gradient for Mentalizing in Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, 24(8), 1742-1752. https://doi.org/10.1162/jocn_a_00233
 23. Doerfler, S. M., Tajmiriyahi, M., Ickes, W., & Jonason, P. K. (2021). The self-concepts of people with Dark Triad traits tend to be weaker, less clearly defined, and more state-related. *Personality and Individual Differences*, 180, 110977. <https://doi.org/10/gjzxr Psychological Science>, 16(4), 328-335. <https://doi.org/10.1111/j.0956-7976.2005.01535.x>
 24. Ermer, E., Cope, L. M., Nyalakanti, P. K., Calhoun, V. D., & Kiehl, K. A. (2012). Aberrant paralimbic gray matter in criminal psychopathy. *Journal of Abnormal Psychology*, 121(3), 649-658. <https://doi.org/10.1037/a0026371>
 25. Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., Freitag, C. M., & De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers*, 5(1), 1-25. <https://doi.org/10.1038/s41572-019-0095-y>
 26. Falquez, R., Couto, B., Ibanez, A., Freitag, M. T., Berger, M., Arens, E. A., Lang, S., & Barnow, S. (2014). Detaching from the negative by reappraisal: The role of right superior frontal gyrus (BA9/32). *Frontiers in Behavioral Neuroscience*, 8. <https://www.frontiersin.org/article/10.3389/fnbeh.2014.00165>
 27. Fanti, K. A., Kyranides, M. N., Petridou, M., Demetriou, C. A., & Georgiou, G. (2018). Neurophysiological markers associated with heterogeneity in conduct problems, callous unemotional traits, and

- anxiety: Comparing children to young adults. *Developmental Psychology*, 54(9), 1634–1649. <https://doi.org/10.1037/dev0000505>
28. Garvey, M., Avenevoli, S., & Anderson, K. (2016). The National Institute of Mental Health Research Domain Criteria and Clinical Research in Child and Adolescent Psychiatry. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(2), 93–98. <https://doi.org/10.1016/j.jaac.2015.11.002>
 29. Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00008>
 30. Gudjonsson, G. H., & Roberts, J. C. (1983). Guilt and self-concept in 'secondary psychopaths'. *Personality and Individual Differences*, 4(1), 65–70. [https://doi.org/10.1016/0191-8869\(83\)90053-3](https://doi.org/10.1016/0191-8869(83)90053-3)
 31. Hart, W., Richardson, K., & Breeden, C. (2019). An interactive model of narcissism, self-esteem, and provocation extent on aggression. *Personality and Individual Differences*, 145, 112–118. <https://doi.org/10.1016/j.paid.2019.03.032>
 32. Herpertz, S. C., Bertsch, K., & Jeung, H. (2018). Neurobiology of Criterion A: Self and interpersonal personality functioning. *Current Opinion in Psychology*, 21, 23–27. <https://doi.org/10.1016/j.copsyc.2017.08.032>
 33. Hoppenbrouwers, S. S., De Jesus, D. R., Stirpe, T., Fitzgerald, P. B., Voineskos, A. N., Schutter, D. J. L. G., & Daskalakis, Z. J. (2013). Inhibitory deficits in the dorsolateral prefrontal cortex in psychopathic offenders. *Cortex*, 49(5), 1377–1385. <https://doi.org/10.1016/j.cortex.2012.06.003>
 34. Hsieh, M., & Stright, A. D. (2012). Adolescents' Emotion Regulation Strategies, Self-Concept, and Internalizing Problems. *The Journal of Early Adolescence*, 32(6), 876–901. <https://doi.org/10.1177/0272431611433848>
 35. Huffman, L. G., & Oshri, A. (2022). Neural signatures of the development of antisocial behaviors & callous-unemotional traits among youth: the moderating role of parental support. *International Journal of Developmental Neuroscience*, 82(3), 205–221. doi: <https://doi-org.ezproxy.leidenuniv.nl/10.1002/jdn.10173>
 36. Johanson, M., Vaurio, O., Tiihonen, J., & Lähteenvuo, M. (2020). A Systematic Literature Review of Neuroimaging of Psychopathic Traits. *Frontiers in Psychiatry*, 10. <https://doi.org/10.3389/fpsy.2019.01027>
 37. Kita, Y., & Inoue, Y. (2017). The Direct/Indirect Association of ADHD/ODD Symptoms with Self-esteem, Self-perception, and Depression in Early Adolescents. *Frontiers in Psychiatry*, 8. <https://doi.org/10.3389/fpsy.2017.00137>
 38. Koenigs, M. (2012). The role of prefrontal cortex in psychopathy. *Reviews in the Neurosciences*, 23(3), 253–262. <https://doi.org/10.1515/revneuro-2012-0036>
 39. Lecrubier, Y., Sheehan, D., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J., & Dunbar, G. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CID-I. *European Psychiatry*, 12(5), 224–231. [https://doi.org/10.1016/S0924-9338\(97\)83296-8](https://doi.org/10.1016/S0924-9338(97)83296-8)
 40. Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, 99, 311–328. <https://doi.org/10.1016/j.neubiorev.2018.12.021>

41. Marsh, A. A., Finger, E. C., Mitchell, D. G. V., Reid, M. E., Sims, C., Kosson, D. S., Towbin, K. E., Leibenluft, E., Pine, D. S., & Blair, R. J. r. (2008). Reduced Amygdala Response to Fearful Expressions in Children and Adolescents With Callous-Unemotional Traits and Disruptive Behavior Disorders. *American Journal of Psychiatry*, 165(6), 712–720. <https://doi.org/10.1176/appi.ajp.2007.07071145>
42. Marsh, H., Parada, R., Yeung, A., & Healey, J. (2001). Aggressive school troublemakers and victims: A longitudinal model examining the pivotal role of self-concept. *Journal of Educational Psychology*, 93, 411–419. <https://doi.org/10.1037/0022-0663.93.2.411>
43. McCuish, E. C., Corrado, R., Lussier, P., & Hart, S. D. (2014). Psychopathic traits and offending trajectories from early adolescence to adulthood. *Journal of Criminal Justice*, 42(1), 66–76. <https://doi.org/10.1016/j.jcrimjus.2013.12.004>
44. Meshi, D., Mamerow, L., Kirilina, E., Morawetz, C., Margulies, D. S., & Heekeren, H. R. (2016). Sharing self-related information is associated with intrinsic functional connectivity of cortical midline brain regions. *Scientific Reports*, 6(1), 22491. <https://doi.org/10.1038/srep22491>
45. Miglin, R., Rodriguez, S., Bounoua, N., & Sadeh, N. (2021). A multidimensional examination of psychopathy traits and gray matter volume in adults. *Social Cognitive and Affective Neuroscience*, nsab131. <https://doi.org/10/gnsvfb>
46. Murray, R. J., Schaer, M., & Debbané, M. (2012). Degrees of separation: A quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self- and other-reflection. *Neuroscience & Biobehavioral Reviews*, 36(3), 1043–1059. <https://doi.org/10.1016/j.neubiorev.2011.12.013>
47. Na, C., & Jang, S. J. (2019). Positive Expected Selves and Desistance among Serious Adolescent Offenders. *Journal of Developmental and Life-Course Criminology*, 5(3), 310–334. <https://doi.org/10.1007/s40865-019-00109-4>
48. Northoff, G., & Tumati, S. (2019). "Average is good, extremes are bad" – Non-linear inverted U-shaped relationship between neural mechanisms and functionality of mental features. *Neuroscience & Biobehavioral Reviews*, 104, 11–25. <https://doi.org/10/gjfdw4>
49. Oberlader, V., Banse, R., Beier, S., & Schmidt, A. F. (2021). *Law-abiding versus criminal identity and self-efficacy: A quantitative approach to unravel psychological factors supporting desistance from crime*. PsyArXiv. <https://doi.org/10.31234/osf.io/wp6m8>
50. Parise, M., Canzi, E., Olivari, M. G., & Ferrari, L. (2019). Self-concept clarity and psychological adjustment in adolescence: The mediating role of emotion regulation. *Personality and Individual Differences*, 138, 363–365. <https://doi.org/10.1016/j.paid.2018.10.023>
51. Paternoster, R., Bachman, R., Kerrison, E., O'Connell, D., & Smith, L. (2016). Desistance from Crime and Identity: An Empirical Test With Survival Time. *Criminal Justice and Behavior*, 43(9), 1204–1224. <https://doi.org/10.1177/0093854816651905>
52. Paternoster, R., & Bushway, S. (2009). Desistance and the "feared self": Toward an identity theory of criminal desistance. *The Journal of criminal law and criminology*, 1103–1156.
53. Philippi, C. L., & Koenigs, M. (2014). The neuropsychology of self-reflection in psychiatric illness. *Journal of Psychiatric Research*, 54, 55–63. <https://doi.org/10.1016/j.jpsychires.2014.03.004>
54. Raine, A. (2008). From Genes to Brain to Antisocial Behavior. *Current Directions in Psychological Science*, 17(5), 323–328. <https://doi.org/10.1111/j.1467-8721.2008.00599.x>
55. Ribeiro da Silva, D., Rijo, D., & Salekin, R. T. (2020). Psychopathic traits in children and youth: The state-of-the-art after 30 years of research. *Aggression and Violent Behavior*, 55, 101454. <https://doi.org/10.1016/j.avb.2020.101454>

56. Rocque, M., Posick, C., & Paternoster, R. (2016). Identities through time: An exploration of identity change as a cause of desistance. *Justice Quarterly*, 33(1), 45-72, doi: 10.1080/07418825.2014.894111.
57. Salekin, R. T., Andershed, H., & Clark, A. P. (2018). Psychopathy in children and adolescents: Assessment and critical questions regarding conceptualization. In *Handbook of psychopathy, 2nd ed* (pp. 479–508). The Guilford Press.
58. Salihovic, S., Özdemir, M., & Kerr, M. (2014). Trajectories of Adolescent Psychopathic Traits. *Journal of Psychopathology and Behavioral Assessment*, 36(1), 47–59. <https://doi.org/10.1007/s10862-013-9375-0>
59. Schettini, E., Wilson, S., & Beauchaine, T. P. (2021). Internalizing–externalizing comorbidity and regional brain volumes in the ABCD study. *Development and Psychopathology*, 33(5), 1620–1633. <https://doi.org/10.1017/S0954579421000560>
60. Shaffer, D., Fisher, P., Lucas, C., Dulcan, M., & Schwab-Stone, M. (2000). Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry* 39: 28-38. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 28–38. <https://doi.org/10.1097/00004583-200001000-00014>
61. Sifneos P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and psychosomatics*, 22(2), 255–262. <https://doi-org.ezproxy.leidenuniv.nl/10.1159/000286529>
62. Silvers, J. A., & Guassi Moreira, J. F. (2019). Capacity and tendency: A neuroscientific framework for the study of emotion regulation. *Neuroscience Letters*, 693, 35–39. <https://doi.org/10.1016/j.neulet.2017.09.017>
63. Silvers, J. A., Hubbard, A. D., Biggs, E., Shu, J., Fertuck, E., Chaudhury, S., Grunebaum, M. F., Weber, J., Kober, H., Chesin, M., Brodsky, B. S., Koenigsberg, H., Ochsner, K. N., & Stanley, B. (2016). Affective lability and difficulties with regulation are differentially associated with amygdala and prefrontal response in women with Borderline Personality Disorder. *Psychiatry Research: Neuroimaging*, 254, 74–82. <https://doi.org/10.1016/j.psychres.2016.06.009>
64. Trzesniewski, K., Donnellan, M., Moffitt, T., Robins, R., Poulton, R., & Caspi, A. (2006). Low Self-Esteem During Adolescence Predicts Poor Health, Criminal Behavior, and Limited Economic Prospects during Adulthood. *Developmental Psychology*, 42, 381–390. <https://doi.org/10.1037/0012-1649.42.2.381>
65. van de Groep, I. H., Bos, M. G. N., Jansen, L. M. C., Achterberg, M., Popma, A., & Crone, E. A. (2021). Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others. *Neuropsychologia*, 161, 108000. <https://doi.org/10/gn267m>
66. van de Groep, I. H., Bos, M. G. N., Jansen, L. M. C., Kocevská, D., Bexkens, A., Cohn, M., van Domburgh, L., Popma, A., & Crone, E. A. (2022). Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback. *NeuroImage: Clinical*, 34, 102973. <https://doi.org/10.1016/j.nicl.2022.102973>
67. van de Schoot, R., & Wong, T. M. L. (2012). Do Delinquent Young Adults Have a High or a Low Level of Self-concept? *Self and Identity*, 11(2), 148–169. <https://doi.org/10.1080/15298868.2010.517713>
68. van der Aar, L. P. E., Crone, E. A., & Peters, S. (2019). What Characterizes Adolescents Struggling With Educational Decision-Making?: The Role of Behavioral and Neural Correlates of Self-Concept and Self-Esteem. *Mind, Brain, and Education*, mbe.12209. <https://doi.org/10.1111/mbe.12209>

69. van der Cruijisen, R., Peters, S., & Crone, E. A. (2017). Neural correlates of evaluating self and close-other in physical, academic and prosocial domains. *Brain and Cognition*, 118, 45–53. <https://doi.org/10.1016/j.bandc.2017.07.008>
70. van der Cruijisen, R., Peters, S., Zoetendaal, K. P. M., Pfeifer, J. H., & Crone, E. A. (2019). Direct and reflected self-concept show increasing similarity across adolescence: A functional neuroimaging study. *Neuropsychologia*, 129, 407–417. <https://doi.org/10.1016/j.neuropsychologia.2019.05.001>
71. van der Cruijisen, R., Peters, S., van der Aar, L.P.E. & Crone, E.A. (2018). The neural signature of self-concept development in adolescence: The role of domain and valence distinctions. *Developmental Cognitive Neuroscience*. <https://doi.org/10.1016/j.dcn.2017.11.005>
72. van Drunen, L., Dobbelaar, S., van der Cruijisen, R., van der Meulen, M., Achterberg, M., Wierenga, L. M., & Crone, E. A. (2021). The nature of the self: Neural analyses and heritability estimates of self-evaluations in middle childhood. *Human Brain Mapping*, 42(17), 5609–5625. <https://doi.org/10.1002/hbm.25641>
73. van Zalk, M. H. W., & Van Zalk, N. (2015). Violent peer influence: The roles of self-esteem and psychopathic traits. *Development and Psychopathology*, 27(4pt1), 1077–1088. <https://doi.org/10.1017/S0954579415000693>
74. Ybrandt, H. (2008). The relation between self-concept and social functioning in adolescence. *Journal of Adolescence*, 31(1), 1–16. <https://doi.org/10.1016/j.adolescence.2007.03.004>
75. Zamani, A., Carhart-Harris, R., & Christoff, K. (2022). Prefrontal contributions to the stability and variability of thought and conscious experience. *Neuropsychopharmacology*, 47(1), 329–348. <https://doi.org/10/gmw45f>
76. Zhao, W., Luo, L., Li, Q., & Kendrick, K. (2013). What Can Psychiatric Disorders Tell Us about Neural Processing of the Self? *Frontiers in Human Neuroscience*, 7. <https://doi.org/10.3389/fnhum.2013.00485>

Acknowledgements

We would like to thank all participants for their participation, and Veronica Baljé, Merel Spaander, Mara Giamboi, Cevdet Acarsoy, Sjors Bootsman, Esmee Haverkort, Angelique de Groot, Eline Vink, Fleur van der Linden, Hannah Woerdman, Julie Tournier, Lenore van Dijk, Mark Cnossen, Merlijn Smis, Marjolein Beemster and Bram Zwanenburg for their help with data collection.

SUPPLEMENTARY MATERIALS

APPENDIX A

A.1 Introduction

A.1.1 The association between self-concept and psychopathic traits

Various studies indicate that behavioral responses related to self-evaluations may also be differentially affected by psychopathic traits. For instance, Callous-Unemotional traits have been associated with low self-esteem in adolescents (aged 12-14; Fanti, 2013). On the contrary, high levels of grandiosity might cause people to have a positive self-regard, albeit one that is primarily defensive in nature – while people who score low on Grandiose-Manipulative traits may also display a positive self-concept, but one that is more secure and stable in nature (Ostrowsky, 2010). According to Baumeister et al. (1996), the former type of self-concept positivity causes people to engage in self-preserving antisocial behavior towards others who threaten or dispute their inflated self-view (i.e., in response to ego-threat). However, it should be noted that in children and adolescents (aged 9-18), maladaptive grandiosity or narcissism (e.g. characterized by a sense of entitlement, manipulative behavior to exploit others) was not consistently related to self-esteem in previous research, while adaptive narcissism (characterized by authority and self-sufficiency) was positively associated with self-esteem in children (Barry et al., 2003, 2007). Moreover, it is also possible that a combination of high grandiosity and low self-esteem (vs. high grandiosity and self-esteem) is particularly present among individuals who persistently show antisocial behavior throughout their development (Barry et al., 2003; Fanti, 2013; Fanti & Henrich, 2014) – although some have suggested that their aggressiveness stems from their intent to harm others, rather than from ego-threat (Hart et al., 2019). Given these conflicting findings, it remains unclear whether high or low levels of grandiosity and self-concept are to be expected in young adults with antisocial profiles. Hence, on a behavioral level, we tested two competing hypotheses (Hypothesis 3a) for Grandiose-Manipulative traits: that these traits were (1) positively associated with endorsement of positive self-evaluations (Baumeister et al., 1996; Horvath & Morf, 2010), or (2) negatively associated with endorsement of negative self-evaluations (Barry et al., 2003; Fanti, 2013; Fanti & Henrich, 2014). We expected that Callous-Unemotional traits would be positively associated with endorsement of negative traits (Fanti, 2013). We explored the relationship between Impulsive-Irresponsible traits and self-concept but have no specific hypothesis about this association.

A.2 Methods

A.2.1 Participants

Participants were recruited from all over the Netherlands, with a primary focus on three areas where participants from the childhood arrestee cohort grew up and had been registered for a police offence prior to the age of twelve (Gelderland-Midden, Utrecht and Rotterdam-Rijnmond, cf. van Domburgh et al., 2011). Note that these areas cover different SES and included both rural and urban areas.

A.2.2 Procedure

Prior to participation, participants were informed about the study by telephone and through a digital information letter. Subjects from the control sample completed the study protocol in one session between June and September 2019. For participants in the childhood arrestee sample, data was collected across two sessions (a 'home visit' and scan session). However, because data collection for the childhood arrestee sample was ongoing during the outbreak of the COVID-19 pandemic, these 'home visits' were only conducted at participants' homes between November 2019 and March 13th 2020 (i.e., start of the first lockdown in the Netherlands); and subsequently conducted through skype for business between March 14th 2020 and February 2021. From March 13th 2020 onwards, for the remainder of the childhood arrestee participants, IQ tests were administered during the MRI session, instead of the 'home visit' session (see also van de Groep et al., 2022).

In both subsamples, the aforementioned questionnaires (i.e., YPI and MINI) were administered prior to the scan session. During the scanning session, participants first received instructions about the protocol and performed practice versions of the fMRI task, and subsequently completed a scan protocol that included the self-concept fMRI task.

APPENDIX B

B.1 Self-evaluative, domain and valence specific neural activation

To examine which neural activation was specific for self-evaluations, we examined the following contrasts within a whole brain t-test. First, we tested the contrasts self > control and control > self. As expected, the contrast self > control revealed significant activation in the mPFC, while the reverse contrast resulted in activity in the left Insula, Occipital Gyrus and Precuneus. Next, to examine which neural activation was specific for the specified domains and valences within the self-condition, we examined main and interaction contrasts within a 2 (domain) x 2 (valence) whole brain ANOVA (see Table 2 and Figure 2 for an overview of the results).

First, the contrast positive > negative resulted in increased activity in several cortical midline areas, including the mPFC and Inferior Parietal. The reverse contrast yielded increased activity in the Cerebellum and Inferior Frontal Gyrus. Second, we examined domain-specific neural activation by testing the contrasts prosocial > physical, and physical > prosocial. For the contrast physical > prosocial, we found significant activation in the dlPFC, as well as the Parietal and Temporal Inferior Gyrus. The reversed contrast resulted in activity in the Supplementary Motor Area and Superior Occipital Gyrus. Finally, the interaction between Domain x Valence showed that there was significantly more activity for positive vs negative traits in the physical domain compared to the prosocial domain in the following regions: Left lingual, Right Supplementary Motor Area and Superior Temporal Gyrus.

B2. Individual differences in domain and valence-specific self-evaluations: Psychopathic trait sub-dimensions

All psychopathic trait sub-dimensions showed a similar pattern to the total psychopathic traits scores (see also section 3.1.2 and Figure 2). Accordingly, ANCOVAs on average applicability scores revealed (1) a significant interaction between Domain, Valence and Callous-Unemotional trait scores ($F(1, 91)=22.17, p < .001, \eta_p^2 = 0.198$), (2) a significant interaction between Domain, Valence and Grandiose-Manipulative trait scores ($F(1, 91) = 14.85, p < .001, \eta_p^2 = 0.142$), and (3) a significant interaction between Domain, Valence and Impulsive-Irresponsible trait scores ($F(1, 91) = 14.00, p < .001, \eta_p^2 = 0.135$).

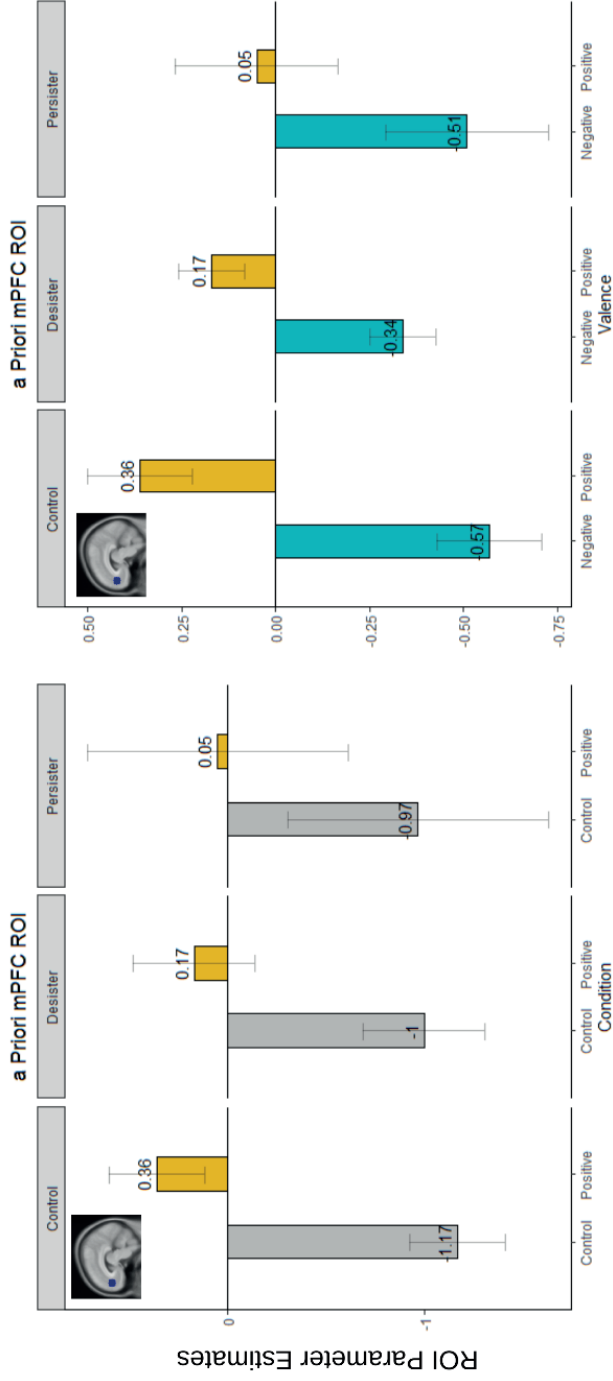


Figure S1. Task condition effects for the contrast Positive Self > Control in the a prior selected mPFC based on prior meta-analyses (see methods) were similar between groups (left panel). Valence effects for the contrast Positive > Negative in the mPFC also did not differ between groups (right panel).

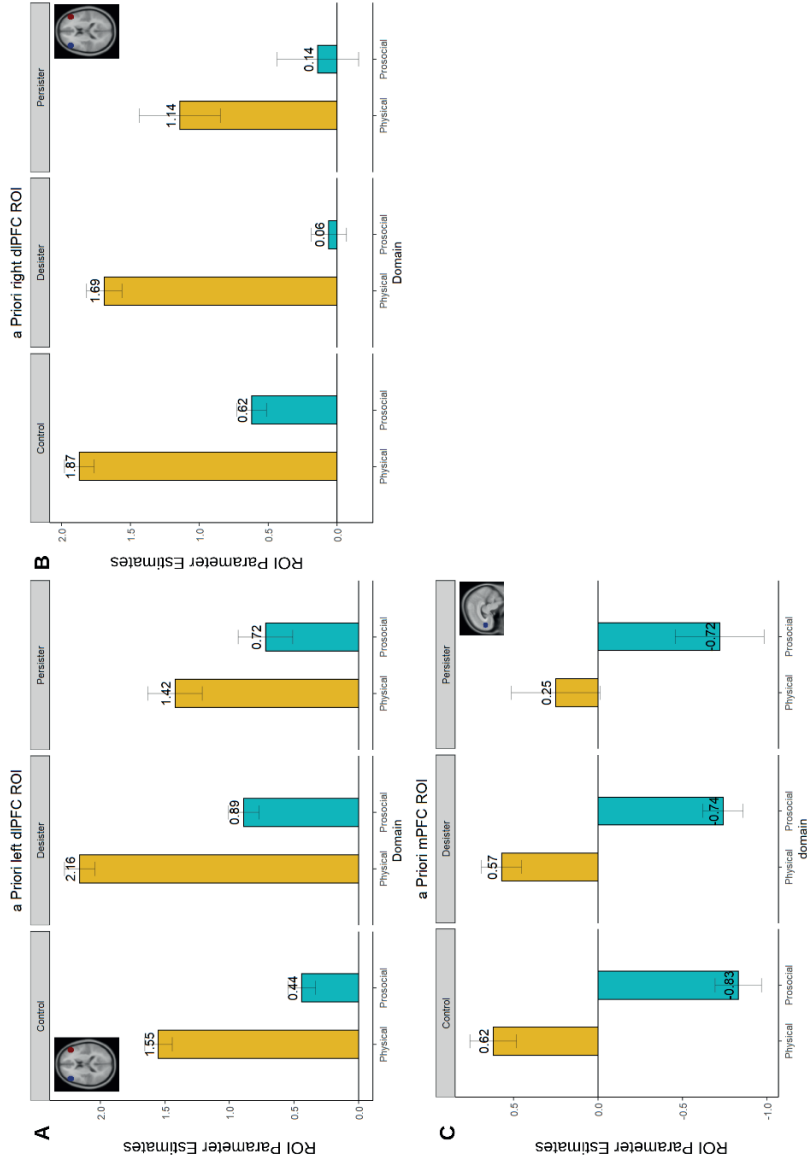


Figure S2. (A) Domain Specific effects for the contrast Physical > Prosocial in the left dlPFC (blue) were similar between groups. (B) Domain Specific effects for the contrast Physical > Prosocial in the right dlPFC (red) were similar between groups. (C) Domain Specific effects for the contrast Physical > Prosocial in the mpFC (blue) were similar between groups.

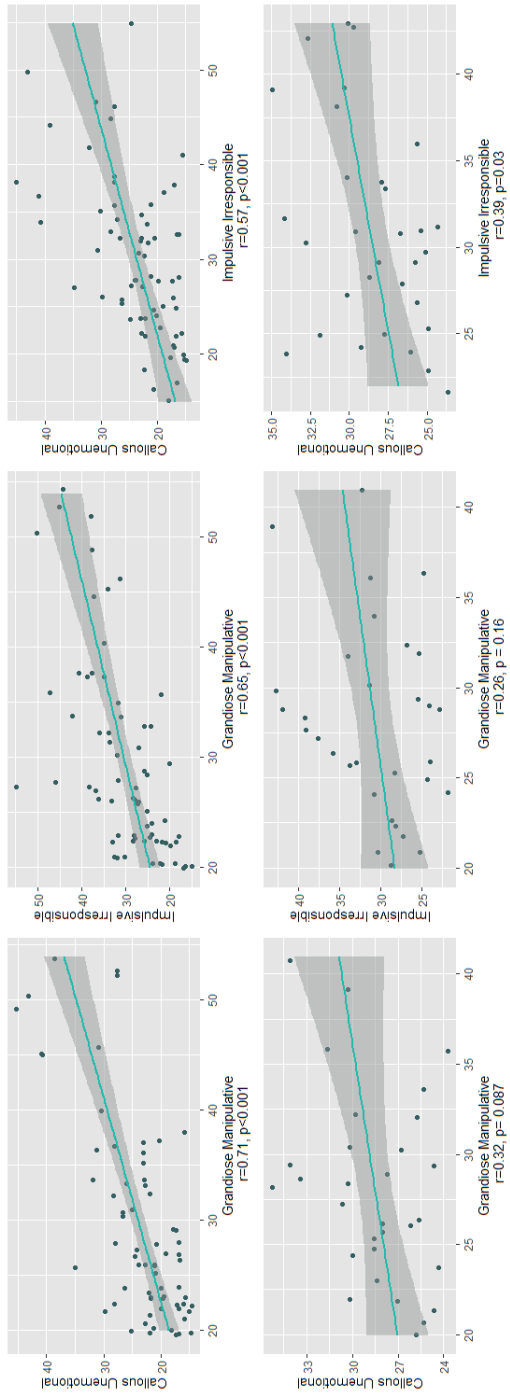


Figure S3. Correlations between the YPI subdimensions (Callous Unemotional, Grandiose Manipulative and Impulsive Irresponsible traits) in the childhood arrestee cohort (panel A) and in the control group (Panel B).

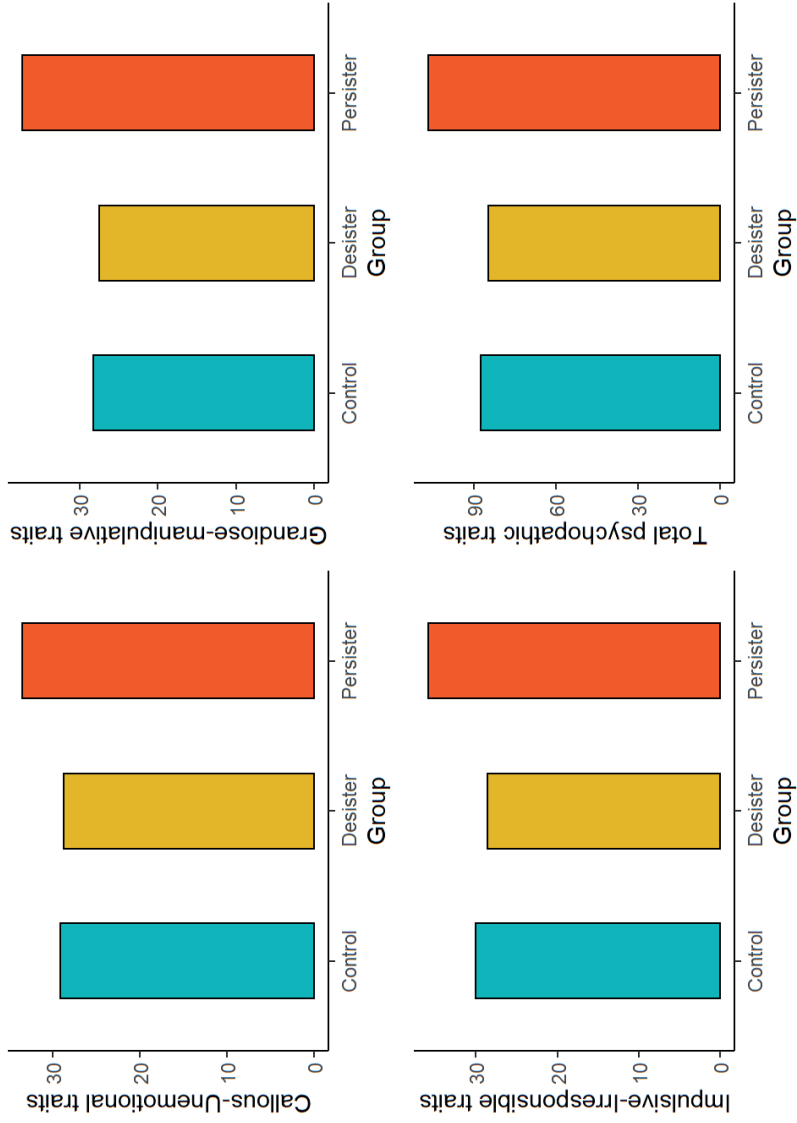
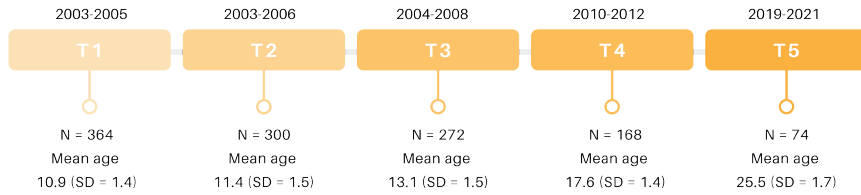


Figure S4. Mean scores for the YPI subdimensions (Callous Unemotional, Grandiose Manipulative and Impulsive Irresponsible traits) and total YPI in the Control, Desister and Persister groups.

A



B

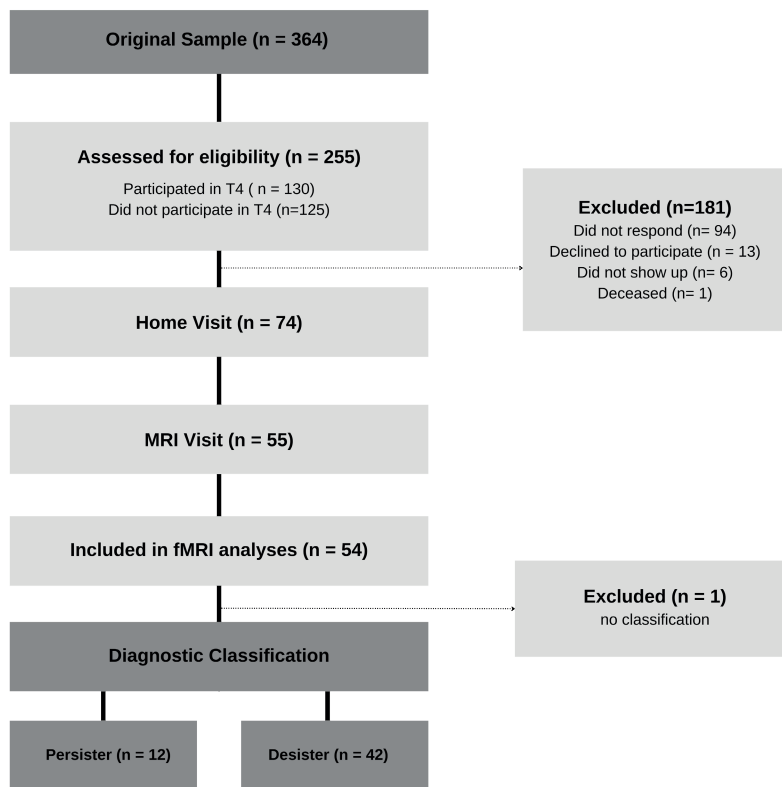


Figure S5. (A) Overview of the longitudinal RESIST study (Research on Individual (Anti-) Social Trajectories), aimed at investigating predictors and consequences of (anti-)social behavior across development. **(B)** Participant flow chart diagram for the fifth wave (T5) of the RESIST study.

Table S1. Attrition analyses

	Excluded Sample (N = 19)	Included Sample (N = 54)	Statistics
No. Persisters	8	12	$\chi^2 = 2.28, p = 0.13$
Age	M = 26.43, SD = 1.17	M = 26.29, SD = 1.53	T(71) = 0.35, $p = 0.72$
IQ ¹	M = 93.72, SD = 10.98	M = 102.89, SD = 13.46	T(71) = -2.67, $p = 0.009$
No. Males	15	46	$\chi^2 < 0.001, p = 1$
YPI Callous-Unemotional Traits ²	M = 29.82, SD = 4.45	M = 29.83, SD = 4.42	T(71) = -0.003, $p = 0.99$
YPI Grandiose-Manipulative Traits ²	M = 27.68, SD = 8.08	M = 29.72, SD = 9.27	T(71) = -0.85, $p = 0.399$
YPI Impulsive-Irresponsible Traits ²	M = 32.21, SD = 9.15	M = 30.20, SD = 7.99	T(71) = 0.91, $p = 0.367$
YPI total score ²	M = 89.72, SD = 17.51	M = 89.74, SD = 19.23	T(71) = -0.005, $p = 0.99$

¹ Note that for 17 participants ($n_{\text{fullcomplete}} = 3, n_{\text{partialcomplete}} = 14$), the IQ tests at T5 were not completed, due to time constraints ($n_{\text{fullcomplete}} = 3$), or limitations imposed by the COVID-19 pandemic to complete the WAIS in person ($n_{\text{partialcomplete}} = 14$). Therefore, we estimated these scores using multiple imputation (mice package v3.13.0; van Buuren & Groothuis-Oudshoorn, 2011), based on the other variables reported in this table, as well as prior IQ scores (T4).

² Note that for three participants ($n_{\text{fullcomplete}} = 2, n_{\text{partialcomplete}} = 1$), the YPI was not completed. Therefore, we estimated these scores using multiple imputation, based on the other variables reported in this table, as well as prior IQ scores (T4).

Table S2. MINI diagnoses in the persister and desister groups¹

MINI Diagnosis	Desister (N = 42)	Persister (N = 12)
Past Major Depressive Disorder	6 (14.29%)	5 (45.45%)
Current Mood Disorder due to physical condition	1 (2.38%)	0 (0.00%)
Past Mood Disorder due to drug use	0 (0.00%)	1 (9.09%)
Agoraphobia	1 (2.38%)	3 (27.27%)
Obsessive-Compulsive Disorder	1 (2.38%)	1 (9.09%)
Generalized Anxiety Disorder	7 (16.6%)	3 (27.27%)
Alcohol dependence / abuse	15 (35.71%)	5 (45.45%)
Drug (non-alcohol) dependence / abuse	10 (23.8%)	7 (63.6%)
Attention Deficit Hyperactivity Disorder	0 (0.00%)	1 (9.09%)
Posttraumatic Stress Disorder	1 (2.38%)	2 (18.18%)

¹ Note that screening for behavioral issues and clinical disorders was performed on the same day of the MRI for the controls (using the diagnostic interview), and for participants from the persister/desister group 0-375 days before the MRI session, with an average of 108 days.

Note that the Chi-Square Test revealed no significant differences between groups, $\chi^2 = 9.12, p = .43$.

Table S3. Diagnoses, mean age and types of index crimes in the persister and desister groups (N = 54)

DISC / MINI Diagnosis ¹	Timepoint																	
	T1			T4			T5			T5								
	Desister	Persister	Total	Desister	Persister	Total	Desister	Persister	Total	Desister	Persister	Total						
DBD ²	9	3	12 ³	0	4	4	-	-	-	-	-	-						
ASPD	-	-	-	-	-	-	0	12	12	-	-	-						
ADHD	4	3	7	4	0	4	-	-	-	-	-	-						
PTSS	-	-	-	1	0	1	-	-	-	-	-	-						
None	29	6	35	29	3	32	-	-	-	-	-	-						
Missing	-	-	-	8	5	13	-	-	-	-	-	-						
Age	T0 ⁴			T1			T2			T3			T4			T5		
Mean	10.49	10.9	10.9	12.01	12.01	13.01	18.11	18.11	26.52	18.11	18.11	26.52						
SD	1.43	1.47	1.47	1.53	1.53	1.58	1.31	1.31	1.63	1.31	1.31	1.63						
Min	5.95	6.21	6.21	7.47	7.47	8.37	14.76	14.76	21.6	14.76	14.76	21.6						
Max	11.97	12.78	12.78	14.38	14.38	15.15	20.38	20.38	29.14	20.38	20.38	29.14						
Type of index crime	Desister (n = 42)			Persister (n = 12)			Total (n = 54)			Total (n = 54)								
Aison	7	7	14	0	0	0	0	0	0	0	0	0						
Theft	7	7	14	1	1	2	8	8	16	8	8	16						
Violent crime	4	4	8	4	4	8	8	8	16	8	8	16						
Public nuisance	11	11	22	0	0	0	11	11	22	11	11	22						
Vandalism	8	8	16	4	4	8	12	12	24	12	12	24						
Missing	5	5	10	3	3	6	8	8	16	8	8	16						

¹ Note that the DISC was only administered at T2 or T3 if it was not administered at T1

² Means participant had at least a DBD diagnosis (DBD CD / ODD / CD + ODD / ADHD + OD, ADHD + CD, ADHD + OD + CD)

³ Note that of these 12 participants, 7 were diagnosed with DBD before the age of 12, and 5 after the age of 12 (during the current study).

⁴ Note that T0 corresponds to the timepoint at which participants were arrested.

Table 4. MNI coordinates of local maxima activated the contrasts Self > Control, Control > Self, Positive > Negative, Negative > Positive, Physical > Prosocial and Prosocial > Physical, and the interaction between Valence and Domain. Results were calculated using a primary voxel-wise threshold of $p < .001$ (uncorrected).

Area of activation	MNI Coordinates			z	T	Test statistic	Cluster Size
	x	y	y				
Self > Control							
SupraMarginal_L	60	-24	38	5.02	307		
Frontal_Inf_Tri_R	52	10	14	4.71	103		
Frontal_Sup_Medial_R	-2	56	2	4.66	366		
Frontal_Mid_R	22	4	58	4.36	230		
Control > Self							
Insula_L	-42	24	18	8.25	1758		
Occipital_Mid_L	-20	-82	-14	7.08	3419		
Precuneus_R	-4	-64	40	4.55	357		
Temporal_Inf_L	-50	-42	2	4.37	374		
Positive > Negative							
Parietal_Inf_L	-40	-28	56	9.66	20189		
Cerebellum_6_R	6	-82	-8	7.31	1963		
Temporal_Sub_R	64	-36	20	5.92	2598		
Frontal_Mid_R	-28	36	42	4.86	372		
Temporal_Inf_R	-52	-66	2	4.28	292		
Negative > Positive							
Cerebellum_Crus1_L	-8	-80	-8	8.84	679		
Temporal_Sub_L	-54	-34	-2	4.25	144		

Area of activation	MNI Coordinates	Test statistic	Cluster Size
Frontal_Inf_Tri_L	-52 24 8	4.15	151
Physical > Prosocial			
Frontal_Inf_Tri_L	-46 36 10	12.82	22647
Temporal_Inf_L	-60 -30 -8	7.67	691
Parietal_Inf_L	-32 -72 48	7.57	1999
Occipital_Mid_R	40 -68 44	7.31	1487
Temporal_Inf_R	56 -40 -8	6.22	510
Precuneus_L	12 -64 30	4.27	405
Prosocial > Physical			
Supp_Motor_Area_R	-6 4 18	5.14	329
Occipital_Sup_R	16 -48 -4	4.36	229
Interaction Valence x Domain			
Lingual_L	-8 -84 0	9.77	12175
Supp_Motor_Area_R	-4 2 62	9.03	11129
Temporal_Sup_L	-50 -34 -2	6.73	1287

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed. See <https://neurovault.org/collections/DNPF5QNK/> for a full, unthresholded overview of activation.

Effects of Age, Sex, IQ and Education on self-evaluations

To test whether Age, Sex, IQ and Education (i.e., the variables that differed between groups) influenced how participants in the different groups evaluated themselves, we repeated the mixed-measures ANOVA, with Valence (positive vs. negative), Domain (prosocial vs. physical), and Group (Persister vs. Desister vs. Control) as independent variables, Age, Sex and IQ and Education as covariates, and self-ratings as dependent measure.

Similar to what we reported in the main text, we observed main effect of valence, $F(1, 89) = 357.09$, $p < .001$, $\eta^2 = 0.800$, and a main effect of domain, $F(1, 89) = 44.96$, $p < .001$, $\eta^2 = 0.336$. There was no significant Group x Valence interaction, $F(2, 89) = 0.31$, $p = .731$, $\eta^2 = 0.007$ – nor a significant Valence, Group and Domain interaction $F(2, 89) = 2.71$, $p = .072$, $\eta^2 = 0.057$.

Effects of Age, Sex, IQ and Education on mPFC activity during self-evaluations

We repeated the three mixed measure ANOVAs using Group as between-subjects factor, and Condition (for the contrast positive self > control), Valence (positive vs. negative) or Domain (prosocial vs. physical) as within-subjects factors, with Age, Sex, IQ and Education as covariates. Once again there was no interaction effect between Group and Condition for the contrast positive > control, $F(2, 89) = 0.27$, $p = .76$, nor between Group and Valence, $F(2, 89) = 1.87$, $p = .160$, for the contrast positive > negative valence. The main effect of Domain remained significant, $F(1, 89) = 70.74$, $p < .001$, $\eta^2 = 0.451$ (see Figure S3C), and the Group x Domain interaction remained non-significant ($F(1, 89) = 0.70$, $p = .501$). Hence, accounting for the covariates did not change the mPFC activity results.

SUPPLEMENTARY REFERENCES

1. Barry, C. T., Frick, P. J., & Killian, A. L. (2003). The relation of narcissism and self-esteem to conduct problems in children: A preliminary investigation. *Journal of Clinical Child and Adolescent Psychology, 32*(1), 139-152. https://doi.org/10.1207/S15374424JCCP3201_13
2. Barry, T. D., Thompson, A., Barry, C. T., Lochman, J. E., Adler, K., & Hill, K. (2007). The importance of narcissism in predicting proactive and reactive aggression in moderately to highly aggressive children. *Aggressive Behavior, 33*(3), 185–197. <https://doi.org/10.1002/ab.20198>
3. Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103*(1), 5–33. <https://doi.org/10.1037/0033-295x.103.1.5>
4. Fanti, K. A., Demetriou, C. A., & Kimonis, E. R. (2013). Variants of callous-unemotional conduct problems in a community sample of adolescents. *Journal of youth and adolescence, 42*(7), 964-979. <https://doi.org/10.1007/s10964-013-9958-9>
5. Fanti, K. A., & Henrich, C. C. (2015). Effects of Self-Esteem and Narcissism on Bullying and Victimization During Early Adolescence. *The Journal of Early Adolescence, 35*(1), 5–29. <https://doi.org/10.1177/0272431613519498>
6. Hart, W., Richardson, K., & Breeden, C. (2019). An interactive model of narcissism, self-esteem, and provocation extent on aggression. *Personality and Individual Differences, 145*, 112–118. <https://doi.org/10.1016/j.paid.2019.03.032>
7. Ostrowsky, M. K. (2010). Are violent people more likely to have low self-esteem or high self-esteem?. *Aggression and Violent Behavior, 15*(1), 69-75. <https://doi.org/10.1016/j.avb.2009.08.004>



Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback

This chapter is published as: van de Groep, I. H., Bos, M. G. N., Jansen, L. M. C., Kocevka, D., Bexkens, A., Cohn, M., van Domburgh, L., Popma, A., & Crone, E. A. (2022). Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback. *NeuroImage: Clinical*, 34, 102973. <https://doi.org/10.1016/j.nicl.2022.102973>

ABSTRACT

Early adulthood has long been recognized as a potential turning point for the development of antisocial behavior, due to changes in social contexts and ongoing psychological and neurobiological maturation. However, it remains unclear how different developmental trajectories of antisocial behavior, their neural underpinnings, and individual differences in psychopathic traits may help explain the distinct developmental outcomes of individuals who persist in or desist from antisocial behavior in early adulthood - such as how they respond to others in social contexts. Therefore, in the current study, young adults (aged 18-30, 68% male) with a persistent or desistant antisocial trajectory (N = 54), as well as healthy controls (N = 39), completed the Social Network Aggression Task, during which they received positive, neutral, or negative feedback on a personal profile and got the opportunity to retaliate by blasting a loud noise. On a behavioral level, results indicated that in all groups, negative peer feedback evoked higher retaliatory aggression, compared to positive and neutral feedback. On a neural level, when receiving social feedback, individuals with persistent or desistant trajectories showed both similar and dissociable patterns of neural activity; desisting and persisting trajectory groups showed higher activity in the Insula, and the desisting trajectory group showed higher activity in dlPFC. Finally, when participants retaliated, they showed increased dlPFC and ACC activity following positive relative to neutral and negative feedback, where ACC activity correlated most strongly with inhibition of retaliatory responses in the desisting trajectory group. Together, these findings provide novel insights in dissociable patterns of brain activity that may increase our understanding of the mechanisms underlying different developmental trajectories of antisocial behavior.

Keywords: antisocial behavior; aggression; social evaluation; fMRI; antisocial developmental trajectories

INTRODUCTION

Antisocial behavior is defined as behavior that violates the rights or wellbeing of others, and often conflicts with age-appropriate norms and rules (Frick et al., 2018). Children who show antisocial behavior at an early age are at risk of developing persistent antisocial behavior (i.e., antisocial behavior that continues throughout adolescence and adulthood; also known as life-course or early-onset persistent antisocial behavior Moffitt, 1993, 2018, and for poorer outcomes in various life domains related to health, finances and social relationships (Moffitt, 2018; Pulkkinen et al., 2009). However, only a small group of children with an early onset of antisocial behavior actually show persistent antisocial behavior throughout their lives, and in general, desistance from antisocial development is the norm (Bersani & Doherty, 2018; Moffitt, 2018). In line with this idea, longitudinal research has identified people who show desistant, *childhood-limited* antisocial behavior, who are likewise characterized by high levels of conduct problems early in life (Aguilar et al., 2000; Bevilacqua et al., 2018; Fairchild et al., 2013; Odgers et al., 2007, 2008). These children typically desist from antisocial behavior in adolescence and early adulthood, and show better life outcomes in the majority of (but not all) domains compared to the life-course persistent and adolescence limited groups (Bevilacqua et al., 2018; Carlisi et al., 2020, 2021; Odgers et al., 2007, 2008).

Hence, while early developmental experiences are considered important to the etiology and maintenance of antisocial behavior (Moffitt, 2018), an early age of onset of antisocial behavior itself is not a strong predictor of antisocial development, since it has been associated with both relatively adverse outcomes (in the case of life-course persistent antisocial behavior), and relatively positive outcomes (in the case of childhood-limited antisocial behavior, see Bevilacqua et al., 2017; but see Carlisi et al., 2021; Moffitt et al., 2002). Therefore, it is important to identify possible candidate mechanisms that allow us to gain a better understanding why such developmental outcome differences (i.e., desisting or persisting antisocial trajectories) arise between groups with an early onset. Here, we propose that investigating different behavioral and neural responses to social rejection between these subgroups may contribute to a better understanding of the mechanisms that underlie persistence and desistance of antisocial behaviors in early adulthood.

Behavioral and neural responses to social rejection in adolescence and early adulthood

An emerging body of evidence suggests that social context-related factors, such as social (peer) rejection, may be important to understand why differences in developmental outcomes emerge between persisters and desisters in early adulthood (Moffitt, 1993; Monahan et al., 2009; Veenstra et al., 2009). Changes in social context can either provide positive developmental opportunities and desistance of antisocial behavior, or aggravate existing patterns of antisocial behavior (Cyr et al., 2020; Hyde et al., 2018). In the transition

from childhood to adolescence, there is a reorientation in the social context toward peers (Crone & Dahl, 2012) and adolescents become increasingly susceptible to peer influence (Prinstein & Dodge, 2008). Accordingly, peers often have a profound influence on the development of antisocial behavior (Mohanan et al., 2009; Moffitt 1993), which manifests itself through peer selection (i.e., who people tend to affiliate with; e.g. antisocial or prosocial peers; Kandel, 1978; Monahan et al., 2009) and peer socialization (i.e., how affiliation and social interactions influence subsequent (anti-)social behavior; Monahan et al., 2009; Prinstein & Dodge, 2008)). The transition from adolescence to early adulthood is also characterized by a shift in social context, marked by taking on new social roles and establishing long-term relationships (Arnett, 2007), more freedom and less social control (Arnett, 2005). Generally, when people develop into young adults, they become less susceptible to effects of peer selection and socialization (Monahan et al., 2009; Arnett, 2007). However, there are also remarkable individual differences in the speed of, and the actual development of this capacity (Monahan et al., 2009).

Possibly, such individual differences may coincide with persistent or desistant developmental trajectories. In children and adolescents who display life-course persistent antisocial behavior, repeated social rejection often triggers maladaptive behaviors such as aggression (Veenstra et al., 2009), likely in an attempt to gain social acceptance or maintain positive self-views (David & Kistner, 2000; van de Groep et al., 2021; Veenstra et al., 2009). This aggressive behavior, in turn, is likely to elicit more social rejection by prosocial peers, and affiliation with deviant peers, which may result in a vicious cycle of maladaptive antisocial behavior throughout development (Veenstra et al., 2009). In contrast, people who desist from antisocial behavior may be more likely to have positive, prosocial experiences throughout development that allow them to deflect from antisocial responses (Cyr et al., 2020; Hyde et al., 2018). However, much less is known about whether these group differences extend into early adulthood (Hyde et al., 2018; Moffitt et al., 2002).

To study (immediate) behavioral and neural responses to social rejection, previous studies in adults have used social exclusion (Cyberball; e.g. Chester et al., 2014) and social feedback paradigms (e.g. Social Network Aggression Task (SNAT); Achterberg et al., 2016). In the SNAT, participants are evaluated on their personal profile and receive acceptance, neutral or rejection feedback by age-matched peers. Subsequently, they can blast a noise towards the peer in response to the feedback (see also Chester et al., 2014 for a similar approach). Rejection feedback was associated with longer noise blasts (Achterberg et al., 2016, van de Groep et al., 2021; Chester et al., 2014), which is indicative of more retaliatory / aggressive responses.

Interestingly, the SNAT has also been used to examine the neural underpinnings of feedback processing and retaliatory responses in children, adolescents and young adults. On a neural level, social rejection and acceptance feedback led to increased activity in the Insula, ACC and Medial Prefrontal Cortex (mPFC), regions often associated with saliency processing (Achterberg et al., 2016, 2017, 2018, 2020). Increased neural activation in the

dIPFC after negative feedback (relative to positive or neutral) has been associated with less aggressive behavior after negative feedback (Achterberg et al., 2016, 2017, 2018, 2020; van de Groep et al., 2021). These findings are consistent with other social feedback paradigms that demonstrated a causal relation between dIPFC stimulation through transcranial magnetic stimulation and aggression following rejection (Riva et al., 2015).

It has recently been suggested that the aforementioned brain areas (i.e., Insula, ACC, mPFC and dIPFC) may be important for differentiating between positive and negative development opportunities in early adulthood (Taber-Thomas & Pérez-Edgar, 2015), and they have been implicated in the stability and severity of antisocial behavior (Alegria et al., 2016; Aoki et al., 2014; Carlisi et al., 2020, 2021; Dugré et al., 2020; Fairchild et al., 2011; Yang & Raine, 2009). More specifically, social changes in early adulthood are accompanied by continuous changes in brain function and structure (Herting et al., 2018; Tamnes et al., 2017), and the interaction between the emerging social context and neural development may give rise to developmental opportunities and vulnerabilities (Taber-Thomas & Perez-Edgar, 2015). Moreover, recent neuroimaging studies indicate that life-course persistent antisocial behavior is characterized by abnormal functional and structural development of both cortical and subcortical brain areas, whereas adolescence-limited and childhood-limited antisocial behavior are not (Fairchild et al., 2011; Carlisi et al., 2020; 2021). Together, these findings suggest that investigating functional imaging during social rejection may further elucidate possible mechanisms underlying different developmental trajectories of antisocial behavior into early adulthood.

Accounting for heterogeneity in antisocial behavior: individual differences in psychopathic traits

Over the past few years, it has become increasingly clear that aggression is heterogeneous, in its causes, underlying motivations and expression (Girard et al., 2019). Accordingly, researchers have argued against a categorical approach of investigating antisocial and aggressive behavior in individuals who desist or persist in these behavioral profiles, and have argued for a more dimensional perspective on psychopathology, which allows for more nuanced approaches to investigate individual differences (Garvey et al., 2016; Insel et al., 2010).

One factor that has repeatedly been linked to differences in frequency, severity and persistence of aggressive behavior in social contexts is psychopathy (Blair, 2015; Brennan et al., 2018). For instance, higher levels of psychopathic traits have been associated with aberrant processing during social rejection, and psychopathic trait levels moderated the links between social rejection processing and subsequent self-reported emotional and behavioral responses, such as anger and aggression (Brennan et al., 2018). However, several interrelated dimensions of psychopathy may differentially influence aggressive behavior and its underlying behavioral and neural underpinnings. Indeed, Grandiose-Manipulative interpersonal characteristics (marked by lying, manipulating and

a grandiose sense of self-worth), Impulsive-Irresponsible traits (characterized by impulsivity and irresponsibility) and Callous-Unemotional traits (characterized by a lack of empathy, remorse and shallow affect; Andershed et al., 2002) have all been differently associated with aggression (Jambroes et al., 2018; Orue et al., 2016; Orue & Andershed, 2015) and with altered brain structure and function in the ACC, Insula and dlPFC (Poepl et al., 2019; Yang & Raine, 2009). Hence, considering individual differences in psychopathic traits may further elucidate why young adults behave aggressively in social contexts.

The Current Study

In this pre-registered study, we examined 94 young adults (aged 18-30) who were subtyped according to their history of antisocial behavior as showing (1) persistent antisocial behavior, (2) desistant antisocial behavior or (3) no history of antisocial behavior (henceforth referred to as the control group), with two aims.

Our first aim was to examine (the neural correlates of) aggression regulation following social rejection in early adulthood comparing individuals with different types of antisocial profiles. On a behavioral level, we hypothesized that (1a) across all participants, social rejection results in stronger aggressive responses than positive or neutral social feedback (Achterberg et al., 2016, van de Groep et al., 2021). When comparing groups, we expected that (1b) social rejection results in increased aggression in persisters when compared to desisters and controls (Chester et al., 2014, Achterberg et al., 2016). Second, on a neural level, across all participants, we (2a) expected increased brain activation in the Insula and Anterior Cingulate Cortex (ACC) following positive and negative feedback, when compared to neutral feedback (Achterberg et al., 2016). When comparing groups (2b), we expected that these effects would be stronger in persisters when compared to desisters and controls (Achterberg et al., 2016). Third, we hypothesized that (3a) across all participants, less aggression would be related to increased dlPFC activity, especially during negative feedback (Achterberg et al., 2016). When comparing groups, we expected that (3b) dlPFC activity would be stronger in desisters and controls when compared with persisters (Achterberg et al., 2016). Finally, when considering brain-behavior associations we expected that (3c) the aforementioned relationship between dlPFC activity and aggression would be more negative in desisters and controls when compared to persisters.

The second aim of this study was to examine whether behavioral and neural responses to social rejection differ depending on levels of psychopathic traits. On a behavioral level, we hypothesized that (1c) the three psychopathic trait dimensions (Grandiose-Manipulative, Callous-Unemotional, Impulsive-Irresponsible) are differentially related to aggressive responses following negative feedback. In addition, on a neural level, we expected that (2c) activity in the Insula and ACC are differentially related to the three psychopathic traits (see supplement for the specific behavioral and neural hypotheses for each dimension). The hypotheses, the design and analysis plan were pre-registered prior to data analysis and are available on the Open Science Framework (<https://osf.io/d6fku/>).

2. METHODS

2.1 Participants

The current study was part of a larger longitudinal study on the development of antisocial behavior from late childhood to early adulthood in the Netherlands (Cohn et al., 2012; Cohn et al., 2013; Cohn, Pape, et al., 2015; Cohn, van Lith, et al., 2016; Cohn, Veltman, et al., 2015; Cohn, Viding, et al., 2016; van Domburgh, 2009; Pape et al., 2015; van Domburgh et al., 2008, 2011, 2019; Tielbeek et al., 2018), called 'RESIST' (see Figure S1A for an overview of the five different time points, T1 (2003-2006, mean age 10.9 (SD = 1.4)), T2 (2004-2008, mean age 11.4 (SD = 1.5)), T3 (2005-2008, mean age 13.1 (SD = 1.5)), T4 (2010-2012, mean age 17.6 (SD = 1.4)), T5 (2019-2021, mean age 25.5 (SD = 1.7))). For the current study, we approached participants from the original sample ($N = 364$, prioritizing participants who had participated in and up to the previous wave (T4: $N = 130$), resulting in a sample of 74 participants (see Figure S1A-B). Of these 74 individuals, 55 completed the MRI protocol. Demographic and clinical data did not differ between participants included and excluded for the MRI session (see supplemental methods and Table S1), except for IQ scores, which were higher in the included than the excluded group. In addition, we recruited 40 healthy controls, without a history of antisocial behavior, who also completed the same measures and MRI protocol (see van de Groep et al., 2021). Note that the current study primarily reports cross-sectional data, but uses longitudinal data to determine whether participants desisted or persisted in antisocial behavior (see section 2.2.3).

All participants were screened for fMRI contra-indications and had normal to corrected vision. Participants were excluded from fMRI analyses in case they did not perform or complete the task ($N_{control} = 1, N_{cases} = 0$), if the MRI data was corrupted ($N_{control} = 1, N_{cases} = 0$) or showed excessive head motion ($>3\text{mm}$; $N_{control} = 3; N_{cases} = 1$), resulting in final fMRI sample of 35 controls and 53 cases (42 desisters and 11 persisters), respectively. Head motion did not differ between individuals from the control ($M = 0.095, SD = .052$), desister ($M = 0.087, SD = .060$) or persister groups ($M = 0.105, SD = .102$), $F(2, 93) = 0.43, p = .67$. Analyses on behavioral results were conducted for those participants who completed the task and required questionnaires ($N_{control} = 39$ and $N_{cases} = 54$ (42 desisters and 12 persisters), respectively). See Table 1 for an overview of the descriptive data (total sample and sub-groups).

The study protocol was approved by the VU University Medical Center Medical Ethical Committee (registration number 2009.268 - NL28844.029.09), with local approval from the Leiden Institute for Brain and Cognition. All subjects gave written informed consent in accordance with the Declaration of Helsinki. After completing the experiment, participants were debriefed about the aim of the study and received a financial reimbursement for their participation (75 euros for controls, 100 euros for cases).

2.2 Materials

2.2.1 Social Network Aggression Task.

To investigate the neural basis of social evaluation and subsequent aggressive responses, we used the Social Network Aggression Task (Achterberg et al., 2016). During this task, participants received social feedback (Positive, Negative, Neutral) from unknown same-aged peers, based on a personal profile completed by the participants prior to the experiment. Social feedback valence was signaled by different icons (green thumbs up for Positive feedback, grey circle for Neutral feedback, red thumbs down for Negative feedback; see Figure 1A), with superimposed neutral pictures of same-aged peers. After receiving social feedback, participants were asked to respond to the evaluations by sending hypothetical noise blasts to the same-aged peers. Participants were instructed to press the button always, but could control the loudness of the noise blast with a button press. A longer button press corresponded with a longer noise blast duration (i.e., louder white noise). Noise blast duration was visualized by a volume bar (see Figure 1B for a schematic trial representation). The SNAT consisted of three blocks of 20 trials (60 in total, van de Groep et al., 2021), with three social feedback conditions (i.e., Neutral, Positive, Negative) being semi-randomized across these blocks. Participants could not receive feedback from the same type more than three times in a row. Trial order and jitter timing were optimized using Optseq2 (Dale, 1999).

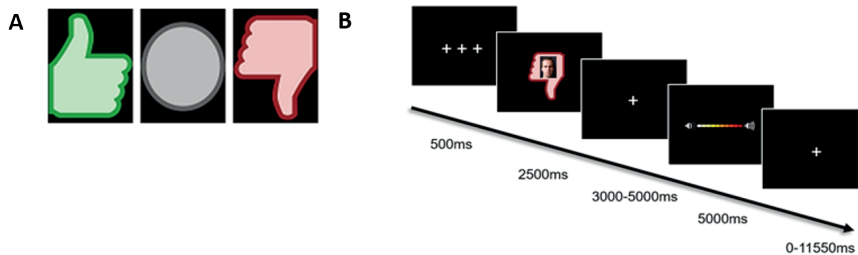


Figure 1. (A) Participants received Positive, Neutral and Negative feedback from same-aged peers. (B) Schematic representation of a Negative feedback trial in the Social Network Aggression Task (SNAT).

2.2.2. Youth Psychopathic Traits Inventory (YPI).

Psychopathic traits were assessed using the Youth Psychopathy Inventory (Andershed et al., 2002), a 50-item self-report questionnaire that distinguishes three trait dimensions: Grandiose-Manipulative, Callous-Unemotional, and Impulsive-Irresponsible traits. Although the questionnaire was originally developed to assess psychopathic traits in adolescents, the YPI has also been validated in young adults (see e.g. Campbell et al., 2009; Neumann & Pardini, 2012). Each item is scored on a 4-point Likert scale (1 = does not apply at all, to 4 = applies very well). For both samples, the reliability of the total YPI score,

Grandiose-Manipulative and Impulsive-Irresponsible subscales was good to excellent, and reliability of Callous-Unemotional traits was poor for both samples. Total and dimensional sub scores are displayed in Table 1.

2.2.3 Antisocial behavior.

All participants with a history of antisocial behavior were arrested by the police before the age of 12. Hence, in the current sample, all individuals with a history of antisocial behavior showed an early onset of such behavior (in the form of a convicted criminal offense, but not necessarily in the form of a disruptive behavior disorder (DBD) (see Table S5), and none of them could be characterized as showing adolescence-limited antisocial behavior. Participants with a history of antisocial behavior were subtyped into different developmental trajectories using diagnostic interviews conducted at ages 14-20 (T4) and ages 21-29 (T5). DBD diagnoses were determined using the National Institute of Mental Health DISC-IV (Shaffer et al., 2000). Antisocial personality disorder was determined by using the MINI-PLUS (Lecrubier et al., 1997), a brief structured diagnostic interview to diagnose psychiatric disorders according to the DSM-IV. Participants were classified as showing persistent antisocial behavior when they received a diagnosis of disruptive behavior disorder at wave 4 (T4) of the longitudinal study, and / or antisocial personality disorder at wave 5 (T5). Of the 54 participants who completed the experimental task, 12 were classified as persister, and 42 as desister (see Figure S1B). One participant did not complete the MINI and could not be classified. Hence, this participant was excluded from all analyses involving subgroup comparisons.

2.3 Procedure

Prior to participation, participants received information about the study by telephone and through a digital information letter. Participants in the control sample completed the whole procedure in one session (June - September 2019, van de Groep et al., 2021). For participants in the 'case' sample, data collection (for the fifth timepoint, T5) was split across two sessions (a 'home visit' and scan session). Both aforementioned questionnaires (i.e., YPI and MINI) were administered during the home visit. Given that data collection for this sample was ongoing during the outbreak of the COVID-19 pandemic, these 'home visits' were conducted at participants' homes between November 2019 and March 13th 2020; and subsequently conducted through skype for business between March 14th 2020 and February 2021. The only other procedural difference was that for the part of the sample who participated after March 13th 2020, IQ tests were completed during the MRI session, instead of the 'home visit' session.

After signing informed consent, all participants filled out several questionnaires prior to the scanning session. During the scanning session, participants first received instructions about the tasks and performed practice versions of the fMRI tasks. Since the current study was part of a larger project, several additional measures were taken during the MRI session.

2.4 Neuroimaging Methods

2.4.1 Neuroimaging Methods: MRI Data Acquisition.

We acquired MRI data using a 3T MRI scanner (Philips Achieva TX, Erlangen, Germany) with a standard whole-head coil. For functional MRI scans, T2*-weighted gradient echo-planar images were collected (repetition time = 2.2 sec, echo time = 30 msec, flip angle = 8 degrees, sequential acquisition: 38 slices, voxel size = 2.75 x 2.75 x 2.75 mm, 80 x 80 matrix, field of view = 220 x 220 x 115 mm). Functional scans were acquired during three runs (corresponding to the three task blocks), which consisted of 150 dynamic scans each. Prior to the first functional scan of each run, we acquired five dummy scans. Stimuli were displayed on a screen that participants could view through a mirror attached to the head coil. Participants' head movements were restricted by using foam inserts at one or both sides of the head. In addition to the fMRI sequences, we collected structural images for anatomical reference (duration of 4 minutes and 12 seconds, high resolution 3D T1, repetition time = 7.9 ms, echo time = 3.5 ms, flip angle = 8 degrees, 3D matrix size for 3D acquisitions: 228 x 177 x 155 slices, axial slice orientation, voxel size = 1.1 x 1.1 x 1.1 mm, field of view = 250 x 196 x 170 mm). T1 dummy scans for stabilization were automatically discarded by the scanner.

2.4.2 Neuroimaging Methods: Preprocessing.

Data were analyzed using SPM12 (Wellcome Department of Cognitive Neurology, London, United Kingdom) using the following steps: realignment, slice-time correction, spatial normalization to T1 templates, spatial smoothing with a 6-mm FWHM isotropic Gaussian kernel. Subsequently, all volumes were resampled to voxels of 3x3x3 millimeters. Our templates were based on the MNI305 stereotaxic space (Cocosco et al., 1997).

2.4.3 Neuroimaging Methods: First level analyses.

To perform first-level individual analyses, we used the general linear model in SPM12. We modelled the fMRI time series as a series of two events convolved with the hemodynamic response function (HRF). More specifically, we first modelled social feedback onset with a zero duration and with separate regressors for the feedback conditions (i.e., Positive, Negative, Neutral). Second, we modelled the noise blast start for the length of the noise blast duration, with separate regressors for noise blasts following Positive, Negative, and Neutral feedback. Each run was modeled as a separate block. In addition, six motion parameters were included as nuisance regressors. Invalid trials (on which participants failed to respond, 1.72% of trials) were modeled separately as a covariate of no interest and were excluded from further analyses. Least-square parameter estimates of the height of the best-fitting canonical hemodynamic response function were used for each condition in pairwise contrasts. These pairwise comparisons led to participant-specific contrast images, which were subsequently submitted to second-level group analyses.

2.4.4 Neuroimaging Methods: Second level analyses.

We first performed a full factorial analysis of variance (ANOVA) with three levels (Positive, Negative, and Neutral feedback) to examine the neural responses to social feedback on a whole-brain level. More specifically, we calculated and tested the contrasts “Positive vs. Negative valence,” “Positive vs. Neutral valence,” “Negative vs. Neutral valence” (and the reversed contrasts) to investigate which brain regions that were specifically activated for social rejection or social acceptance. In addition, we calculated the conjunction “(Positive + Negative) vs. Neutral valence” (and the reversed contrast) to examine which brain regions were specifically activated in response to valenced evaluations.

Second, we exploratively performed another full factorial ANOVA three levels (Positive, Negative, and Neutral feedback) to examine the neural responses during the noise blast on a whole-brain level, using the contrasts “Positive vs. Negative Noise Blast,” “Positive vs. Neutral Noise Blast,” “Negative vs. Neutral Noise Blast,” “(Positive + Negative) vs. Neutral Noise Blast,” (and the reversed contrasts). Finally, we also explored whether brain activity during the noise blast event following positive feedback was associated with the noise blast duration after positive feedback (relative to negative feedback), using a whole brain regression analysis, using the contrasts “Positive vs. Negative Noise Blast,” and “Negative vs. Positive Noise Blast”. All results were corrected using a FDR cluster-corrected threshold of $p < .001$. Coordinates for local maxima are reported in MNI space. Unthresholded statistical maps of all reported whole-brain analyses are available on Neurovault (Gorgolewski et al., 2015); see <https://neurovault.org/collections/THUHIXAC/>.

2.4.5 Neuroimaging Methods: Region-of-Interest analyses (ROIs).

To test for neural differences related to social feedback evaluation, we created 4 ROIs using the MarsBaR toolbox (Brett, Anton, Valabregue, & Poline, 2002) for SPM12 for which we extracted parameter estimates for the left Insula (coordinates $x = -36$, $y = 23$, $z = -2$), right Insula ($x = 33$, $y = 20$, $z = -11$), ACC ($x = 0$, $y = 38$, $z = 16$) (Achterberg et al., 2016), and dlPFC ($x = 48$, $y = 17$, $z = 37$) (Achterberg et al., 2018, van de Groep et al., 2021), based on a-priori hypotheses. All ROIs were created by extracting 10mm spheres around the specified coordinates. For the 4 a-priori defined ROIs, we applied Bonferroni correction for correlated variables with a threshold of $\alpha = 0.0287$ (Perneger, 1998). A more detailed description of ROI analyses can be found in the supplement.

2.5 Statistical analyses

We followed all analyses steps as detailed in our pre-registration on the Open Science Framework (<https://osf.io/d6fku/>). Behavioral and ROI data were analyzed using R (Version 4.0.1, R Core team, 2020). Prior to analyses, assumptions were checked. We identified two univariate noise blast duration outliers for positive feedback. These univariate outlier scores were winsorized (Tabachnick & Fidell, 2013). Results did not change before and after winsorizing. Here, we report the winsorized results.

Table 1 Sample description, group comparisons and reliability estimates

Measure	Group				Statistical comparison
	Desister (n = 42)	Persister (n = 12)	Control (n = 39)		
Gender [n males/females]	36/6	11/1	16/23		
Age [M (SD)]	26.20 (1.63)	26.62 (1.13)	22.7 (3.07)		$F(2,90) = 26.42, p < .001^a$
Education [n]					$\chi^2 = 9.38, p = 0.15$
Vocational	25	6	16		
College	9	2	18		
University	4	1	4		
Other	4	3	1		
IQ [M (SD)] ¹	103.47 (14.13)	100.87 (11.08)	107.47 (11.64)		$F(2,90) = 1.65, p = .19$
YPI Total psychopathic traits [M (SD)] ²	84.91 (16.37)	106.67 (19.51)	87.57 (9.72)		$F(2,90) = 10.85, p < .001^b$
YPI Callous-Unemotional Traits [M (SD)]	28.77 (3.64)	33.5 (5.03)	29.18 (3.09)		$F(2,90) = 8.21, p < .001^b$
YPI Grandiose-Manipulative Traits [M (SD)]	27.53 (7.47)	37.33 (11.16)	28.29 (4.99)		$F(2,90) = 9.20, p < .001^b$
YPI Impulsive-Irresponsible Traits [M (SD)]	28.59 (7.53)	35.83 (7.19)	30.03 (5.70)		$F(2,90) = 5.34, p = .006^b$
Cronbach's Alpha	Cases		Controls		
YPI Total	0.92		0.81		
YPI Callous-Unemotional Traits	0.52		0.39		
YPI Grandiose-Manipulative Traits	0.91		0.78		
YPI Impulsive-Irresponsible Traits	0.86		0.77		

Note. IQ, estimated IQ based on two subscales of the Wechsler Adult Intelligence Scale-IV (Similarities and Block Design), YPI = Youth Psychopathic traits Inventory.

^a Significant differences between controls and desisters, and controls and persisters

^b Significant differences between persisters and desisters, and persisters and controls

¹ Note that for three participants who completed the fMRI session ($n_{\text{persister}} = 1, n_{\text{desister}} = 2$), the IQ tests at T5 were not completed. Therefore, we estimated these scores using multiple imputation based on the other variables reported in this table, as well as prior IQ scores (T4).

² Note that for two participants who completed the fMRI session ($n_{\text{persister}} = 1, n_{\text{desister}} = 1$), the YPI was not completed. Therefore, we estimated these scores using multiple imputation, based on the other variables reported in this table, as well as prior IQ scores (T4).

Table 2 MNI coordinates of local maxima activated the contrasts (1) "Positive > Negative", (2) "Positive > Neutral", (3) " (Positive + Negative) > Neutral", (4) "Positive > (Negative + Neutral)", (5) "(Positive + Neutral) > Negative", (6) "Negative > Neutral", (7) "Negative > (Positive + Neutral)", and (8) "Neutral > Negative" during the feedback event for the Social Network Aggression Task. Results were FDR cluster-corrected using $p < 0.001$. The reversed contrasts "Negative > Positive", "Neutral > Positive" and "(Negative + Neutral) > Positive", "Neutral > (Positive + Negative)" did not result in significant effects. See <https://neurovault.org/collections/THUHIXAC/> for a full, unthresholded overview of activation.

Area of activation	MNI Coordinates			Z	Test statistic T	Cluster Size
	X	Y	Z			
Positive > Negative feedback						
Lingual_R	4	-78	-8		7.90	3591
Angular_R	40	-70	42		5.68	755
Frontal_Mid_R	30	22	52		5.64	848
Putamen_L	-28	4	-2		5.11	730
Cingulum_Post_L	0	-34	24		4.95	798
Frontal_Inf_Oper_L	-44	6	26		4.87	437
Frontal_Inf_Tri_L	-46	30	20		4.70	341
Frontal_mid_L	-28	10	54		4.46	472
Parietal_Inf_L	-34	-74	44		4.46	207
Precuneus_R	4	-68	48		4.39	213
Parietal_Inf_L	-50	-36	48		4.30	242
Supp_Motor_Area_L	-6	-10	52		4.19	215
Positive > Neutral feedback						
Fusiform_R	24	-74	-10		7.62	4841
Frontal_Inf_Tri_L	-34	28	2		5.93	541
Cingulum_Ant_R	6	42	4		4.09	233

Area of activation	MNI Coordinates			Test statistic	Cluster Size
Positive + Negative > Neutral feedback					
Temporal_Inf_R	48	-66	-6	7.8	2638
Occipital_Mid_L	-42	-80	6	7.28	1609
Frontal_Inf_Tri_L	-36	28	2	5.90	335
Positive > Negative + Neutral feedback					
Lingual_L	-20	-76	-10	8.29	3038
Insula_L	-30	18	6	5.43	857
Occipital_Mid_L	-28	-82	20	5.19	239
Cingulum_Post_L	-2	-36	24	5.11	421
Supp_Motor_Area_R	4	6	48	4.36	383
Parietal_Inf_R	52	-56	40	4.26	186
Frontal_Inf_Oper_L	-44	6	26	4.14	176
Frontal_Inf_Tri_L	-40	44	12	4.07	184
Cingulum_Ant_R	10	42	10	3.85	204
Positive + Neutral > Negative feedback					
Frontal_Mid_R	30	2	52	5.99	857
Lingual_R	4	-80	-4	5.79	2370
Angular_R	40	-72	40	5.77	529
Frontal_Mid_L	-28	10	54	5.49	436
Cingulum_Mid_R	4	-38	38	5.07	358
Frontal_Inf_Tri_L	-46	30	20	4.60	204

Area of activation	MNI Coordinates	Test statistic	Cluster Size
Parietal_Inf_L	-34 -74 44	4.50	168
Frontal_Inf_Oper_L	-44 4 24	4.49	310
Negative > Neutral feedback			
Occipital_Sup_L	-12 -96 8	7.67	1041
Temporal_Inf_R	48 -66 -6	7.11	1508
Frontal_Inf_Orb_R	44 26 -1	4.53	172
Negative > Positive + Neutral feedback			
Occipital_Sup_L	-12 -96 8	8.87	165
Calcarine_R	10 -94 10	5.59	326
Occipital_Mid_L	-46 -78 8	5.31	171
Temporal_Inf_R	46 -66 -8	4.51	176
<i>Neutral > Negative feedback</i>			
Frontal_Mid_R	30 12 56	4.99	433

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed.

Table 3 MNI coordinates of local maxima activated the contrasts (1) "Positive > Negative", (2) "Positive > Neutral" and (3) "Positive > (Negative + Neutral)" during the noise blast event for the Social Network Aggression Task. Results were FDR cluster-corrected using $p < 0.001$. The reversed contrasts did not result in significant effects. See <https://neurovault.org/collections/THUHIXAC/> for a full, unthresholded overview of activation.

Area of activation	MNI Coordinates			Test statistic <i>T</i>	Cluster Size
	x	y	z		
Positive > Negative (during Noise Blast)					
Occipital_Sup_L	-14	-78	-2	6.56	12184
Frontal_Inf_Oper_L	-56	4	24	5.61	4404
Frontal_Mid_R	42	6	54	5.05	968
Frontal_Inf_Oper_L	-40	28	34	4.79	793
Putamen_L	-38	-16	10	4.11	324
Positive > Neutral (during Noise Blast)					
Calcarine_R	-2	70	8	5.47	2859
SupraMarginal_L	-42	-50	36	4.64	509
Angular_R	46	-58	46	4.44	466
Positive > Negative + Neutral (during Noise Blast)					
Cerebellum_6_L	-10	-76	-2	6.45	9551
SupraMarginal_L	-44	-52	40	5.23	2711
Frontal_Mid_R	42	6	54	5.10	1247
Fusiform_L	-46	-66	-8	5.01	307
Frontal_Inf_Oper_L	-40	30	34	4.61	495

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed.

Table 4 MNI coordinates of local maxima activated the contrast “Negative > Positive” for the Whole brain regression Social Network Aggression Task. Results were FDR cluster-corrected using $p < 0.001$. The reversed contrasts “Positive > Negative” did not result in significant effects. See <https://neurovault.org/collections/THUHIXAC/> for a full, unthresholded overview of activation.

Area of activation	MNI Coordinates			Test statistic <i>T</i>	Cluster Size
	x	y	z		
Negative > Positive During Noiseblast					
Cuneus_L	-12	-86	2	7.66	9647
Putamen_R	24	-10	2	5.01	242
Parietal_Sup_R	50	-32	54	4.81	232
Supp_Motor_Area_R	0	-4	68	4.75	232
Cingulum_Ant_R	-6	30	26	4.67	371
Caudate_L	-10	4	4	4.37	221

Note: Names were based on the aal toolbox in SPM. For functional regions discussed throughout the paper, both the aal label and functional label (between brackets) are displayed.

3. RESULTS

3.1 Behavioral Results

3.1.1 Behavioral results: Social feedback x Group.

To test whether social feedback and Group status interactively influenced noise blast duration, we performed a repeated-measures ANOVA with Feedback type (Positive vs. Neutral vs. Negative) and Group (Persister vs. Desister vs. Control) as independent variables, and noise blast duration as dependent variable. As can be seen in Figure 2A, there was a main effect of Feedback type, $F(1.18, 104.26) = 44.45$, $p < .001$, $\eta_p^2 = 0.34$, indicating that noise blasts were longest following negative feedback ($M = 1274.04$, $SD = 992.45$), shorter for neutral feedback ($M = 759.86$, $SD = 542.18$), and shortest for positive feedback ($M = 476.93$, $SD = 346.42$; all post hoc comparisons (Bonferroni-corrected), p 's $< .001$), in line with hypothesis 1a. We found no main effect of Group, nor an interaction effect between Feedback type and Group, all p 's $> .05$ (Figure 2B). Hence, in line with this omnibus test and contrary to our hypothesis 1b, we did not find differences in noise blast duration following social rejection (i.e., negative social feedback) between persisters and desisters, or persisters and controls, all p 's $> .05$.

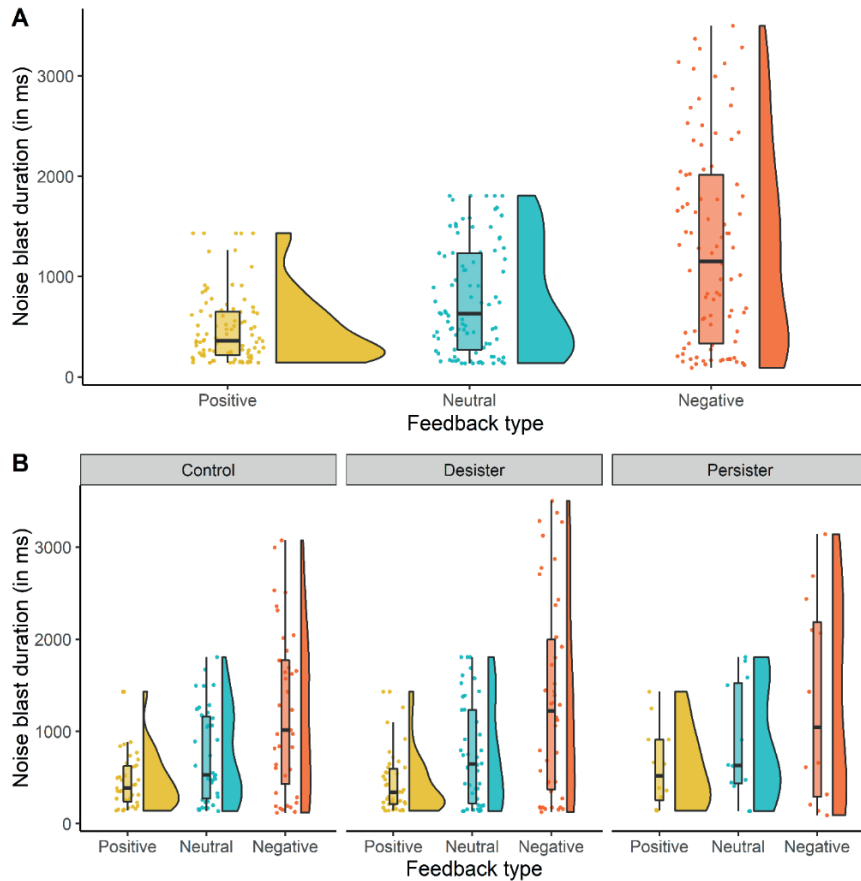


Figure 2. (A) Average noise blast duration following social feedback during the SNAT. Noiseblast duration was longest following Negative feedback, shorter for Neutral feedback and shortest for Positive feedback. (B) Average noise blast duration following social feedback in the different groups (Persisters, Desisters and Controls).

3.1.3 Behavioral results: Social feedback x psychopathic traits.

To investigate whether psychopathic traits influence noise blast duration (hypothesis 1c), we performed repeated measures ANOVAs with Feedback type and Psychopathic traits as independent variables, separately for each trait dimension and the total YPI score (Callous-Unemotional, Grandiose-Manipulative, Impulsive-Irresponsible, YPI Total). Visual inspection and correlation analysis of the association between Grandiose-Manipulative and Impulsive-Irresponsible traits and noise blast duration indicated a positive association, implying that higher Grandiose-Manipulative traits ($R = .13, p = .028$), and higher Impulsive-Irresponsible traits ($R = .12, p = .047$) are associated with longer noise blast

durations (see Figure 3B-C). However, it should be noted that these associations did not survive corrections for multiple testing. Likewise, the ANOVA with psychopathic traits and Feedback type as independent variables did not reveal a significant main effect of Grandiose-Manipulative traits, $F(1, 91) = 3.03, p = .085, \eta_p^2 = 0.032$, Callous-Unemotional traits, $F(1, 91) = 0.68, p = .41, \eta_p^2 = 0.07$, or Impulsive-Irresponsible traits, $F(1, 91) = 2.48, p = .19, \eta_p^2 = 0.03$ on noise blast duration, nor any interactions between feedback type and the three trait dimensions, all p 's $> .68$. Also for the total YPI score, visual inspection and correlation analysis revealed a positive correlation ($R = .13, p = .026$, see Figure 3D), but the ANOVA with YPI score and Feedback type as independent variables revealed no significant main effect of YPI score, $F(1, 91) = 3.13, p = .08, \eta_p^2 = 0.03$, nor interaction effect between feedback type and total YPI score, $F(1.19, 108.08) = .05, p > .87$.

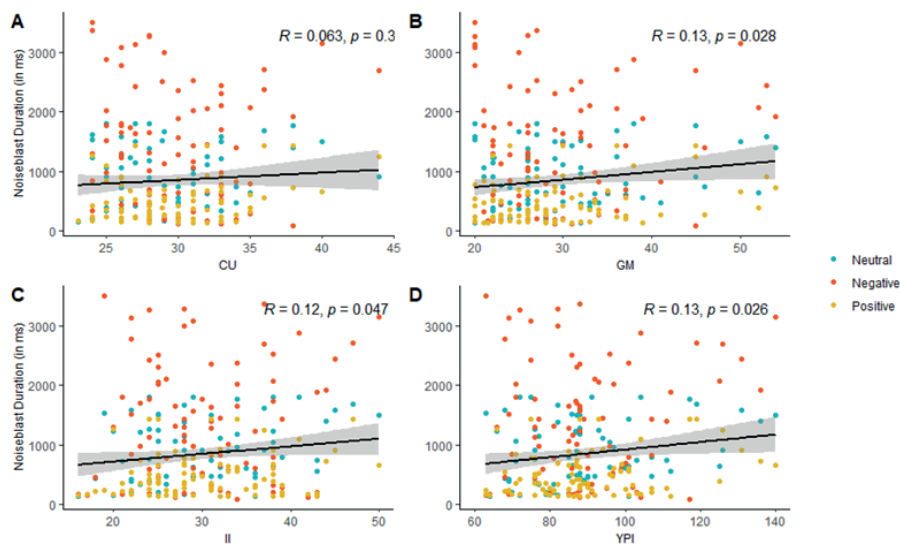


Figure 3. (A-C) Association between Callous-Unemotional (CU), Grandiose-Manipulative (GM) and Impulsive-Irresponsible (II) traits and noise blast duration. **(D)** Association between Total psychopathic traits (YPI) scores and noise blast duration.

3.2 Neural Results Feedback processing

3.2.1 Confirmatory Whole brain Analysis.

To examine neural responses on the whole brain level, we performed a whole brain full-factorial ANOVA with Feedback type (Negative, Positive, Neutral) as within-subject factor (see Table 2 for an overview of the results). First, the “Positive > Neutral” feedback contrast resulted in significant activation in the right ACC / mPFC and left Insula / IFG (Figure 4E). Second, the Valence “(Positive + Negative) vs. Neutral” contrast showed significant activity in the left Insula / IFG (Figure 4G). Third, the “Negative > Neutral”

Feedback contrast yielded significant activation in the right Insula / IFG (Figure 4F). Together, these results indicate increased brain activation in the left and right Insula and Anterior Cingulate Cortex (ACC) following positive and/or negative feedback, when compared to neutral feedback (consistent with hypothesis 2a).

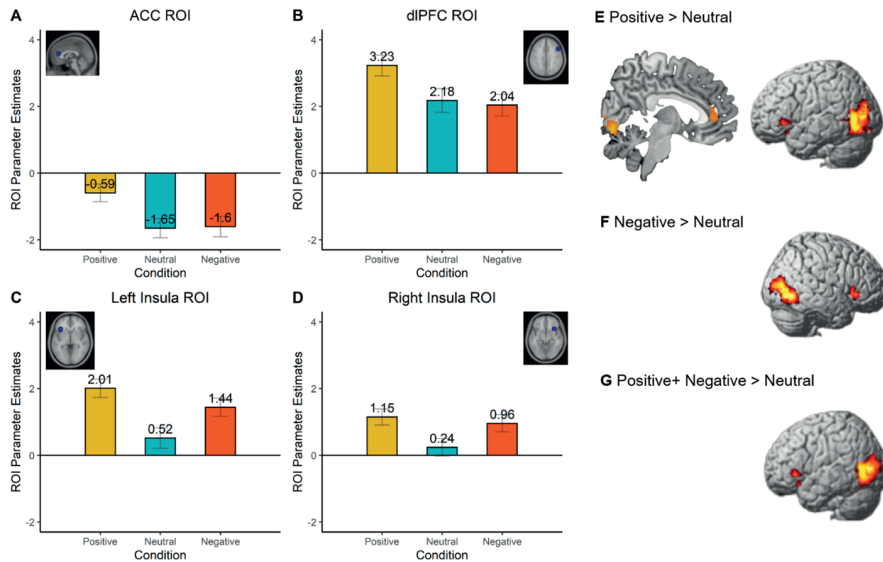


Figure 4. (A-D) Task condition effects (for social feedback) in three pre-defined ROIs (ACC, left and right Insula) and one exploratory ROI (dIPFC). In general, activation was highest for Positive feedback than for Negative and Neutral feedback. (E-G) Whole brain full factorial ANOVA conducted at the group level for the contrasts Positive vs. Neutral feedback (E), Negative vs. Neutral feedback (F) and Positive + Negative vs. Neutral (G).

3.2.2 Confirmatory ROI Analyses.

3.2.2.1 Confirmatory ROI Analyses: The effect of Feedback type (Salience) on Insula and ACC activity

To test whether receiving positive or negative feedback (compared to neutral feedback) resulted in increased brain activation in the Insula and ACC, we also performed repeated measures ANOVAs for the three a priori defined ROIs based on the Achterberg et al. (2016); ACC, left Insula and right Insula (Figure 4A-D). The analyses resulted in main effects of Feedback type on ACC activation, $F(2, 176) = 4.39, p = .013, \eta_p^2 = .048$, the left Insula activation, $F(2, 176) = 6.91, p = .001, \eta_p^2 = .073$, and the right Insula, $F(2, 176) = 3.54, p = .031, \eta_p^2 = .039$, although the latter did not survive Bonferroni correction. As can be seen in Figure 4A-D, for all a priori ROIs, activation was highest for positive feedback, and lowest for neutral feedback. Post-hoc tests (Bonferroni-corrected) yielded increased activation in the ACC following positive feedback compared to neutral feedback ($p = .017$). In addition, we observed significant higher activity in the left Insula, $p = .002$, following positive feedback

vs. neutral feedback. The other differences between conditions in the ACC, left and right Insula were not significant, all other p 's > .036 (see supplementary materials Table S2).

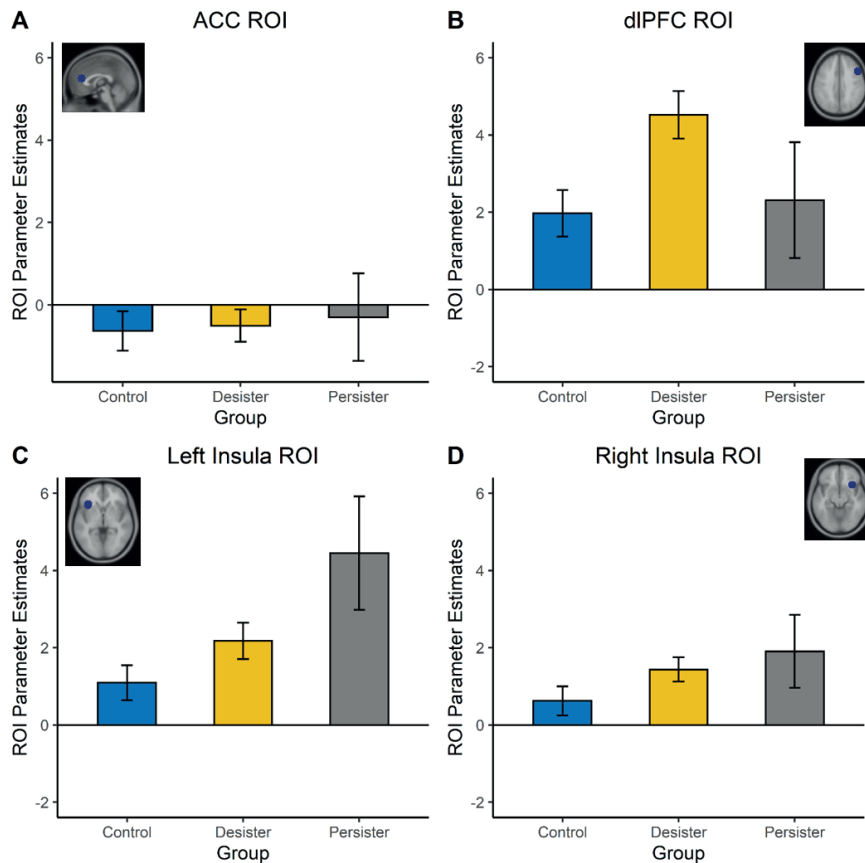


Figure 5. Group effects for three pre-defined ROIs (ACC, left and right Insula) and one exploratory ROI (dIPFC) during feedback processing. For the left Insula (B), right Insula (C), and dIPFC (D), there was a main effect of Group.

3.2.2.2 Confirmatory ROI Analyses: The interactive effects of Feedback type x Group (Salience) on Insula and ACC activity

To test whether the aforementioned saliency effects in the ACC and Insula would be stronger in persisters when compared to desisters and controls, we performed repeated measures ANOVAs with Feedback type and Group status as independent variables (see Figure 5). For the left Insula, we observed a main effect of Group, $F(2, 85) = 3.42, p = .037, \eta_p^2 = .074$, although this effect did not survive Bonferroni correction. For the right Insula, we also found a main effect of Group, $F(2, 85) = 7.37, p = .001, \eta_p^2 = .148$. Post hoc tests (Bonferroni-corrected) showed a difference in right Insula activation between persisters

and controls, $p = .031$, and a significant difference between desisters and controls, $p = .002$ with more activation in persisters and desisters than controls. There was no main effect of Group for the ACC, $F(2, 85) = 1.46, p = .238$, and no Feedback \times Group interactions for the ACC, $F(3.87, 164.34) = 0.55, p = .69$, left Insula, $F(3.89, 165.41) = 1.14, p = .34$, or right Insula, $F(3.95, 167.98) = 0.62, p = .65$.

3.2.2.3 Confirmatory ROI Analyses: The interactive effects between Feedback type \times Psychopathic traits (Salience) on Insula and ACC activity.

To investigate whether psychopathic traits influence brain activation in the Insula and ACC, we tested whether the three psychopathic trait dimensions and total scores (i.e., Callous-Unemotional, Grandiose-Manipulative, Impulsive-Irresponsible and YPI Total) influenced saliency difference scores (i.e., parameter estimates for “(Positive + Negative) > Neutral”) for the ACC, left and right Insula. Contrary to our hypothesis 2c, we found no evidence that psychopathic traits differentially influence activity in these areas (see supplementary materials, table S3).

3.3 Confirmatory ROI Analyses: The effect of Feedback type on dlPFC activity

The next question was to test whether there were significant correlations between dlPFC activity and noise blast for the contrasts “Negative > Positive” and “Negative > Neutral”. Before testing these associations, we explored whether there were main effects of Group or Group \times Feedback type interactions in this ROI.

3.3.1 Confirmatory ROI Analyses: The interactive effect of Feedback type \times Group status on dlPFC activity

To test whether Feedback type and Group interactively influenced activity in the dlPFC, we performed repeated measures ANOVA with Feedback type and Group as the independent variables, and dlPFC parameter estimates as the dependent variable. This analysis yielded a main effect of Group, $F(2, 85) = 9.37, p < .001, \eta_p^2 = .181$, see Figure 5B. Post hoc tests indicated that desisters showed increased dlPFC activity compared to controls, $p < .001$. The other differences between conditions and groups were not significant, all other p 's $> .12$.

3.3.2 Confirmatory Brain Behavior Associations: dlPFC Activity during Feedback processing following negative feedback

Contrary to our expectations (hypothesis 3a), there were no significant correlations between dlPFC activity following negative feedback and noise blast duration following negative feedback (relative to positive and neutral feedback), all p 's $> .57$ (corrected for multiple-testing).

3.3.3 Confirmatory Brain Behavior Associations: dlPFC Activity during Feedback processing following negative feedback between groups

To test for differences between groups with regard to the observed associations between dlPFC activity and noise blast duration following negative feedback (compared to both neutral and positive feedback), we computed fisher r-to-z transformations. Subsequently, we tested whether the correlations were significantly different between groups (i.e., control vs. persisters, control vs. desisters, and persisters vs. desisters) (Lenhard & Lenhard, 2014), corrected for multiple-testing. Contrary to our hypothesis (3c), group status did not influence associations between dlPFC activity following negative feedback and noise blast duration following negative feedback (relative to both positive and neutral feedback). Hence, contrary to what we expected, there were no differences between persisters and controls, $z = -.65, p = .26$ (Negative vs. Positive), $z = -1.03, p = .15$ (Negative vs. Neutral), nor between persisters and desisters, $z = -.35, p = .36$ (Negative vs. Positive), $z = 1.64, p = .051$ (Negative vs. Neutral).

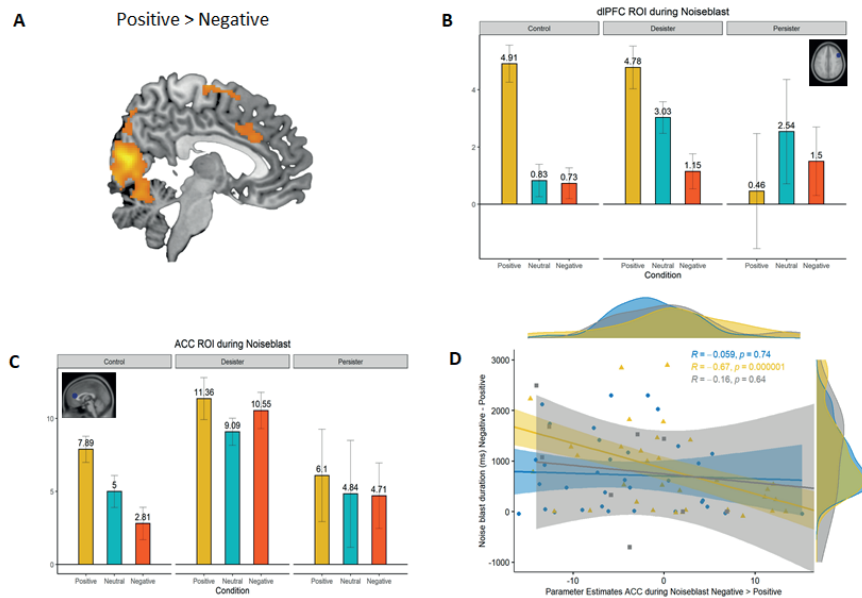


Figure 6. (A) Whole brain regression conducted at the group level for the contrasts Positive vs. Negative feedback. (B) dlPFC parameter estimates during the noise blast event. There was a significant interaction effect between condition and group. (C) ACC parameter estimates during the noise blast event. (D) Difference scores in ACC activity (Negative > Positive feedback) and noise blast duration (Negative > Positive feedback).

3.4 Neural Results Aggressive Responses

3.4.1 Exploratory Whole brain Analyses: Neural activity during the Noise Blast

Based on prior findings, we explored neural activity during the noise blast event (van de Groep et al., 2021). Several contrasts showed significant differences in activation during the whole brain analyses during the noise blast event (see Table 3). First, the contrast “Positive > Negative Noise Blast” resulted in more activity in the left IFG, right frontal middle gyrus and left putamen (see Figure 6A). Second, the contrast “Positive > Neutral Noise Blast” resulted in increased activity in the right Calcarine, left Supramarginal gyrus and right Angular gyrus. Finally, the “Positive > (Negative + Neutral) Noise Blast” contrast resulted in more activity in the left IFG, left Cerebellum, left Supramarginal gyrus, left Fusiform gyrus and right Frontal middle gyrus.

3.4.2 Exploratory ROI Analyses: dlPFC activity during the Noise Blast

Based on earlier findings that suggest differential reactivity to social feedback of the dlPFC during the noise blast event (van de Groep et al., 2021), we extracted ROI values from the whole brain analysis, (coordinates: $x = -34$, $y = 36$, $z = 16$; using a 10mm sphere) to examine this possibility in more detail, and explore whether this reactivity differed between groups, using a Feedback type x Group ANOVA. We found a main effect of Feedback type, $F(1.81, 153.66) = 4.74$, $p = .012$, $\eta_p^2 = .053$. Post hoc tests revealed significantly higher dlPFC activity during noise blast responses following positive feedback compared to negative feedback, $p < .001$, and neutral feedback, $p = .008$. In addition, we observed a Feedback type x Group interaction (See Figure 6B), $F(3.62, 153.66) = 3.11$, $p = .021$, $\eta_p^2 = .068$, which indicated that the persister group showed less dlPFC activity during the noise blast event following positive feedback compared to controls, $p = .02$, and desisters, $p = .02$. Follow-up ANOVAs also revealed that persisters did not respond differently to the different feedback types, $F(2, 20) = .38$, $p = .696$, $\eta_p^2 = .004$, unlike the controls, $F(2, 68) = 16.51$, $p < .001$, $\eta_p^2 = .327$, and desisters, $F(2, 82) = 8.03$, $p < .001$, $\eta_p^2 = .163$. The latter two groups both showed most activity following positive feedback, less after neutral feedback, and least after negative feedback.

3.4.3 Exploratory Whole Brain Regression: Activity during the Noise Blast

Finally, we explored whether brain activity during the noise blast event following positive feedback was associated with the noise blast duration after positive feedback (relative to negative and neutral feedback), using a whole brain regression analysis. We found increased activity following positive feedback was associated with shorter noise blast duration in several areas (see Table 4 and Figure 6C), including the ACC and Dorsal Striatum (Caudate and Putamen). Visual inspection revealed that these associations were mainly driven by the desister subgroup. To further explore this effect, we examined whether brain-behavior associations for these areas differed between groups, using the same approach as described in the previous section. These analyses revealed that desisters

showed stronger negative associations in the ACC compared to the controls (see Figure 6D), $z = 5.037, p < .001$, and persisters, $z = 2.543, p = .005$, as well as in the left Caudate, compared to controls, $z = 4.116, p < .001$, and persisters, $z = 2.572, p = .005$. There were no significant differences between persisters and controls.

4. DISCUSSION

An important developmental question concerns why some people who show antisocial behavior in childhood persist in antisocial behavior into early adulthood, whereas others desist from this trajectory (Hyde et al., 2018; Laub & Sampson, 2021). In this study, we addressed this question using a social aggression paradigm to examine behavioral and neural responses to social feedback in young adults with and without a history of antisocial behavior. We examined the role of social context in retaliatory aggressive behavior using two different, but complementary approaches: a developmental group trajectory approach (i.e., comparing desisters/persisters/controls) and an individual differences approach by examining the association with psychopathic traits. We showed three important behavioral and neural development findings. First, when participants received rejection relative to neutral and positive feedback, they showed higher retaliatory aggression (noise blasts), regardless of group. Moreover, higher retaliatory aggression responses were associated with higher levels of psychopathic traits. Second, when receiving social feedback, individuals with persistent or desistant trajectory of antisocial behavior showed dissociable patterns of neural activity; with higher activity in the Insula for the desisting and persisting trajectory groups (compared to controls) and higher activity in dlPFC only for the desisting trajectory group (compared to the persistent and control groups). Third, when administering the noise blast, participants in the desister and control groups showed increased activity in dlPFC and ACC for positive relative to neutral and negative feedback, whereas ACC activity correlated most strongly with inhibiting noise blasts in the desisting trajectory group. Together, these findings provide novel insights in similar and dissociable patterns of brain activity that suggest differences between various subgroups in how people process social information, and preliminary insights in whether and how they adapt their behavior accordingly in social situations during development.

Research on antisocial behavior is building an increasingly detailed picture of the etiology and maintenance of aggressive behavior throughout development (Moffitt, 2018). Although aggressive behavior in social contexts has been well characterized in childhood and adolescence (Achterberg et al., 2017, 2018, 2020; Bertsch et al., 2020), far less is known about such behavior in early adulthood, particularly in high-risk groups (Bertsch et al., 2020) – even though this developmental period may be a crucial period for the (dis)continuity of antisocial behavior (Hyde et al., 2018; Monahan et al., 2009). This study used a social aggression paradigm that combined social feedback with the

possibility to retaliate by pressing a noise blast (Achterberg et al., 2016; Chester et al., 2014). As expected, noise blasts were longer following rejection feedback, shorter for neutral and shortest for positive feedback, replicating prior findings (Achterberg et al., 2016). Interestingly, this pattern was not different between the persisting, desisting and control subgroups, showing that the basic retaliation response is observed in participants with and without a history of antisocial behavior. Yet, the overall noise blast duration correlated with individual differences in psychopathic traits. More specifically, Grandiose-Manipulative and Impulsive-Irresponsible traits, as well as the total YPI score, were positively associated with noise blast duration. These findings are consistent with prior studies showing positive associations between (subdimensions of) psychopathic traits and reactive aggression (Blais et al., 2014). This pattern also speaks to prior studies noting that a dimensional approach can provide a more sensitive index compared to a categorical approach of DSM diagnoses (Garvey et al., 2016). However, it should be noted that these associations between psychopathic traits and reactive aggression were small in size, and no longer significant when correcting for multiple testing, or when social feedback was added to the ANOVA model, signaling that future research is warranted to better understand how social context influences the link between psychopathy and aggression (Brennan et al., 2018; Van Baardewijk et al., 2009). In addition, the similarity in behavioral patterns between groups raises the question whether future studies should employ stronger social context manipulations that result in more pronounced differences between groups.

Examining the neural basis of information processing can provide a better understanding of underlying neural responses that cannot always be observed at the level of behavior. Indeed, this study replicated the neural pattern observed in prior studies showing that feedback that signals acceptance or rejection leads to increased activity in the Insula and ACC (Dalglish et al., 2017), possibly indicating higher saliency for feedback that has valence information (Dalglish et al., 2017; Eisenberger et al., 2011), or increased monitoring of such socially salient cues, which facilitates updating and selecting appropriate action plans (Puiu et al., 2020). As predicted, we observed that the subgroups differed in neural responses to feedback, although this was observed at the level of general feedback processing and was not valence specific. That is, individuals with a persisting and desisting antisocial trajectory showed higher activity overall in the Insula to all types of social feedback, relative to the control group. Possibly, this exaggerated activity indicates increased salience of social cues in these groups, and/or increased allocation of processing resources to self-relevant and motivational social information (Baskin-Sommers & Newman, 2014; Perini et al., 2018; Puiu et al., 2020). These findings fit with earlier studies showing that antisocial behavior is associated with altered anterior Insula function and structure (Dugré et al., 2020; Noordermeer et al., 2016). However, evidence on the direction of this alteration is currently inconclusive, given that other functional studies tend to find anterior Insula hypoactivity during emotional processing in antisocial populations (Dugré

et al., 2020), rather than hyperresponsiveness. Possibly, the direction of these anterior Insula effects may be context-dependent, contingent on whether the social cues are self-relevant and require a behavioral response (Perini et al., 2018). Our finding that increased anterior Insula activity during social feedback processing seems specific to individuals with a history of antisocial behavior also raises the question whether this neural sensitivity is already apparent early in development, whether it arises as a consequence of repeated antisocial behavior, repeated negative social interactions, or a combination. As such, future research should investigate when and how environmental factors and social interactions shape neural sensitivity to social feedback in populations who display early-onset antisocial behavior during different developmental stages (Ellis et al., 2011; Foulkes & Blakemore, 2018; Muscatello et al., 2020; Schriber & Guyer, 2016).

A novel finding that was not predicted in the pre-registration was that the individuals with a desisting trajectory recruited the dlPFC more strongly during general feedback processing, relative to control and persisting subgroups. Possibly, this increased activity in the desister group reflects increased attention to changing task demands (i.e., to context-dependent changes in feedback presentation between trials), which supports subsequent top-down cognitive control or emotion regulation by preparing response maintenance, selection or inhibition (Niendam et al., 2012). In line with this idea, dlPFC activity during feedback processing in the desister group was highest during positive feedback, compared to neutral and negative feedback. Together with the notion of structural and functional dlPFC impairments in antisocial populations (Yang & Raine, 2009), our finding suggests that increased dlPFC activity may play a role in desisters' ability to successfully adapt their responses and refrain from aggression and other forms of antisocial behavior. However, as of yet, it remains unclear whether this increased dlPFC activity underlies successful behavioral adaptation itself, or reflects increased effortful control which is initiated by desisters after learning that aggressive, retaliatory behavior may not be an optimal, socially adaptive strategy. Future studies may employ transcranial magnetic stimulation (TMS) to further test whether the dlPFC is indeed causally involved in behavioral adaptation, and whether altering dlPFC function in individuals with persistent antisocial behaviors may help them to successfully adapt their behavior.

A final exploratory focus concerned the neural correlates of delivering the noise blast following positive, neutral and negative feedback. Direct comparisons revealed increased activity in the dlPFC and ACC specifically for positive feedback (in the desister and control groups, but not in the persister group) which is the condition where the participants gave the shortest noise blasts. This observation led to the hypothesis that these regions may be involved in the inhibition of retaliation following positive feedback (Brockett et al., 2020; Crew et al., 2021; van Heukelum et al., 2021). Whole brain regression analyses confirmed that successfully being able to regulate aggression after social acceptance was associated with increased activity in the ACC and dorsal striatum (caudate and putamen) during retaliatory responses. Our findings fit with earlier studies showing a negative association

between retaliation and ACC activity (Alegria et al., 2016; Gavita et al., 2012; Krämer et al., 2007; Yang & Raine, 2009) and dorsal striatal activity during retaliatory responses (Kose et al., 2015; Krämer et al., 2007; Lotze et al., 2007), corroborating that these areas are important for adaptive behavioral control of retaliatory responses in a social context (Brockett et al., 2020; Crew et al., 2021; van Heukelum et al., 2021; Bertsch et al., 2020; Grahn et al., 2008). Moreover, the cluster we identified in the ACC was also similar to the dorsal-frontomedial cortex area that has been implicated in the voluntary, intentional inhibition of actions (Filevich et al., 2012), which was confirmed through visual inspection. Interestingly, our findings further revealed the negative association between aggression following positive feedback and activity in the ACC and dorsal striatum during retaliation was stronger in desisters, compared to controls and persisters. Together, these findings point towards a possible adaptive behavioral control mechanism that enables adolescents to desist from antisocial behavior in early adulthood (Bersani & Doherty, 2018; Bertsch et al., 2020; Krämer et al., 2007; Moffit et al., 2002), albeit one that requires significantly more effort to adjust behavior compared to individuals without a history of antisocial behavior.

Finally, our finding that the persistent antisocial behavior group did not show these patterns suggests that they may be less motivated to adapt their behavior, possibly because affiliative and prosocial behavior is not rewarding for them (Foulkes et al., 2014), or they have failed to learn to inhibit aggressive responses, due to problems in stimulus-reinforcement learning and response outcome learning (e.g. cf. Violence Inhibition Mechanism (VIM; Blair, 1995, 2001, 2013)). Future research should further examine these possibilities, and investigate whether this is especially true in contexts that lack interpersonal signals that cue distress in others (Blair, 1995; Van Baardewijk et al., 2009).

Limitations and future directions

Although the current study has many strengths, including a relatively large sample with varying risk of severe antisocial behavior, and assessment of social rejection and subsequent aggression within one experimental fMRI paradigm, and a combination of a dimensional and developmental group trajectory approach, the results should be interpreted in the context of limitations. First, in both the antisocial groups (persisters and desisters), some individuals had (other) mental health problems and/or comorbidities (see Table S4 for more details). While such comorbidities are common in the population of interest (Nichita & Buckley, 2020), and our sample is thus representative in that sense, results should be interpreted with this in mind (Rappaport & Barch, 2020; Simmons et al., 2008; Vetter et al., 2018). Likewise, we cannot rule out that our findings might be influenced by other differences in demographic characteristics between groups, such as gender or age. For example, even though all participants were in the developmental stage of early adulthood, participants in the control group were significantly younger than participants in the other groups. Second, the sub-sample of individuals with a persistent history of antisocial behavior was relatively small in size, which may have limited our power to

detect significant effects. Third, while our aim was to test differences between people who persisted versus desisted in antisocial behavior regardless of age of onset, we acknowledge that age of onset is an important construct to understand the development of antisocial behavior, which should be taken into account in future studies. Although all participants in the persister and desister groups showed an early onset of antisocial behavior (indicated by the young age at which they committed a reported index crime), our sample (size) is not well suited to fully take this factor and possible differences between early vs. later onset into consideration. Fourth, the reliability of the Callous-Unemotional scales in both samples (controls and childhood arrestee cohort) was poor, which fits with earlier research suggesting that affective dimensions of psychopathy are difficult to accurately assess using self-report measures (Hillege et al., 2010; see also Cardinale & Marsh, 2020). Hence, while our study provides preliminary evidence for links between individual differences and social aggression, future work needs to replicate our findings using different types of assessments (Boonmann et al., 2015). Finally, in the current paper, we did not specifically examine heterogeneity within persisting or desisting developmental trajectories. However, recent studies indicate that while many people who desist from antisocial behavior show a positive social development, this is not always the case (Moffitt et al., 2002). For instance, some people who desist might no longer show antisocial behavior, but nevertheless display abnormal social behavior, that is characterized by social isolation and internalizing problems (Moffitt et al., 2002; Carlisi et al., 2021). Future studies should try to further disentangle heterogeneous patterns within persistent and desistant trajectories, and examine how the interplay between neural vulnerabilities and social interactions give rise to diverging patterns of social behavior in early adulthood (Carlisi et al., 2021).

Conclusion

In conclusion, our study provides new evidence of both similar and dissociable patterns of neural activation in individuals with persisting and desisting antisocial trajectories in brain areas that signal socially salient and self-relevant information (Perini et al., 2018), including the bilateral Insula and ACC, and brain areas that are important for behavioral control, such as the dlPFC (Taber-Thomas & Perez-Edgar, 2015). Given that early adulthood is characterized by continuous neurodevelopment in brain areas that are important for adaptive social behavior, this study may help to unravel sensitivities that allow us to understand why children and adolescents desist from negative developmental trajectories before they enter adulthood (Bersani & Doherty, 2018; van Goozen & Fairchild, 2008).

REFERENCES

1. Achterberg, M., Duijvenvoorde, A. C. K. van, IJzendoorn, M. H. van, Bakermans-Kranenburg, M. J., & Crone, E. A. (2020). Longitudinal changes in DLPFC activation during childhood are related to decreased aggression following social rejection. *Proceedings of the National Academy of Sciences*, *117*(15), 8602–8610. <https://doi.org/10.1073/pnas.1915124117>
2. Achterberg, M., Duijvenvoorde, A. C. K. van, Meulen, M. van der, Bakermans-Kranenburg, M. J., & Crone, E. A. (2018). Heritability of aggression following social evaluation in middle childhood: An fMRI study. *Human Brain Mapping*, *39*(7), 2828–2841. <https://doi.org/10.1002/hbm.24043>
3. 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. <https://doi.org/10.1093/scan/nsv154>
4. Achterberg, M., van Duijvenvoorde, A. C. K., van der Meulen, M., Euser, S., Bakermans-Kranenburg, M. J., & Crone, E. A. (2017). The neural and behavioral correlates of social evaluation in childhood. *Developmental Cognitive Neuroscience*, *24*, 107–117. <https://doi.org/10.1016/j.dcn.2017.02.007>
5. Aguilar, B., Sroufe, L. A., Egeland, B., & Carlson, E. (2000). Distinguishing the early-onset/persistent and adolescence-onset antisocial behavior types: From birth to 16 years. *Development and Psychopathology*, *12*(2), 109–132. <https://doi.org/10.1017/S0954579400002017>
6. Alegria, A. A., Radua, J., & Rubia, K. (2016). Meta-Analysis of fMRI Studies of Disruptive Behavior Disorders. *American Journal of Psychiatry*, *173*(11), 1119–1130. <https://doi.org/10.1176/appi.ajp.2016.15081089>
7. Andershed, H. A., Kerr, M., Stattin, H., & Levander, S. (2002). *Psychopathic traits in non-referred youths: A new assessment tool* (pp. 131–158). Elsevier. <http://urn.kb.se/resolve?urn=urn:nbn:se:oru:diva-6748>
8. Aoki, Y., Inokuchi, R., Nakao, T., & Yamasue, H. (2014). Neural bases of antisocial behavior: A voxel-based meta-analysis. *Social Cognitive and Affective Neuroscience*, *9*(8), 1223–1231. <https://doi.org/10.1093/scan/nst104>
9. Arnett, J. J. (2005). The Developmental Context of Substance use in Emerging Adulthood. *Journal of Drug Issues*, *35*(2), 235–254. <https://doi.org/10.1177/002204260503500202>
10. Arnett, J. J. (2007). Emerging Adulthood: What Is It, and What Is It Good For? *Child Development Perspectives*, *1*(2), 68–73. <https://doi.org/10.1111/j.1750-8606.2007.00016.x>
11. Baskin-Sommers, A. R., & Newman, J. P. (2014). Psychopathic and externalizing offenders display dissociable dysfunctions when responding to facial affect. *Personality Disorders*, *5*(4), 369–379. <https://doi.org/10.1037/per0000077>
12. Bersani, B. E., & Doherty, E. E. (2018). Desistance from Offending in the Twenty-First Century. *Annual Review of Criminology*, *1*(1), 311–334. <https://doi.org/10.1146/annurev-criminol-032317-092112>
13. Bertsch, K., Florange, J., & Herpertz, S. C. (2020). Understanding Brain Mechanisms of Reactive Aggression. *Current Psychiatry Reports*, *22*(12), 81. <https://doi.org/10.1007/s11920-020-01208-6>
14. Bevilacqua, L., Hale, D., Barker, E. D., & Viner, R. (2018). Conduct problems trajectories and psychosocial outcomes: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *27*(10), 1239–1260. <https://doi.org/10.1007/s00787-017-1053-4>
15. Blair, R. (2015). Psychopathic traits from an RDoC perspective. *Current Opinion in Neurobiology*, *30*, 79–84. <https://doi.org/10.1016/j.conb.2014.09.011>

16. Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 1–29.
17. Blair, R. J. R. (2001). Advances in neuropsychiatry: Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry*, 71(6), 727–731. <https://doi.org/10.1136/jnnp.71.6.727>
18. Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience*, 14(11), 786–799. <https://doi.org/10.1038/nrn3577>
19. Blais, J., Solodukhin, E., & Forth, A. E. (2014). A Meta-Analysis Exploring the Relationship Between Psychopathy and Instrumental Versus Reactive Violence. *Criminal Justice and Behavior*, 41(7), 797–821. <https://doi.org/10.1177/0093854813519629>
20. Boonmann, C., Jansen, L. M. C., 't Hart-Kerkhoffs, L. A., Vahl, P., Hillege, S. L., Doreleijers, T. A. H., & Vermeiren, R. R. J. M. (2015). Self-Reported Psychopathic Traits in Sexually Offending Juveniles Compared With Generally Offending Juveniles and General Population Youth. *International Journal of Offender Therapy and Comparative Criminology*, 59(1), 85–95. <https://doi.org/10.1177/0306624X13508612>
21. Brennan, G. M., Crowley, M. J., Wu, J., Mayes, L. C., & Baskin-Sommers, A. R. (2018). Neural processing of social exclusion in individuals with psychopathic traits: Links to anger and aggression. *Psychiatry Research*, 268, 263–271. <https://doi.org/10.1016/j.psychres.2018.07.024>
22. Brockett, A. T., Tennyson, S. S., deBettencourt, C. A., Gaye, F., & Roesch, M. R. (2020). Anterior cingulate cortex is necessary for adaptation of action plans. *Proceedings of the National Academy of Sciences*, 117(11), 6196–6204. <https://doi.org/10.1073/pnas.1919303117>
23. Campbell, M. A., Doucette, N. L., & French, S. (2009). Validity and stability of the Youth Psychopathic Traits Inventory in a nonforensic sample of young adults. *Journal of Personality Assessment*, 91(6), 584–592. <https://doi.org/10.1080/00223890903228679>
24. Cardinale, E. M., & Marsh, A. A. (2020). The Reliability and Validity of the Inventory of Callous-Unemotional Traits: A Meta-Analytic Review. *Assessment*, 27(1), 57–71. <https://doi.org/10.1177/1073191117747392>
25. Carlisi, C. O., Moffitt, T. E., Knodt, A. R., Harrington, H., Ireland, D., Melzer, T. R., Poulton, R., Ramrakha, S., Caspi, A., Hariri, A. R., & Viding, E. (2020). Associations between life-course-persistent antisocial behaviour and brain structure in a population-representative longitudinal birth cohort. *The Lancet Psychiatry*, 7(3), 245–253. [https://doi.org/10.1016/S2215-0366\(20\)30002-X](https://doi.org/10.1016/S2215-0366(20)30002-X)
26. Carlisi, C. O., Moffitt, T. E., Knodt, A. R., Harrington, H., Langevin, S., Ireland, D., Melzer, T. R., Poulton, R., Ramrakha, S., Caspi, A., Hariri, A. R., & Viding, E. (2021). Association of subcortical gray-matter volumes with life-course-persistent antisocial behavior in a population-representative longitudinal birth cohort. *Development and Psychopathology*, 1–11. <https://doi.org/10.1017/S0954579421000377>
27. Chester, D. S., Eisenberger, N. I., Pond, R. S., Richman, S. B., Bushman, B. J., & DeWall, C. N. (2014). The interactive effect of social pain and executive functioning on aggression: An fMRI experiment. *Social Cognitive and Affective Neuroscience*, 9(5), 699–704. <https://doi.org/10.1093/scan/nst038>
28. Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B., & Evans, A. C. (1997). BrainWeb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage*, 5, 425.
29. Cohn, M. D., Pape, L. E., Schmaal, L., Brink, W. van den, Wingen, G. van, Vermeiren, R. R. J. M., Doreleijers, T. A. H., Veltman, D. J., & Popma, A. (2015). Differential relations between juvenile psychopathic traits and resting state network connectivity. *Human Brain Mapping*, 36(6), 2396–2405. <https://doi.org/10.1002/hbm.22779>

30. Cohn, M. D., Popma, A., van den Brink, W., Pape, L. E., Kindt, M., van Domburgh, L., Doreleijers, T. a. H., & Veltman, D. J. (2013). Fear conditioning, persistence of disruptive behavior and psychopathic traits: An fMRI study. *Translational Psychiatry*, 3(10), e319–e319. <https://doi.org/10.1038/tp.2013.89>
31. Cohn, M. D., van Lith, K., Kindt, M., Pape, L. E., Doreleijers, T. A. H., van den Brink, W., Veltman, D. J., & Popma, A. (2016). Fear extinction, persistent disruptive behavior and psychopathic traits: FMRI in late adolescence. *Social Cognitive and Affective Neuroscience*, 11(7), 1027–1035. <https://doi.org/10.1093/scan/nsv067>
32. Cohn, M. D., Veltman, D. J., Pape, L. E., van Lith, K., Vermeiren, R. R. J. M., van den Brink, W., Doreleijers, T. A. H., & Popma, A. (2015). Incentive Processing in Persistent Disruptive Behavior and Psychopathic Traits: A Functional Magnetic Resonance Imaging Study in Adolescents. *Biological Psychiatry*, 78(9), 615–624. <https://doi.org/10.1016/j.biopsych.2014.08.017>
33. Cohn, M. D., Viding, E., McCrory, E., Pape, L., van den Brink, W., Doreleijers, T. A. H., Veltman, D. J., & Popma, A. (2016). Regional grey matter volume and concentration in at-risk adolescents: Untangling associations with callous-unemotional traits and conduct disorder symptoms. *Psychiatry Research: Neuroimaging*, 254, 180–187. <https://doi.org/10.1016/j.psychresns.2016.07.003>
34. Cohn, M., van Domburgh, L., Vermeiren, R., Geluk, C., & Doreleijers, T. (2012). Externalizing psychopathology and persistence of offending in childhood first-time arrestees. *European Child & Adolescent Psychiatry*, 21(5), 243–251. <https://doi.org/10.1007/s00787-012-0257-x>
35. Crew, L. A., Covington, H. E., & Hyman, J. M. (2021). Aggression: How the anterior cingulate cortex helps to ensure a fair fight. *Current Biology*, 31(11), R716–R718. <https://doi.org/10.1016/j.cub.2021.04.024>
36. Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, 13(9), 636–650. <https://doi.org/10.1038/nrn3313>
37. Cyr, M., Zheng, Y., & McMahon, R. J. (2020). A long-term look at “early starters”: Predicting adult psychosocial outcomes from childhood conduct problem trajectories. *Development and Psychopathology*, 1–16. <https://doi.org/10.1017/S0954579420000760>
38. Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, 8(2–3), 109–114. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)8:2/3<109::AID-HBM7>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W)
39. Dalgleish, T., Walsh, N. D., Mobbs, D., Schweizer, S., van Harmelen, A.-L., Dunn, B., Dunn, V., Goodyer, I., & Stretton, J. (2017). Social pain and social gain in the adolescent brain: A common neural circuitry underlying both positive and negative social evaluation. *Scientific Reports*, 7(1), 42010. <https://doi.org/10.1038/srep42010>
40. David, C. F., & Kistner, J. A. (2000). Do Positive Self-Perceptions Have a “Dark Side”? Examination of the Link between Perceptual Bias and Aggression. *Journal of Abnormal Child Psychology*, 28(4), 327–337.
41. Domburgh, L. van. (2009). *Very Young Offenders: Characteristics of children and their environment in relation to(re)-offending*. LAP Lambert Academic Publishing.
42. Dugré, J. R., Radua, J., Carignan-Allard, M., Dumais, A., Rubia, K., & Potvin, S. (2020). Neurofunctional abnormalities in antisocial spectrum: A meta-analysis of fMRI studies on Five distinct neurocognitive research domains. *Neuroscience & Biobehavioral Reviews*, 119, 168–183. <https://doi.org/10.1016/j.neubiorev.2020.09.013>
43. Eisenberger, N. I., Inagaki, T. K., Muscatell, K. A., Byrne Haltom, K. E., & Leary, M. R. (2011). The Neural Sociometer: Brain Mechanisms Underlying State Self-esteem. *Journal of Cognitive Neuroscience*, 23(11), 3448–3455. https://doi.org/10.1162/jocn_a_00027

44. Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary--neurodevelopmental theory. *Development and Psychopathology*, 23(1), 7–28. <https://doi.org/10.1017/S0954579410000611>
45. Fairchild, G., Passamonti, L., Hurford, G., Hagan, C. C., von dem Hagen, E. A. H., van Goozen, S. H. M., Goodyer, I. M., & Calder, A. J. (2011). Brain structure abnormalities in early-onset and adolescent-onset conduct disorder. *The American Journal of Psychiatry*, 168(6), 624–633. <https://doi.org/10.1176/appi.ajp.2010.10081184>
46. Fairchild, G., van Goozen, S. H. M., Calder, A. J., & Goodyer, I. M. (2013). Research Review: Evaluating and reformulating the developmental taxonomic theory of antisocial behaviour. *Journal of Child Psychology and Psychiatry*, 54(9), 924–940. <https://doi.org/10.1111/jcpp.12102>
47. Farrington, D. P. (2019). The Duration of Criminal Careers: How Many Offenders Do Not Desist up to Age 61? *Journal of Developmental and Life-Course Criminology*, 5(1), 4–21. <https://doi.org/10.1007/s40865-018-0098-5>
48. Filevich, E., Kühn, S., & Haggard, P. (2012). Intentional inhibition in human action: The power of 'no.' *Neuroscience & Biobehavioral Reviews*, 36(4), 1107–1118. <https://doi.org/10.1016/j.neubiorev.2012.01.006>
49. Foulkes, L., & Blakemore, S.-J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience*, 21(3), 315–323. <https://doi.org/10.1038/s41593-018-0078-4>
50. Foulkes, L., McCrory, E. J., Neumann, C. S., & Viding, E. (2014). Inverted Social Reward: Associations between Psychopathic Traits and Self-Report and Experimental Measures of Social Reward. *PLOS ONE*, 9(8), e106000. <https://doi.org/10.1371/journal.pone.0106000>
51. Frick, P. J., Robertson, E. L., & Clark, J. E. (2018). 6—Callous–unemotional traits. In M. M. Martel (Ed.), *Developmental Pathways to Disruptive, Impulse-Control and Conduct Disorders* (pp. 139–160). Academic Press. <https://doi.org/10.1016/B978-0-12-811323-3.00006-7>
52. Garvey, M., Avenevoli, S., & Anderson, K. (2016). The National Institute of Mental Health Research Domain Criteria and Clinical Research in Child and Adolescent Psychiatry. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(2), 93–98. <https://doi.org/10.1016/j.jaac.2015.11.002>
53. Gavita, O. A., Capris, D., Bolno, J., & David, D. (2012). Anterior cingulate cortex findings in child disruptive behavior disorders: A meta-analysis. *Aggression and Violent Behavior*, 17(6), 507–513. <https://doi.org/10.1016/j.avb.2012.07.002>
54. Girard, L.-C., Tremblay, R. E., Nagin, D., & Côté, S. M. (2019). Development of Aggression Subtypes from Childhood to Adolescence: A Group-Based Multi-Trajectory Modelling Perspective. *Journal of Abnormal Child Psychology*, 47(5), 825–838. <https://doi.org/10.1007/s10802-018-0488-5>
55. Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00008>
56. Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, 86(3), 141–155. <https://doi.org/10.1016/j.pneurobio.2008.09.004>
57. Herting, M. M., Johnson, C., Mills, K. L., Vijayakumar, N., Dennison, M., Liu, C., Goddings, A.-L., Dahl, R. E., Sowell, E. R., Whittle, S., Allen, N. B., & Tamnes, C. K. (2018). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *NeuroImage*, 172, 194–205. <https://doi.org/10.1016/j.neuroimage.2018.01.020>

58. Hillege, S., Das, J., & de Ruiter, C. (2010). The Youth Psychopathic traits Inventory: Psychometric properties and its relation to substance use and interpersonal style in a Dutch sample of non-referred adolescents. *Journal of Adolescence*, 33(1), 83–91. <https://doi.org/10.1016/j.adolescence.2009.05.006>
59. Hyde, L. W., Waller, R., Shaw, D. S., Murray, L., & Forbes, E. E. (2018). Deflections from adolescent trajectories of antisocial behavior: Contextual and neural moderators of antisocial behavior stability into emerging adulthood. *Journal of Child Psychology and Psychiatry*, 59(10), 1073–1082. <https://doi.org/10.1111/jcpp.12931>
60. Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, 167(7), 748–751. <https://doi.org/10.1176/appi.ajp.2010.09091379>
61. Jambroes, T., Jansen, L. M. C., v.d. Ven, P. M., Claassen, T., Glennon, J. C., Vermeiren, R. R. J. M., Doreleijers, T. A. H., & Popma, A. (2018). Dimensions of psychopathy in relation to proactive and reactive aggression: Does intelligence matter? *Personality and Individual Differences*, 129, 76–82. <https://doi.org/10.1016/j.paid.2018.03.001>
62. Kandel, D. B. (1978). Similarity in real-life adolescent friendship pairs. *Journal of Personality and Social Psychology*, 36(3), 306–312. <https://doi.org/10.1037/0022-3514.36.3.306>
63. Kose, S., Steinberg, J. L., Moeller, F. G., Gowin, J. L., Zuniga, E., Kamdar, Z. N., Schmitz, J. M., & Lane, S. D. (2015). Neural Correlates of Impulsive Aggressive Behavior in Subjects With a History of Alcohol Dependence. *Behavioral Neuroscience*, 129(2), 183. <https://doi.org/10.1037/bne0000038>
64. 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. <https://doi.org/10.1016/j.neuroimage.2007.07.029>
65. Laub, J. H., & Sampson, R. J. (2021). 2 Persistence or Desistance? In *Shared Beginnings, Divergent Lives* (pp. 13–35). Harvard University Press. <https://www.degruyter.com/document/doi/10.4159/9780674039971-003/html>
66. Lecrubier, Y., Sheehan, D., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J., & Dunbar, G. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, 12(5), 224–231. [https://doi.org/10.1016/S0924-9338\(97\)83296-8](https://doi.org/10.1016/S0924-9338(97)83296-8)
67. Lenhard, W., & Lenhard, A. (2014). *Testing the Significance of Correlations*. <https://doi.org/10.13140/RG.2.1.2954.1367>
68. 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(1), 470–478. <https://doi.org/10.1016/j.neuroimage.2006.09.028>
69. Moffitt, T. E. (n.d.). *Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy*. 28.
70. Moffitt, T. E. (2018). Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, 2(3), 177–186. <https://doi.org/10.1038/s41562-018-0309-4>
71. Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology*, 14(1), 179–207. <https://doi.org/10.1017/s0954579402001104>

72. Monahan, K. C., Steinberg, L., & Cauffman, E. (2009). Affiliation With Antisocial Peers, Susceptibility to Peer Influence, and Antisocial Behavior During the Transition to Adulthood. *Developmental Psychology, 45*(6), 1520–1530. <https://doi.org/10.1037/a0017417>
73. Muscatello, M. R. A., Rizzo, A., Celebre, L., Mento, C., Pandolfo, G., Cedro, C., Battaglia, F., Zoccali, R. A., & Bruno, A. (2020). The wounds of childhood: Early trauma subtypes, salience and hyperarousal in a sample of adult psychiatric patients. *International Journal of Social Psychiatry, 66*(1), 3–9. <https://doi.org/10.1177/0020764019872227>
74. Neumann, C., & Pardini, D. (2012). Factor Structure and Construct Validity of the Self-Report Psychopathy (SRP) Scale and the Youth Psychopathic Traits Inventory (YPI) in Young Men. *Journal of Personality Disorders, 28*. https://doi.org/10.1521/pedi_2012_26_063
75. Nichita, E. C., & Buckley, P. F. (2020). Comorbidities of Antisocial Personality Disorder. In *The Wiley International Handbook on Psychopathic Disorders and the Law* (pp. 645–670). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119159322.ch28>
76. Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience, 12*(2), 241–268. <https://doi.org/10.3758/s13415-011-0083-5>
77. Noordermeer, S. D. S., Luman, M., & Oosterlaan, J. (2016). A Systematic Review and Meta-analysis of Neuroimaging in Oppositional Defiant Disorder (ODD) and Conduct Disorder (CD) Taking Attention-Deficit Hyperactivity Disorder (ADHD) Into Account. *Neuropsychology Review, 26*(1), 44–72. <https://doi.org/10.1007/s11065-015-9315-8>
78. Odgers, C. L., Caspi, A., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Moffitt, T. E. (2007). Prediction of differential adult health burden by conduct problem subtypes in males. *Archives of General Psychiatry, 64*(4), 476–484. <https://doi.org/10.1001/archpsyc.64.4.476>
79. Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology, 20*(2), 673–716. <https://doi.org/10.1017/S0954579408000333>
80. Orue, I., & Andershed, H. (2015). The Youth Psychopathic Traits Inventory-Short Version in Spanish Adolescents—Factor Structure, Reliability, and Relation with Aggression, Bullying, and Cyber Bullying. *Journal of Psychopathology and Behavioral Assessment, 37*(4), 563–575. <https://doi.org/10.1007/s10862-015-9489-7>
81. Orue, I., Calvete, E., & Gamez-Guadix, M. (2016). Gender moderates the association between psychopathic traits and aggressive behavior in adolescents. *Personality and Individual Differences, 94*, 266–271. <https://doi.org/10.1016/j.paid.2016.01.043>
82. Pape, L. E., Cohn, M. D., Caan, M. W. A., van Wingen, G., van den Brink, W., Veltman, D. J., & Popma, A. (2015). Psychopathic traits in adolescents are associated with higher structural connectivity. *Psychiatry Research: Neuroimaging, 233*(3), 474–480. <https://doi.org/10.1016/j.pscychresns.2015.07.023>
83. Perini, I., Gustafsson, P. A., Hamilton, J. P., Kämpe, R., Zetterqvist, M., & Heilig, M. (2018). The salience of self, not social pain, is encoded by dorsal anterior cingulate and insula. *Scientific Reports, 8*(1), 6165. <https://doi.org/10.1038/s41598-018-24658-8>

84. Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, *316*(7139), 1236–1238. <https://doi.org/10.1136/bmj.316.7139.1236>
85. Poepl, T. B., Donges, M. R., Mokros, A., Rupprecht, R., Fox, P. T., Laird, A. R., Bzdok, D., Langguth, B., & Eickhoff, S. B. (2019). A view behind the mask of sanity: Meta-analysis of aberrant brain activity in psychopaths. *Molecular Psychiatry*, *24*(3), 463–470. <https://doi.org/10.1038/s41380-018-0122-5>
86. Prinstein, M. J., & Dodge, K. A. (2008). *Understanding Peer Influence in Children and Adolescents*. Guilford Press.
87. Puiu, A. A., Wudarczyk, O., Kohls, G., Bzdok, D., Herpertz-Dahlmann, B., & Konrad, K. (2020). Meta-analytic evidence for a joint neural mechanism underlying response inhibition and state anger. *Human Brain Mapping*, *41*(11), 3147–3160. <https://doi.org/10.1002/hbm.25004>
88. Pulkkinen, L., Lyyra, A.-L., & Kokko, K. (2009). Life success of males on nonoffender, adolescence-limited, persistent, and adult-onset antisocial pathways: Follow-up from age 8 to 42. *Aggressive Behavior*, *35*(2), 117–135. <https://doi.org/10.1002/ab.20297>
89. Rappaport, B. I., & Barch, D. M. (2020). Brain responses to social feedback in internalizing disorders: A comprehensive review. *Neuroscience & Biobehavioral Reviews*, *118*, 784–808. <https://doi.org/10.1016/j.neubiorev.2020.09.012>
90. Riva, P., Romero Lauro, L. J., DeWall, C. N., Chester, D. S., & Bushman, B. J. (2015). Reducing aggressive responses to social exclusion using transcranial direct current stimulation. *Social Cognitive and Affective Neuroscience*, *10*(3), 352–356. <https://doi.org/10.1093/scan/nsu053>
91. Schriber, R. A., & Guyer, A. E. (2016). Adolescent neurobiological susceptibility to social context. *Developmental Cognitive Neuroscience*, *19*, 1–18. <https://doi.org/10.1016/j.dcn.2015.12.009>
92. Shaffer, D., Fisher, P., Lucas, C., Dulcan, M., & Schwab-Stone, M. (2000). Shaffer D, Fisher P, Lucas CP, Dulcan MK, Schwab-Stone ME. NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *J Am Acad Child Adolesc Psychiatry* *39*: 28–38. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 28–38. <https://doi.org/10.1097/00004583-200001000-00014>
93. Simmons, A., Matthews, S. C., Paulus, M. P., & Stein, M. B. (2008). Intolerance of uncertainty correlates with insula activation during affective ambiguity. *Neuroscience Letters*, *430*(2), 92–97. <https://doi.org/10.1016/j.neulet.2007.10.030>
94. Tabachnick, B., & Fidell, L. (2013). *Using Multivariate Statistics*, 6th Edn. Northridge, CA: California State University.
95. Taber-Thomas, B., & Pérez-Edgar, K. (2015). *Emerging Adulthood Brain Development*. The Oxford Handbook of Emerging Adulthood. <https://doi.org/10.1093/oxfordhb/9780199795574.013.15>
96. Tamnes, C. K., Herting, M. M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., Crone, E. A., & Mills, K. L. (2017). Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(12), 3402–3412. <https://doi.org/10.1523/JNEUROSCI.3302-16.2017>
97. van Baardewijk, Y., Stegge, H., Bushman, B. J., & Vermeiren, R. (2009). Psychopathic traits, victim distress and aggression in children. *Journal of Child Psychology and Psychiatry*, *50*(6), 718–725. <https://doi.org/10.1111/j.1469-7610.2008.02023.x>
98. van de Groep, I.H., Bos, M.G.N., Jansen, L.M.C., Achterberg, M., Popma, A., & Crone, E.A. (2021) Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others. *Neuropsychologia*.

<https://doi.org/10.1016/j.neuropsychologia.2021.108000>

99. van Domburgh, L., Doreleijers, T. A., Geluk, C., & Vermeiren, R. (2011). Correlates of self-reported offending in children with a first police contact from distinct socio-demographic and ethnic groups. *Child and Adolescent Psychiatry and Mental Health*, 5(1), 22. <https://doi.org/10.1186/1753-2000-5-22>
100. van Domburgh, L., Geluk, C., Jansen, L., Vermeiren, R., & Doreleijers, T. (2019). Antisocial Behavior and Victimization Over 2-Year Follow-Up in Subgroups of Childhood Arrestees. *Journal of Interpersonal Violence*, 34(18), 3780–3806. <https://doi.org/10.1177/0886260516672052>
101. van Domburgh, L., Vermeiren, R., Blokland, A. A. J., & Doreleijers, Th. A. H. (2008). Delinquent Development in Dutch Childhood Arrestees: Developmental Trajectories, Risk Factors and Co-morbidity with Adverse Outcomes during Adolescence. *Journal of Abnormal Child Psychology*, 37(1), 93. <https://doi.org/10.1007/s10802-008-9260-6>
102. van Goozen, S. H. M., & Fairchild, G. (2008). How can the study of biological processes help design new interventions for children with severe antisocial behavior? *Development and Psychopathology*, 20(3), 941–973. <https://doi.org/10.1017/S095457940800045X>
103. van Heukelum, S., Tulva, K., Geers, F. E., van Dulm, S., Ruisch, I. H., Mill, J., Viana, J. F., Beckmann, C. F., Buitelaar, J. K., Poelmans, G., Glennon, J. C., Vogt, B. A., Havenith, M. N., & França, A. S. C. (2021). A central role for anterior cingulate cortex in the control of pathological aggression. *Current Biology*, 31(11), 2321-2333.e5. <https://doi.org/10.1016/j.cub.2021.03.062>
104. Veenstra, R., Lindenberg, S., Verhulst, F. C., & Ormel, J. (2009). Childhood-Limited Versus Persistent Antisocial Behavior: Why Do Some Recover and Others Do Not? The TRAILS Study. *The Journal of Early Adolescence*, 29(5), 718–742. <https://doi.org/10.1177/0272431608325501>
105. Vetter, N. C., Buse, J., Backhausen, L. L., Rubia, K., Smolka, M. N., & Roessner, V. (2018). Anterior insula hyperactivation in ADHD when faced with distracting negative stimuli. *Human Brain Mapping*, 39(7), 2972–2986. <https://doi.org/10.1002/hbm.24053>
106. Yang, Y., & Raine, A. (2009). Prefrontal Structural and Functional Brain Imaging findings in Antisocial, Violent, and Psychopathic Individuals: A Meta-Analysis. *Psychiatry Research*, 174(2), 81–88. <https://doi.org/10.1016/j.psychres.2009.03.012>

SUPPLEMENTARY MATERIALS

Table S1. Attrition analyses

	Excluded Sample (N = 19)	Included Sample (N = 54)	Statistics
No. Persisters	8	12	$X^2 = 2.28, p = 0.13$
Age	M = 26.43, SD = 1.17	M = 26.29, SD = 1.53	T(71) = 0.35, p = 0.72
IQ ¹	M = 93.72, SD = 10.98	M = 102.89, SD = 13.46	T(71) = -2.67, p = 0.009
No. Males	15	46	$X^2 < 0.001, p = 1$
YPI Callous-Unemotional Traits ²	M = 29.82, SD = 4.45	M = 29.83, SD = 4.42	T(71) = -0.003, p = 0.99
YPI Grandiose-Manipulative Traits ²	M = 27.68, SD = 8.08	M = 29.72, SD = 9.27	T(71) = -0.85, p = 0.399
YPI Impulsive-Irresponsible Traits ²	M = 32.21, SD = 9.15	M = 30.20, SD = 7.99	T(71) = 0.91, p = 0.367
YPI total score ²	M = 89.72, SD = 17.51	M = 89.74, SD = 19.23	T(71) = -0.005, p = 0.99

¹ Note that for 17 participants ($n_{\text{Miflcomplete}} = 3, n_{\text{Miflmissing}} = 14$), the IQ tests at T5 were not completed, due to time constraints ($n_{\text{Miflcomplete}} = 3$), or limitations imposed by the COVID-19 pandemic to complete the WAIS in person ($n_{\text{Miflmissing}} = 14$). Therefore, we estimated these scores using multiple imputation (mice package v3.13.0; van Buuren & Groothuis-Oudshoorn, 2011), based on the other variables reported in this table, as well as prior IQ scores (T4).

² Note that for three participants ($n_{\text{Miflcomplete}} = 2, n_{\text{Miflmissing}} = 1$), the YPI was not completed. Therefore, we estimated these scores using multiple imputation, based on the other variables reported in this table, as well as prior IQ scores (T4).

Table S2 Pairwise comparisons using t-tests for confirmatory ROI Analyses: Feedback type (Saliency) for ACC, Left and Right Insula, and dlPFC

	ACC	Left Insula	Right Insula	dlPFC
Pairwise comparison				
Positive vs. Negative	$p = .043$	$p = .0393$	$p = 1.00$	$p = .020$
Positive vs. Neutral	$p = .017^*$	$p = .0019^*$	$p = .036$	$p = .091$
Negative vs. Neutral	$p = 1.00$	$p = .0902$	$p = .177$	$p = 1.00$

* Statistically significant after bonferroni correction for correlated variables (threshold of $\alpha = 0.0287$)

Table S3 ANOVA results for the influence psychopathic traits on saliency difference scores

Predictor	df_{Num}	df_{Den}	F	p
Callous-Unemotional	18	69	1.14	.331
Grandiose-Manipulative	29	58	0.89	.625
Impulsive-Irresponsible	32	55	1.14	.329
YPI	47	40	1.50	.096

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator.

Table S4 MINI diagnoses in the persister and desister groups

MINI Diagnosis	Desister (N = 42)	Persister (N = 12)
Past Major Depressive Disorder	6 (14.29%)	5 (45.45%)
Current Mood Disorder due to physical condition	1 (2.38%)	0 (0.00%)
Past Mood Disorder due to drug use	0 (0.00%)	1 (9.09%)
Agoraphobia	1 (2.38%)	3 (27.27%)
Obsessive-Compulsive Disorder	1 (2.38%)	1 (9.09%)
Generalized Anxiety Disorder	7 (16.6%)	3 (27.27%)
Alcohol dependence / abuse	15 (35.71%)	5 (45.45%)
Drug (non-alcohol) dependence / abuse	10 (23.8%)	7 (63.6%)
Attention Deficit Hyperactivity Disorder	0 (0.00%)	1 (9.09%)
Posttraumatic Stress Disorder	1 (2.38%)	2 (18.18%)

¹ Note that screening for behavioral issues and clinical disorders was performed on the same day of the MRI for the controls (using the diagnostic interview), and for participants from the persister/desister group 0-375 days before the MRI session, with an average of 108 days.

Table S5. DISC and MINI diagnoses across different time points.

	Timepoint									
	T1		T2		T3		T4		T5	
DISC / MINI Diagnosis 1	Desister	Persister	Total	Desister	Persister	Total	Desister	Persister	Total	Total
DBD ²	9	3	12 ³	0	4	4	-	-	-	-
ASPD	-	-	-	-	-	-	0	12	12	12
ADHD	4	3	7	4	0	4	-	-	-	-
PTSS	-	-	-	1	0	1	-	-	-	-
None	29	6	35	29	3	32	-	-	-	-
Missing	-	-	-	8	5	13	-	-	-	-
Age	T0 ⁴	T1	T2	T3	T4	T5				
Mean	10.49	10.9	12.01	13.01	18.11	26.52				
SD	1.43	1.47	1.53	1.58	1.31	1.63				
Min	5.95	6.21	7.47	8.37	14.76	21.6				
Max	11.97	12.78	14.38	15.15	20.38	29.14				
Type of index crime	Desister (n = 42)		Persister (n = 12)		Total (n = 54)					
Arson	7	0	7	1	8	8				
Theft	4	4	8	4	8	8				
Violent crime	7	7	14	7	21	21				

		Timepoint
Public nuisance	11	0
Vandalism	8	4
Missing	5	3
		11
		12
		8

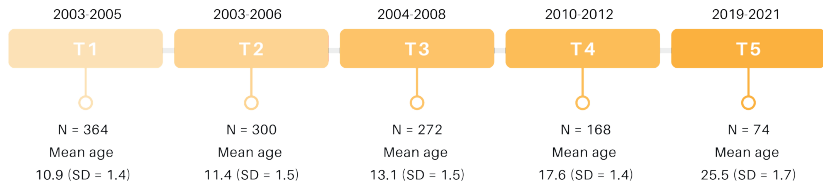
¹ Note that the DISC was only administered at T2 or T3 if it was not administered at T1

² Means participant had at least a DBD diagnosis (DBD CD / ODD / CD + ODD / ADHD + OD, ADHD + CD, ADHD + OD + CD)

³ Note that of these 12 participants, 7 were diagnosed with DBD before the age of 12, and 5 after the age of 12 (during the current study).

⁴ Note that T0 corresponds to the timepoint at which participants were arrested.

A



B

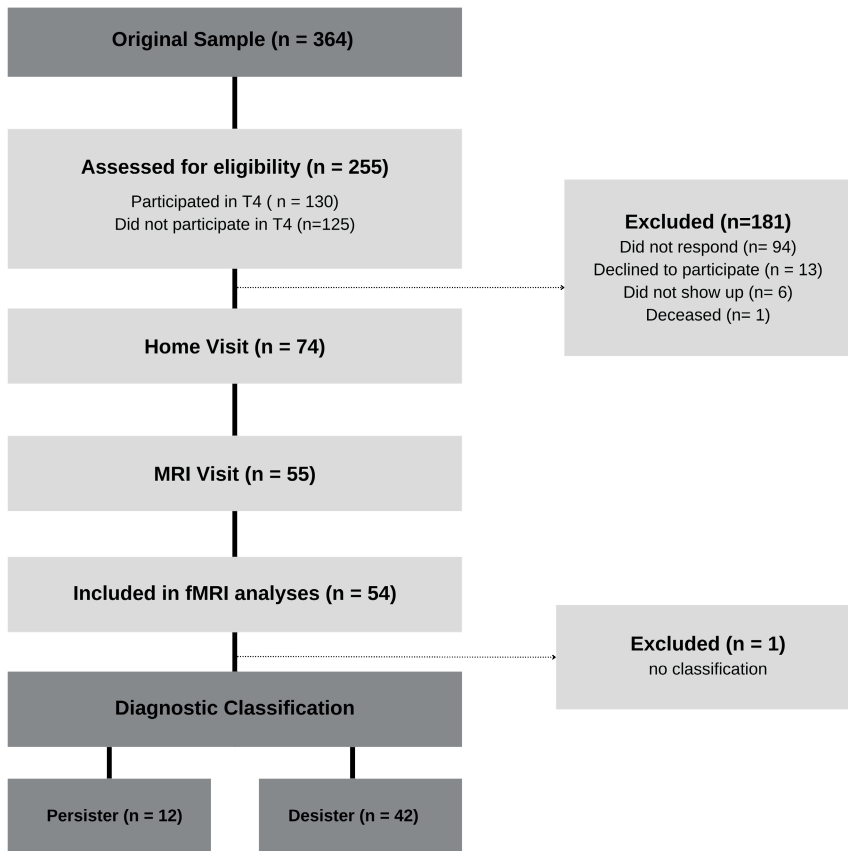


Figure S1. (A) Overview of the longitudinal RESIST study (Research on Individual (Anti-) Social Trajectories), aimed at investigating predictors and consequences of (anti-)social behavior across development. (B) Participant flow chart diagram for the fifth wave (T5) of the RESIST study.

Detailed description ROI Analyses

To test our second hypothesis about neural activation during the receipt of social feedback, next to the whole-brain analyses described above, we performed three ANOVAs with feedback type (Negative, Positive, Neutral) as independent variable, and ROI parameter estimates for the ACC and left and right Insula as dependent variable. To test for interactions with Group status, this categorical variable was included as a covariate. ANOVAs were followed-up by computing pairwise comparisons between conditions and groups. As an additional control analysis, we explored whether results remained the same when Persist and Desist subgroups were combined in one Group (early antisocial experiences). The ANOVAs with Group (i.e., control, early antisocial experiences) did not differ from the ANOVAs with the pre-registered subgroups (i.e., Control, Persistent, Desistant) for any of the behavioral and neural analyses. Therefore, we only report the analyses with the three groups in the result section.

To test how psychopathic trait scores (Grandiose-Manipulative, Callous-Unemotional, Impulsive-Irresponsible, YPI total) influenced neural activation, we calculated saliency difference scores (i.e., parameter estimates for Positive + Negative > Neutral) for the ACC, left and right Insula and subsequently performed ANOVAs with psychopathic trait scores (Grandiose-Manipulative, Callous-Unemotional, Impulsive-Irresponsible, YPI total) as IVs and saliency difference scores as DVs.

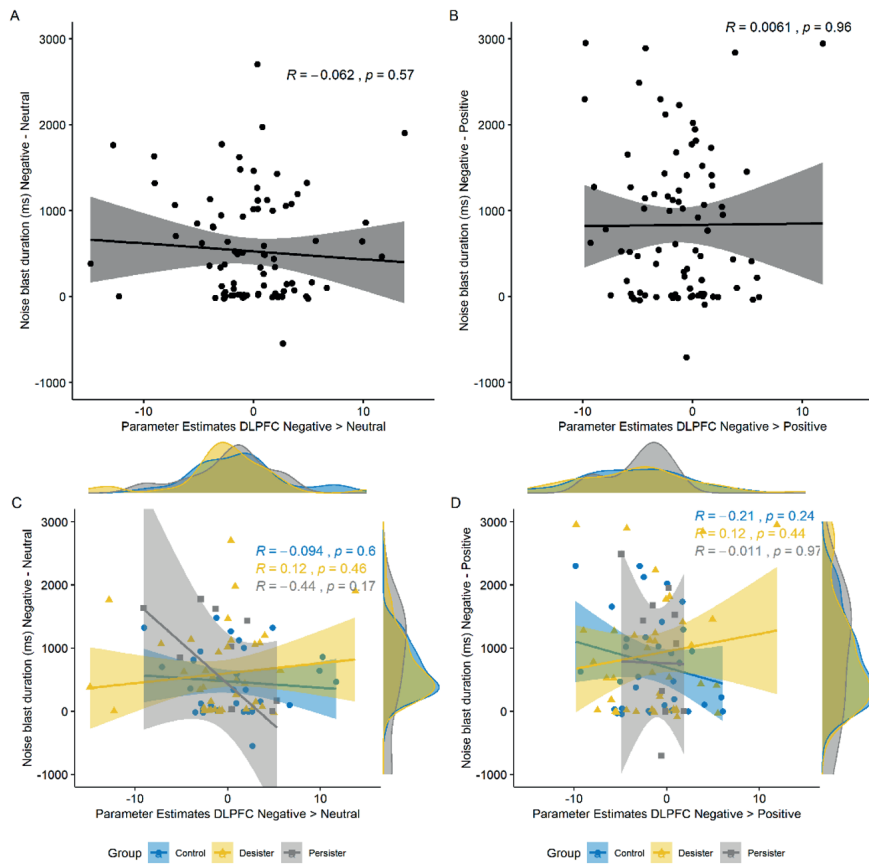


Figure S2. (A) Brain-behavior association between difference scores in dIPFC activity (Negative > Neutral feedback) and noise blast duration (Negative > Neutral feedback), and (B) difference scores in dIPFC activity (Negative > Positive feedback) and noise blast duration (Negative > Positive feedback). Panels C-D show the aforementioned associations split per group.

Acknowledgements

We would like to thank all participants for their participation, and Veronica Baljé, Merel Spaander, Mara Giamboi, Cevdet Acarsoy, Sjors Bootsman, Esmee Haverkort, Angelique de Groot, Eline Vink, Fleur van der Linden, Hannah Woerdman, Julie Tournier, Lenore van Dijk, Mark Cnossen, Merlijn Smis, Marjolein Beemster and Bram Zwanenburg for their help with data collection.



General Discussion

Part of this chapter appeared in modified, preprinted form as: van de Groep, I. H., Bos, M., Popma, A., Crone, E., & Nauta-Jansen, L. (2022). A neurocognitive model of early onset persistent and desistant antisocial behavior in early adulthood. <https://doi.org/10.31234/osf.io/sw73u>

Early adulthood has long been recognized as a turning point for the development of antisocial behavior (Bersani & Doherty, 2018; Moffitt, 1993, 2018). Various studies suggest that during this developmental period, many youth desist from antisocial behavior, even if they did show an early childhood-onset, a factor that has often been associated with various negative life outcomes and continuous patterns of severe antisocial behavior (Moffitt, 1993, 2018; Bersani and Doherty, 2018). Still, a small group of youth with an early onset does indeed persist in antisocial behavior into early adulthood (Moffitt, 1993; 2018; Bersani and Doherty, 2018). Although it is evident that these developmental differences arise, this thesis addressed the question *why* and *how* they arise. One way to start answering these complex scientific questions was by studying neural functional mechanisms that were expected to differentially characterize these persistent and desistant developmental groups in early adulthood, compared to young adults who resist antisocial behavior throughout the lifespan. Understanding these mechanisms is important to inform the development of effective interventions, especially in light of the high treatment resistance associated with persistent antisocial behavior (Gatzke & Raine, 2000) and to help reduce the large costs associated with persistent antisocial development for youths themselves, their victims and society at large (Foster, 2010; Knapp et al., 2002; Romeo et al., 2006; Scott et al., 2001). To understand how young adults with and without antisocial tendencies navigate their rapidly changing social world, the study of neural functional mechanisms focused on important developmental tasks in early adulthood, such as gaining and using knowledge, skills, and self-understanding to balance between environmental constraints and one's personal goals (Arnett 2000, Arnett et al. 2007).

The first aim of this dissertation, derived from prior meta-analyses (Lieberman et al., 2019), was to determine whether there is a common neural substrate for evaluating the self from multiple perspectives in typically developing young adults, and to assess the robustness of functional MRI tasks assessing self-evaluation and regulation (Elliot et al., 2020; Kragel et al., 2021). This first step allowed me to establish a valid basis for the second and main goal of this dissertation, which was to study (1) how young adults with and without a history of antisocial behavior and varying levels of psychopathic traits evaluated themselves, learned *about* themselves and *for* themselves and others simultaneously, and acted upon social feedback information provided by others and (2) the neural underpinnings of these social-cognitive processes and behaviors. Throughout this dissertation, I addressed these aims using a combination of scientific methods including self-report questionnaires, diagnostic interviews, experimental tasks and functional brain activation. Moreover, I used two complementary phenotypic approaches to characterize antisocial behavior in early adulthood: (1) a longitudinal, categorical approach (by comparing persistent, desistant and control groups) and a (2) cross-sectional, dimensional approach (by considering continuous levels of psychopathic traits). In this final chapter, I summarize and discuss the results of the empirical chapters, discuss the scientific and clinical implications, and consider suggestions for future research.

Summary and discussion of main findings

In **Chapter 2**, I describe a neuroimaging study in which I investigated whether a common neural substrate in the medial prefrontal cortex (mPFC) underlies self-evaluation from the perspective of self (self-concept appraisal) and others (social evaluation) in 40 young adults (aged 18-30). To this end, young adults performed two fMRI tasks, the Self-Concept Task and the Social Network Aggression Task (SNAT), while they underwent functional magnetic resonance imaging (fMRI). In the Self-Concept task (van der Crujisen et al., 2018), participants evaluated trait statements in two conditions. In the self-condition, they evaluated to what extent positive and negative trait statements across two different domains (prosocial and physical appearance) described themselves on a four-point scale. In the control-condition, participants categorized similar positive and negative trait statements into different domains (Prosocial, Physical appearance, Academic or Don't know). In the SNAT (Achterberg et al., 2016), participants received social feedback from same-aged peers, that could be positive, negative or neutral and subsequently got the opportunity to retaliate by sending a noise blast towards the peer, where a louder noise blast signaled more aggression.

Using previously used tasks allowed me to assess the robustness and reproducibility of prior behavioral and neurobiological findings on self-evaluation using the Self-Concept Task and SNAT (Achterberg et al., 2016; van der Crujisen et al., 2018), which is important in light of recent concerns about the reproducibility of fMRI results (Elliot et al., 2020). I was able to replicate the previously obtained behavioral patterns, which showed that during internal self-evaluation, young adults are more likely to evaluate positive trait statements (vs. negative) as being more applicable to themselves, and evaluated prosocial traits to be more applicable than physical traits (van der Crujisen et al., 2018). When considering social evaluations, in line with prior research, I found that individuals give the loudest noise blasts following negative feedback, less so after neutral feedback and least following positive feedback (Achterberg et al., 2016, 2017, 2018, 2020; Dobbelaar et al., 2021, 2022). Using Region-of-Interest (ROI) analyses, I also examined whether previously obtained neural activity patterns could be replicated, which was the case for the majority of the findings. For instance, during social feedback processing, young adults showed the expected higher activity in anterior insula and ACC for positive and negative feedback relative to neutral feedback (Achterberg et al., 2016), which fits with the idea that these brain regions signify salient social information. However, I could not replicate previously obtained neural activity patterns during self-evaluations showing more activity in the mPFC during the self-condition compared to the control condition (van der Crujisen et al., 2018) – although I did find differences when comparing positive trait statements to control statements, which fits with the idea that more applicable positive traits may evoke more mPFC activity compared to less applicable negative traits (D'Argembeau, 2013).

A further, explorator aim of the study described in **chapter 2** was to test whether the different types of social feedback (positive, negative and neutral) were related to different

patterns of neural activation during retaliatory responses. I found that receiving positive feedback resulted in more activity in the dorsolateral PFC (dlPFC), compared to negative and neutral feedback, indicating that this brain area might be important for the regulation of retaliatory responses, a finding that has now been corroborated by later studies using the same SNAT paradigm (Dobbelaar et al., 2022, see also **chapter 3**).

On a group level, I found overlap in the anterior/rostral mPFC for positively valenced evaluations from the perspectives of self and others. However, on an individual level, when considering correlations between tasks, I found no evidence for overlap in the mPFC. In **chapter 2**, I argued that this discrepancy fits with other research showing that examining functional brain activation can reliably reveal mechanistic insights at the group level, but not on the individual level in the extent to which participants engaged these regions across tasks, as was evident by a relatively low individual level test-retest reliability (Elliott et al., 2020).

Subsequently, in **Chapter 3**, I focused on how young adults with and without a history of antisocial behavior and with varying levels of psychopathic traits evaluated themselves, and whether they form similar or diverging neural responses during self-evaluation across domains and contexts. For this purpose, a subset of young adults with an early-onset persistent or desistant history of antisocial behavior ($n = 54$; aged 18-30) from the RESIST Cohort study and the sample of typically developing young adults ($n = 40$) described in **chapter 2**, performed the Self-Concept Task described in **chapter 2**. The study described in **chapter 3** revealed increased activity in mPFC for self-appraisals, consistent with prior studies (Denny et al., 2012; van der Crujisen et al., 2018) and showed that the same brain regions are recruited for self-appraisals across groups with various histories of antisocial behaviors. Across the total sample, psychopathic traits (combined total of Callous-Unemotional traits, Grandiose-Manipulative traits and Impulsive-Irresponsible traits) were associated with more negative and less positive self-appraisals in the prosocial domain, and not in the physical appearance domain. In terms of neural activity, Callous-Unemotional traits were associated with less anterior mPFC activity during general self-evaluations, which may suggest differences in how individuals with high levels of these traits process abstract information when thinking about themselves, potentially to maintain stability in their thoughts based on existing beliefs about the self (Zamani et al., 2022). Taken together, these findings suggest that the super-ordinate construct of psychopathy is associated with domain-specific self-appraisals, while specific sub-dimensions (e.g., Callous-Unemotional traits) show distinct neurobiological functional alterations across domains – highlighting that considering both total levels of psychopathic traits and specific subdimensions in future research may reveal more insights into the etiology and complex pathways related to antisocial behavior.

In **chapter 4**, I focused on how young adults learned *about* themselves and acted upon social feedback information provided by others. More specifically, I studied neural and behavioral responses to social feedback in the same groups of young adults with

childhood-onset persistent antisocial behavior, childhood-onset desistant antisocial behavior and no history of antisocial behavior, and varying levels of psychopathic traits described in **chapters 2-3**, using the SNAT described in **chapter 2**. Not only persisters, but both groups of early adults with a history of prior antisocial behavior (persisters and desisters) showed increased Anterior Insula activity during feedback processing, regardless of feedback type (negative, neutral, positive), compared to the typically developing controls. Possibly, this finding reflects neural hypersensitivity to salient and self-relevant social feedback information, and more broadly, difficulties in the ability to differentiate between social feedback cues (Kawamoto, Nittono, et al., 2015; Kawamoto, Ura, et al., 2015). An additional finding was that increased activity in the dlPFC during general feedback processing was specific to the desisting group. This increased dlPFC activity likely reflects attentional processes in response to changing task demands (context-dependent changes in feedback presentation between trials; Bertsch et al., 2020; Niendam et al., 2012), which support cognitive and emotional regulation of subsequent behavior.

During retaliatory responses, I observed that young adults with a persistent developmental trajectory of antisocial behavior showed similar levels of noise blast aggression as the other two groups following negative feedback. However, young adults with a persistent antisocial development did not differentiate in their behavioral responses and showed equally aggressive responses regardless of feedback type, unlike controls and those with a desistant antisocial trajectory. Moreover, after receiving positive feedback, young adults with a persistent antisocial trajectory showed less dlPFC activity during their behavioral response (noise blast delivery), compared to the other two groups. The findings further revealed that individuals with a desistant antisocial trajectory showed specific behavioral and neural mechanisms that may explain why they manage to successfully desist from antisocial behavior and show adaptive behavior in changing social contexts. More specifically, when examining retaliatory behavior, I found a positive association between aggression regulation following positive feedback and activity in the ACC and dorsal striatum (caudate and putamen) during the behavioral response (noise blast), which was strongest in the desistant antisocial trajectory. When considering individual differences on a trait level, I found that higher levels of psychopathic traits were associated with higher retaliatory behavioral responses, but I did not find evidence that psychopathy or its different sub-dimensions were associated with neural activity during feedback processing or retaliatory responses.

Taken together, these findings suggest that persistent antisocial development might be associated with difficulties in differentiating between social feedback types, both while processing this feedback and during subsequent behavioral responses. In contrast, while individuals with desistant developmental patterns might likewise show hypersensitivity to social feedback cues, they do differentiate in their subsequent retaliatory responses, on a behavioral and neural level.

Finally, in **chapter 5**, I studied whether high levels of psychopathic traits were associated with reinforcement learning differences *for* self and others in different positive (reward-related) and negative (loss-related) social contexts, and corresponding alterations in neural activity in the striatum, in the same high-risk group of young adults with prior histories of antisocial behavior ($n = 53$) described in **chapter 3 and 4**. To this end, I adapted a social probabilistic learning fMRI task (Christopoulos & King-Casas, 2015), in which participants were required to learn in different contexts, characterized by different outcome contingencies, resulting in rewards or losses, and influencing outcomes for themselves and an unknown other simultaneously. In line with prior reinforcement learning studies, I found that young adults learned faster when learning had consequences for themselves compared to others, and that they learned better in a reward than in a loss context (Westhoff et al., 2021; Lockwood et al., 2016). Moreover, this study is amongst the first to show that higher levels of psychopathic traits were associated with more self-other neural differentiation in the striatum, which might indicate that individuals who score high on psychopathic tendencies show more differentiation in reward value compared to individuals who score lower on psychopathic traits (Overgaauw et al., 2020; Pujara et al., 2013; Hosking et al., 2017; Buckholtz et al., 2010; Bjork et al., 2012; Lockwood et al., 2017). Together, these findings fit well with the idea that learning depends on social contexts and individual differences in psychopathic traits, although replication is warranted.

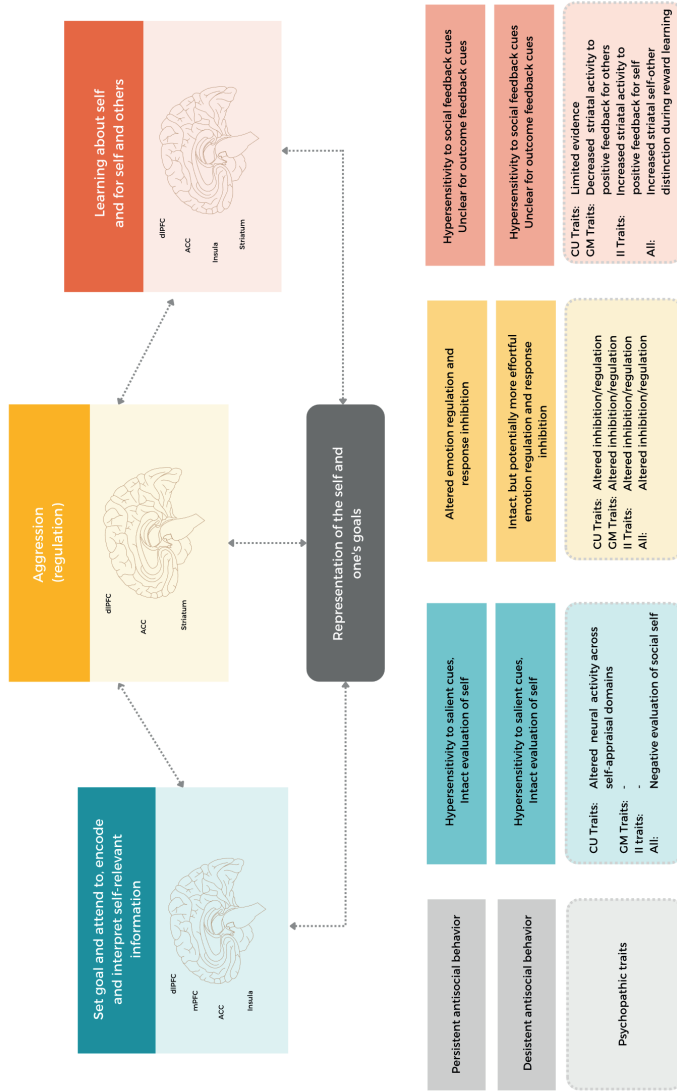


Figure 1. A neurocognitive working model of early onset, persistent and desistent antisocial behavior in early adulthood, based on the summary and discussion of findings in the current dissertation. The working model includes several neurocognitive functions that can be used to understand aggressive, antisocial behavior in current situations (i.e., state-related processes) throughout the course of development (by incorporating learning, or how social information is updated and integrated to guide future behavior and neural processing) and in changing social contexts. Importantly, the model incorporates developmentally salient features of early adulthood, related to understanding and monitoring one's self-traits and goals. (A) A heuristic depiction of the described functions and their development (i.e., (1) goal setting and attending to, encoding and interpreting self-relevant information (Social Information Processing (SIP) model [see General Introduction Figure 3] steps 1-3; input-related steps), (2) monitoring and learning (updating and integration of) information) (SIP steps 3-6, output related steps) and (3) Aggression (regulation) (SIP step 5-6). These functions rely on ongoing development in frontal and subcortical limbic brain areas across early adulthood, including the *dlPFC*, (*vmPFC*, *ACC*, *Insula*, and *Striatum*). (B) Individuals with an early onset, persistent antisocial development show some similarities, but also marked differences in neurocognitive functioning to young adults with a desistent trajectory. In addition, both the overall construct of psychopathy and separate dimensions of psychopathic traits may help to differentiate between heterogeneous antisocial expression in different social contexts and throughout development.

GENERAL DISCUSSION

Characterizing antisocial development in early adulthood

In the current dissertation, I found evidence that a small group of individuals within the RESIST early onset high-risk cohort sample persisted in antisocial behavior into early adulthood (12 out of 54 of the MRI sample [22%], and 20 out of 74 in the full cohort assessed in early adulthood [27%]), while a larger part desisted from antisocial behavior (42 out of 54 of the MRI sample [78%], and 54 out of 74 in the full cohort assessed in early adulthood [73%]) (**chapter 3-4**). Other studies have estimated the prevalence of persistent antisocial behavior to range between 1.6% to 29.1% (Eme, 2020), with discrepancies likely being caused by differences in operationalization and measurement of antisocial development (Eme, 2020). For instance, in the current dissertation, I defined desistance based on the presence of clinical, psychiatric symptoms, while other studies have focused on offending patterns based on police registrations (Bersani & Doherty, 2018; Van Hazebroek et al., 2019). Possibly, estimates based on diagnostic interviews and self-report measures may be higher because they are more likely to capture those individuals who show antisocial behavior, but are not arrested for offenses by the police. However, despite the discrepancies in the exact estimated prevalence, both approaches seem to support the idea proposed in the developmental taxonomy (Moffit, 1993, 2018) that persistent antisocial behavior in early adulthood is relatively rare, and not the norm (Moffit, 2018, Bersani & Doherty, 2018). Similar to other longitudinal cohort studies, the findings in this dissertation also highlight that persistent antisocial behavior was associated with higher levels of psychopathic traits (Moffit et al., 2002; Eme, 2020). The findings described in this dissertation further suggest that antisocial behavior in early adulthood might be characterized by alterations in self-relevant and feedback-related information processing and aggression regulation (**chapter 3-5**). The behavioral differences I observed were associated with functional alterations in several frontolimbic brain regions, including the (v)mPFC, dlPFC, ACC, Insula, and Striatum (**chapter 3-5**; see Figure 1), although other brain areas might also be involved. The neurocognitive difficulties in persistent antisocial individuals seem to be characterized by a limited capacity to differentiate between differently valenced cues and to adapt their behavior to specific and changing social contexts (**chapter 4**).

In contrast, the majority of young adults with an early onset of antisocial behavior desisted after childhood (Odgers et al., 2008; Bersani & Doherty, 2018), in line with recent literature and the revised taxonomy (Bevilacqua et al., 2018; Moffitt, 2018; Moffitt et al., 2002; Monahan et al., 2013; Odgers et al., 2007, 2008), although it should be noted that some of them experienced internalizing and substance use problems (**chapter 3-4**). More specifically, 16.6% of the desister group satisfied the criteria for generalized anxiety disorder, 35.7% reported alcohol dependence or abuse, and 23.8% reported non-alcohol related drug dependence or abuse (**chapter 3,4**). This pattern fits with results from other

developmental studies (Odgers et al., 2008; Moffit et al., 2002) showing heterogeneity within trajectories of early-onset desistant antisocial behavior in early adulthood, where a small part shifts from showing antisocial behavior to others forms of functional and social maladjustment, like internalizing problems and social isolation (Odgers et al., 2008; Moffit et al., 2002). Interestingly, previous studies have shown that such internalizing tendencies might act as a buffer against future antisocial behavior (Moffit et al., 2002). Findings from the current dissertation further suggest that part of desisters' success to refrain from antisocial behavior, at least relative to individuals who persist in antisocial behavior, may also be explained by their ability to regulate aggression and monitor social information, even though they experience similar neural hypersensitivity to salient, self-relevant social feedback information (**chapter 4**).

With regard to such behavioral regulation, an important question that remains unanswered is which specific aspect of cognitive control (i.e., emotion regulation or response inhibition (Bertsch et al., 2020) is more important in determining the observed differences in aggressive behavior between developmental groups. Given that the SNAT paradigm does not allow us to dissect the exact cognitive control process that potentially cause these differences, it remains unclear whether the diverging patterns are the result of differences in inhibitory control, emotion regulation, or both (Bertsch et al., 2020). Based on my findings (**chapter 4**), I hypothesize that young adults with a persistent antisocial history might have problems with both emotion regulation (e.g. downregulating their context-independent emotional and neural hypersensitivity) and response inhibition (e.g. failure to inhibit and adapt their prepotent response to react aggressively, regardless of social context). Conversely, individuals with a desistant antisocial trajectory may show similar difficulties in initial emotional responses (i.e., emotional and neural hypersensitivity), but more successful emotion regulation (e.g. attention to changing task demands and reappraisal of salient information) and response inhibition (e.g. inhibiting responses when such behavior is more appropriate, such as following positive or neutral feedback) (Gross & Levenson, 1993).

However, it should be noted that the current findings (**chapter 3-5**) do not rule out the possibility that antisocial and psychopathic behavior is not necessarily due to an *impaired ability* to regulate behavior, but might instead reflect *reduced motivation* to regulate, monitor or change behavior after receiving feedback (**chapter 4-5**) (Drayton et al. 2018; Foulkes et al., 2014; Gaule et al., 2021.; Viding & McCrory, 2019), or different motives to evaluate and monitor the self in different social contexts (Neiss et al., 2006) (**chapter 3**). Regarding the former, young adults with persistent antisocial development and high levels of psychopathic traits may simply be motivated to use their dominant response set (showing aggression) (**chapter 4**), perhaps even after evaluating alternative response options (i.e., during SIP step 5; Crick & Dodge, 1994). Moreover, the finding described in **chapter 5** that higher psychopathy was associated with more self-other neural differentiation in the striatum, often considered the reward area of the brain (Delgado,

2007; Knutson et al., 2001), might suggest that young adults with high levels of psychopathic traits derive less subjective reward value from choices that benefit others and more subjective reward value for choices that benefit themselves (Overgaauw et al., 2020; Lockwood et al., 2017). Possibly, this subjective value distinction may reinforce their motivation to act self-interested and in disregard of others (Foulkes et al., 2014; Viding & McCrory, 2019; Drayton et al., 2018). A recent study suggests that such devaluation in individuals with high levels of psychopathic traits might occur and manifest itself by exerting less effort when outcomes affect others, instead of oneself (Lockwood et al., 2017). Further evidence for reduced prosocial motivation comes from studies showing that psychopathic individuals can take the perspective of others, but often fail to show the prepotent tendency or motivation to do so (Drayton et al., 2018; Viding & McCrory, 2019), especially when considering the perspective of others is not important for the goal(s) they want to obtain (Drayton et al., 2018; Baskin-Sommers & Brazil, 2022).

Regarding evaluation and monitoring the self, young adults scoring higher on psychopathic traits might use different motives when evaluating and updating their self-concept (Neiss et al., 2006). In particular, our finding that higher psychopathy was associated with more negative self-concept appraisal in the prosocial, but not physical appearance domain (**chapter 3**), suggests that young adults scoring high on psychopathic traits might have a realistic view of their prosocial self (Viding & McCrory, 2019). Possibly, then, their self-evaluations might be biased towards the motivation to achieve an realistic, accurate image of themselves (known as the self-assessment motive, Neiss et al., 2006). However, a future research should first rule out the alternative explanation that these individuals may have difficulties updating beliefs about the self (Elder et al., 2021; Hamilton et al., 2015; Korn et al., 2012; Rodman et al., 2017). Instead, individuals scoring lower on psychopathic traits might be primarily motivated by other self-relevant goals like maintaining positive self-image (i.e., self-enhancement) and a consistent and coherent view of the self (i.e., self-verification) (Neiss et al., 2006; Doerfler et al., 2021). Taken together, future research should further entangle whether difficulties in the *ability* or *motivation* underlie the regulation of antisocial behavior in early adulthood, paying close attention to the potential timing and duration of - and interaction between - neurocognitive processes and behaviors (Sheppes & Gross, 2011)

Understanding heterogeneity in antisocial expressions in early adulthood

The current findings clearly highlight that the behavioral and neurobiological frameworks described in the discussion, like the developmental taxonomy of antisocial behavior (Moffitt, 1993, 2018) and social information processing model (SIP, Crick & Dodge, 1994), provide a somewhat simplistic view of antisocial behavior in early adulthood, given the (1) observed heterogeneity in behavior and neural responses and the (2) observed behavioral

and neural differences associated with antisocial and psychopathic tendencies that were dependent on valence and other salient aspects of the social context (**chapter 3-5**).

Although some behaviors were altered in both persistent antisocial behavior and psychopathy (aggression regulation; **chapter 4**), other behavioral and neural differences seem specific to individual differences at the trait level (e.g. self-evaluation, reinforcement learning; **chapter 3, 5**). This observation fits with earlier research suggesting that using a dimensional, continuous approach may yield higher behavioral and neural specificity than categorical approaches alone (Viding et al., 2012; Carré et al., 2013; Cohn et al., 2015; Hauser et al., 2023), often even more so when separately considering different but correlated psychopathic trait sub-dimensions (Lilienfeld, 2018; Hauser et al., 2023; **chapter 3, 5**). This specificity has been found when considering the entire psychopathic trait spectrum, including the lower and higher ends (Andershed et al.; 2002; Viding et al., 2012; Cohn et al., 2015; Carré et al., 2013; Seara-Cardoso & Viding, 2015), in both the general population and high-risk samples (Andershed et al., 2002, Lockwood et al., 2017; Cohn et al., 2015; Seara-Cardoso & Viding, 2015; **chapter 3, 5**). However, for the behavioral association between psychopathic traits and self-concept appraisal observed in **chapter 3**, I argued that part of this increased specificity might be due to closer conceptual and methodological (i.e., common-method variance) overlap, highlighting the importance of combining several scientific methods to overcome possible limitations imposed by one of them.

The findings from the current dissertation also fit with the idea that considering affective components of social information processing (Gaule et al., 2021), such as valence (Lerner and Arsenio, 2000; Smeijers et al., 2020) can provide more specificity to understand potential neural and behavioral differences in the ability or motivation to evaluate, monitor, or learn social information and subsequently use this information to make decisions and to regulate behavior - and should therefore be incorporated into the SIP model. Indeed, various studies show that cognitive processes, like perspective-taking and cognitive empathy, are generally intact in individuals with persistent antisocial behavior and psychopathy (Gaule et al., 2021; Radke et al., 2013; Viding & McCrory, 2019), while difficulties arise in tasks employing both cognitive and affective social information processes (Viding & McCrory, 2019). In the studies described in this dissertation (**chapter 2-5**), we also found evidence for such affective-cognitive social information processing differences. For instance, in **chapter 4**, delineating between positive, negative and neutrally valenced feedback revealed that the increased insula and dlPFC activity during feedback processing in both the persisters and desisters groups and the desister group respectively, were not specific to the valence of the stimuli, but were general (i.e., similar for all feedback types). Combined with the finding that typically developing young adults only show increased insula activity for positive and negative feedback, but not for neutral feedback (**chapter 2, 4**), this finding might thus suggest that youth with prior antisocial experiences may particularly show feedback processing differences in ambiguous social

contexts (Brennan et al., 2018). Likewise, in **chapter 2-3**, considering both positive and negative trait statements allowed for increased specificity to detect (and replicate) different behavioral and neural signatures associated with valence-specific self-appraisals (van der Crujisen et al., 2018), and uncover diverging behavioral associations with global levels of psychopathic traits for positive and negative trait statements in the prosocial domain (**chapter 3**). Finally, in **chapter 5**, I showed that learning and neural responses to outcome feedback are specific for positive and negative outcomes, and for positive and negative feedback (van den Bos, 2009), and that associations between psychopathic traits and neural activity in the ventral striatum were specific for positive feedback in the reward context.

Taken together, based on the findings described in the current dissertation (**chapter 2-5**), and the general discussion of these findings (**chapter 6**), I propose that studying trait level characteristics, like individual differences in psychopathic traits (**chapter 3-5**; Lockwood et al., 2017) and motivation (**chapter 6**), and emotion- and context- specific characteristics, such valence (**chapter 3-5**; Gaule et al., 2021; Lemerise and Arsenio, 2000; Smeijers et al., 2020), provides a promising avenue to differentiate between heterogeneous antisocial and psychopathic expressions in different social contexts and throughout development.

Clinical implications

The current dissertation mainly described fundamental research focused on uncovering neurobiological mechanisms associated with antisocial and psychopathic tendencies in early adulthood, and thus does not allow for direct implementation in clinical practice. However, I offer recommendations to facilitate communication with and integration of fundamental scientific findings in society (Vandenbroucke et al., 2021; see box 1). Moreover, based on the current findings and the general discussion, several preliminary recommendations can be made to inform the use of personalized approaches in treatment and interventions. In particular, the findings from the current dissertation highlight important mechanistic building blocks that can be used for prevention, intervention, treatment and support (Odgers et al., 2008).

For instance, in light of the identified mechanisms that allow individuals with a desistant trajectory to successfully adapt their retaliatory behavior in social contexts, prevention and treatment might focus on a combination of regulatory strategies and skills that focus on reappraisal and increasing sensitivity to context-specific social information, and to support behavioral regulation (Bertsch et al., 2020), while considering participants' motivation (Foulkes et al., 2014; Viding & McCrory, 2019; Drayton et al., 2018; Gaule et al., 2021; Hawes et al., 2014). Ideally, given the costly, pervasive and heritable nature of persistent antisocial behavior (Dodge & McCourt, 2010), prevention efforts should start early in life and continue intensively throughout development (Dodge & McCourt, 2010; Hawes et al., 2014). Several scientific studies also support the notion that prevention

efforts should especially focus on individuals with a high risk for negative life outcomes (Dodge & McCourt, 2010; Moffit, 2018; Beelmann & Raabe, 2009), like those displaying early indications of psychopathic traits (Waller & Wagner, 2019; Hawes et al., 2014). Recent evidence further suggests that it is important to consider not only individual characteristics, but also focus on contextual factors that may influence how biological vulnerabilities are expressed throughout development (Hawes et al., 2014, Dodge et al., 2009; Byrd et al., 2014). For instance, training of skills and strategies to advance social learning and behavioral regulation might be more successful if parents or peers are also involved (Hawes et al., 2014). Likewise, findings from the current dissertation suggest that prevention efforts requiring learning and behavioral change might benefit from knowing the types of feedback or contexts that allow individuals with persistent antisocial tendencies and higher psychopathic traits to learn best (Westhoff et al., 2021; Lockwood et al., 2016). In particular, for individuals who are at risk for developing persistent antisocial behavior, it might be important to communicate feedback in a manner that minimizes the probability that they interpret this feedback as ambiguous and hostile (Brennan et al., 2018; Crick & Dodge, 1994). Likewise, prevention and intervention strategies for individuals with high levels of psychopathic traits might benefit from using consistent reward-related reinforcement strategies (Byrd et al., 2014).

Box 1 – Communicating with and integrating scientific findings into society

To integrate fundamental scientific findings in clinical practice, policy, law enforcement, and society more broadly, effective science communication is important (Vandenbroucke et al., 2021). Most science communication has been based on the assumption that scientists need to address ignorance, or a knowledge deficit that stakeholders and the general public might have (Bubela et al., 2009). Arguably, scientists have an important role in society, which includes disseminating their latest findings and to support the development of scientific literacy across different audiences (Bubela et al., 2009). However, translation and consolidation of scientific knowledge are likely more effective if youth, clinical experts, policy makers, law enforcers and researchers from different disciplines also actively participate alongside scientists in this process (Vandenbroucke et al., 2021; Bubela et al., 2009). In particular, such a co-creative process may help all those involved to develop more efficacy, motivation, skills and knowledge, which may in turn benefit both their work and future communication efforts (Vandenbroucke et al., 2021; Bubela et al., 2009), result in inclusion of more diverse and inclusive populations (Green, van de Groep et al., 2022) and coincidentally also increase trust in science (Bubela et al., 2009). The effectiveness of science communication also depends on characteristics of the specific audience, highlighting that using a mix of different communication methods and media might be most beneficial (Bubela et al., 2009). One way to ensure the sustainability and effectiveness of science communication is to appoint dedicated 'Knowledge-brokers' (Vandenbroucke et al., 2021), who are trained in fundamental science and have the ability to clearly communicate and integrate findings.

Future Research

Based on the main findings of this dissertation and the general discussion, I highlight three important and related considerations for future research on antisocial and psychopathic development in early adulthood.

First, it is important to gain a more comprehensive understanding of differences (in stability) between – and changes within – antisocial developmental pathways. To this end, future research should go beyond cross-sectional research in early adulthood, and conduct longitudinal studies across development to help identify the onset, rate and consistency of the developmental processes of interest and corresponding neural underpinnings. Longitudinal research may also reveal insights into the causality and temporal order of developmental and life events, and thus provide starting points to understand potential mechanisms of change – which is essential to develop suitable and personalized interventions. Importantly, pointers for change may also be provided by focusing on more immediate, short-term and dynamic adaptations in social contexts (Flechsenshar et al., 2022). For instance, focusing on time-related changes within tasks, and trial by trial changes may further illuminate how the brain computes processes underlying antisocial behavior (Bertsch et al., 2020; Pauli & Lockwood, 2022; Lockwood et al., 2016; 2017), and facilitates the development of both social competence and personal goal attainment (Flechsenshar et al., 2022). Taken together, longitudinal studies and trial-based analyses may shed new light on the exact timing of SIP difficulties, and opportunities for both immediate (short-term) and developmental (long-term) adaptations in social contexts (Flechsenshar et al., 2022).

Second, the studies described in this dissertation clearly stress the importance of considering the complex interaction between characteristics of the social context, the aggressive or antisocial response and individual characteristics to understand the neurodevelopment of antisocial behavior. Different aspects of the social context (e.g. the specific trigger of antisocial behavior - social rejection, frustration or threat) and how many people are involved (Bertsch et al., 2020; Rappaport & Barch, 2020), determine how and why people act in a specific manner (e.g. showing reactive aggression in response to social rejection, based on the affordances of the situation), in interaction with individual characteristics (e.g. achieving self-relevant goals in line with one's self-concept, goal representations and personality). Moreover, the adaptivity of antisocial behaviors and tendencies likely changes over time and as social norms and contexts change. In particular, when growing up in a hostile environment, antisocial and psychopathic tendencies might be adaptive - at least in the short term - but they cease to do so when a variety of different contexts with changing demands and affordances are encountered in adolescence and young adulthood. Thus, an important avenue for future research is to incorporate different social-cognitive, neurobiological and environmental measures into one (longitudinal) approach (Brazil et al., 2018; van der Wal et al., 2021), to do justice to the complexity of these factors and their interplay. Such integrative approaches require a large amount of

data, and thus, studies with large(r) sample sizes. Given that antisocial populations are difficult to recruit and retain, especially in neuroimaging research, these approaches will likely involve the merging of datasets within consortia (Brazil et al., 2018; see for example the ENIGMA project, Thompson et al., 2017). However, it is important to note that although this approach may help to understand the heterogeneity in the display and development of antisocial behavior, much about the (development of) functional mechanisms of interest underlying antisocial development is still unclear. Thus, such approaches should be complemented with targeted (fMRI) studies that aim to identify or clarify (changes in) important functional, computational, and behavioral mechanisms - or situational and personal characteristics. Ultimately, a combination of these approaches, which relies on their combined strengths, is most likely to advance our understanding of the development of antisocial behavior (Brazil et al., 2018).

Third, it is important to consider the arguably complex role of psychopathic personality traits in the neurodevelopment of (persistent) antisocial behavior in more detail. The findings from this dissertation highlight that there might be both overlapping and distinct features of persistent antisocial behavior and psychopathic trait dimensions, that can be used to differentiate between diverse impaired functional and neurobiological social information processing mechanisms (Pauli & Lockwood, 2022; Lilienfeld, 2018), and help further explain heterogenous pathways in antisocial development. Between the specific psychopathic dimensions, current findings (**chapter 3, 5**) and recent literature provide evidence for both overlapping and distinct neurobiological mechanisms (Gillespie et al., 2022, Garafalo, et al., 2018). To improve our understanding in the differences and overlap between persistent antisocial development and psychopathic traits, it is important to also consider how changeable and stable both psychopathic traits and antisocial behavior are – throughout development and across different social contexts. Both persistent antisocial behavior and psychopathic traits are assumed to involve a certain stability, and traditionally, psychopathic traits have been assumed to be relatively insensitive to change throughout development (Nentjes et al., 2022). Although for most individuals, psychopathic traits remain quite stable during the transition from adolescence to young adulthood (Lee & Kim, 2021), there is also evidence that this is not always the case, and not always the same for all psychopathic trait subdimensions (Lee & Kim, 2021; Lynam et al., 2007). In particular, Grandiose-Manipulative traits seem susceptible to change (Lee & Kim, 2021). Moreover, it should be noted that associations with neurocognitive functional mechanisms are not always present across different social contexts (Nentjes et al., 2021). A related question is how central the studied difficulties are to the persistence of psychopathic and antisocial development (Gillespie et al., 2022; Garafalo et al., 2018; Bersani & Doherty, 2018). Thus, future research should examine this centrality, and consider how environmental influences impact the nature and stability of antisocial and multidimensional psychopathic traits throughout development (Blair, 2013), across different social contexts.

Conclusion

To conclude, this dissertation showed that it is important to consider developmental aspects of early adulthood to understand functional and neurobiological mechanisms that underlie different developmental trajectories of antisocial behavior. In particular, this dissertation offers a comprehensive perspective that includes self- and other related processing by considering how young adults evaluate, act upon, monitor and learn about themselves. As such, this dissertation provides valuable starting points to understand how and why some individuals with an early onset of antisocial behavior manage to adapt to changing social environments and balance between situational characteristics and self-relevant goals and motivations, while others fail to do so and persist in antisocial behavior. Importantly, this dissertation also shows that considering functional neuroimaging alongside behavior reveals new insights that help to overcome difficulties (e.g. biases) associated with behavioral (self-report) measures. Finally, considering individual differences such as psychopathic traits, and specific emotional characteristics (e.g. valence of self-traits and feedback) may further illuminate functional and neural mechanisms underlying heterogenous developmental pathways. Additional research should examine changes and stability in antisocial and psychopathic tendencies throughout development and between different social contexts to further clarify whether functional neurocognitive deficits related to the development of antisocial behavior are general or context- and valence specific – and central to antisocial development in young adulthood. Ultimately, such research will provide important advances required to understand and overcome persistent antisocial behavior, and provide starting points for the development of timing-appropriate and personalized interventions.

REFERENCES

1. Baskin-Sommers, A., & Brazil, I. A. (2022). The importance of an exaggerated attention bottleneck for understanding psychopathy. *Trends in Cognitive Sciences*. <https://doi.org/10/gpc9fb>
2. Beelmann, A., & Raabe, T. (2009). The Effects of Preventing Antisocial Behavior and Crime in Childhood and Adolescence: Results and Implications of Research Reviews and Meta-Analyses. *International Journal of Developmental Science*, 3(3), 260–281. <https://doi.org/10.3233/DEV-2009-3305>
3. Bersani, B. E., & Doherty, E. E. (2018). Desistance from Offending in the Twenty-First Century. *Annual Review of Criminology*, 1(1), 311–334. <https://doi.org/10.1146/annurev-criminol-032317-092112>
4. Bertsch, K., Florange, J., & Herpertz, S. C. (2020). Understanding Brain Mechanisms of Reactive Aggression. *Current Psychiatry Reports*, 22(12), 81. <https://doi.org/10.1007/s11920-020-01208-6>
5. Bevilacqua, L., Hale, D., Barker, E. D., & Viner, R. (2018). Conduct problems trajectories and psychosocial outcomes: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, 27(10), 1239–1260. <https://doi.org/10.1007/s00787-017-1053-4>
6. Bubela, T., Nisbet, M. C., Borchelt, R., Brunger, F., Critchley, C., Einsiedel, E., Geller, G., Gupta, A., Hampel, J., Hyde-Lay, R., Jandciu, E. W., Jones, S. A., Kolopack, P., Lane, S., Lougheed, T., Nerlich, B., Ogbogu, U., O’Riordan, K., Ouellette, C., ... Caulfield, T. (2009). Science communication reconsidered. *Nature Biotechnology*, 27(6), Article 6. <https://doi.org/10.1038/nbt0609-514>
7. 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. <https://doi.org/10.1080/17470919.2012.703623>
8. Christopoulos, G. I., & King-Casas, B. (2015). With you or against you: Social orientation dependent learning signals guide actions made for others. *NeuroImage*, 104, 326–335. <https://doi.org/10.1016/j.neuroimage.2014.09.011>
9. Cohn, M. D., Veltman, D. J., Pape, L. E., van Lith, K., Vermeiren, R. R. J. M., van den Brink, W., Doreleijers, T. A. H., & Popma, A. (2015). Incentive Processing in Persistent Disruptive Behavior and Psychopathic Traits: A Functional Magnetic Resonance Imaging Study in Adolescents. *Biological Psychiatry*, 78(9), 615–624. <https://doi.org/10.1016/j.biopsych.2014.08.017>
10. D’Argembeau, A. (2013). On the Role of the Ventromedial Prefrontal Cortex in Self-Processing: The Valuation Hypothesis. *Frontiers in Human Neuroscience*, 7. <https://www.frontiersin.org/article/10.3389/fnhum.2013.00372>
11. Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences*, 1104, 70–88. <https://doi.org/10.1196/annals.1390.002>
12. Dodge, K. A., & McCourt, S. N. (2010). Translating models of antisocial behavioral development into efficacious intervention policy to prevent adolescent violence. *Developmental Psychobiology*, 52(3), 277–285. <https://doi.org/10.1002/dev.20440>
13. Drayton, L. A., Santos, L. R., & Baskin-Sommers, A. (2018). Psychopaths fail to automatically take the perspective of others. *Proceedings of the National Academy of Sciences*, 115(13), 3302–3307. <https://doi.org/10.1073/pnas.1721903115>
14. Elder, J., Davis, T., & Hughes, B. (2021). *Learning About the Self: Motives for coherence and positivity constrain learning from self-relevant feedback*. PsyArXiv. <https://doi.org/10.31234/osf.io/etz56>

15. Eme, R. (2020). Life course persistent antisocial behavior silver anniversary. *Aggression and Violent Behavior, 50*, 101344. <https://doi.org/10.1016/j.avb.2019.101344>
16. Foster, E. M. (2010). Costs and effectiveness of the fast track intervention for antisocial behavior. *The Journal of Mental Health Policy and Economics, 13*(3), 101–119.
17. Foulkes, L., McCrory, E. J., Neumann, C. S., & Viding, E. (2014). Inverted Social Reward: Associations between Psychopathic Traits and Self-Report and Experimental Measures of Social Reward. *PLOS ONE, 9*(8), e106000. <https://doi.org/10.1371/journal.pone.0106000>
18. Gatzke, L. M., & Raine, A. (2000). Treatment and prevention implications of antisocial personality disorder. *Current Psychiatry Reports, 2*(1), 51–55. <https://doi.org/10.1007/s11920-000-0042-2>
19. Gaule, A., Bevilacqua, L., Molleman, L., Roberts, R., Duijvenvoorde, A. C. van, den Bos, W. van, McCrory, E. J., & Viding, E. (n.d.). Social information use in adolescents with conduct problems and varying levels of callous-unemotional traits. *JCPP Advances, n/a*(n/a), e12067. <https://doi.org/10.1002/jcv2.12067>
20. Green, K. H., Van De Groep, I. H., Te Brinke, L. W., van der Crujisen, R., van Rossenberg, F., & El Marroun, H. (2022). A perspective on enhancing representative samples in developmental human neuroscience: Connecting science to society. *Frontiers in Integrative Neuroscience, 16*, 981657. <https://doi.org/10.3389/fnint.2022.981657>
21. Gross, J. J., & Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology, 64*(6), 970. <https://doi.org/10.1037/0022-3514.64.6.970>
22. Hamilton, R. K. B., Hiatt Racer, K., & Newman, J. P. (2015). Impaired integration in psychopathy: A unified theory of psychopathic dysfunction. *Psychological Review, 122*(4), 770–791. <https://doi.org/10.1037/a0039703>
23. Hauser, N. C., Hollerbach, P., & Habermeyer, E. (2023). Social decision-making in highly psychopathic offenders – A systematic literature review. *Aggression and Violent Behavior, 68*, 101797. <https://doi.org/10.1016/j.avb.2022.101797>
24. Hawes, D. J., Price, M. J., & Dadds, M. R. (2014). Callous-Unemotional Traits and the Treatment of Conduct Problems in Childhood and Adolescence: A Comprehensive Review. *Clinical Child and Family Psychology Review, 17*(3), 248–267. <https://doi.org/10.1007/s10567-014-0167-1>
25. Kawamoto, T., Nittono, H., & Ura, M. (2015). Trait rejection sensitivity is associated with vigilance and defensive response rather than detection of social rejection cues. *Frontiers in Psychology, 6*. <https://doi.org/10.3389/fpsyg.2015.01516>
26. Kawamoto, T., Ura, M., & Nittono, H. (2015). Intrapersonal and interpersonal processes of social exclusion. *Frontiers in Neuroscience, 9*. <https://doi.org/10.3389/fnins.2015.00062>
27. Knapp, M., McCrone, P., Fombonne, E., Beecham, J., & Wostear, G. (2002). The Maudsley long-term follow-up of child and adolescent depression: 3. Impact of comorbid conduct disorder on service use and costs in adulthood. *The British Journal of Psychiatry, 180*(1), 19–23. <https://doi.org/10.1192/bjp.180.1.19>
28. Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 21*(16), RC159. <https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001>
29. Korn, C. W., Prehn, K., Park, S. Q., Walter, H., & Heekeren, H. R. (2012). Positively Biased Processing of Self-Relevant Social Feedback. *Journal of Neuroscience, 32*(47), 16832–16844. <https://doi.org/10.1523/JNEUROSCI.3016-12.2012>

30. Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience & Biobehavioral Reviews*, *99*, 311–328. <https://doi.org/10.1016/j.neubiorev.2018.12.021>
31. Moffitt, T. E. (1993). *Adolescence-Limited and Life-Course-Persistent Antisocial Behavior: A Developmental Taxonomy*. 28.
32. Moffitt, T. E. (2018). Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, *2*, 177. <http://s2443.www.ncbi.nlm.nih.gov/pmc/articles/PMC6157602/>
33. Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology*, *14*(1), 179–207. <https://doi.org/10.1017/s0954579402001104>
34. Monahan, K. C., Steinberg, L., Cauffman, E., & Mulvey, E. P. (2013). Psychosocial (im)maturity from adolescence to early adulthood: Distinguishing between adolescence-limited and persisting antisocial behavior. *Development and Psychopathology*, *25*(4pt1), 1093–1105. <https://doi.org/10.1017/S0954579413000394>
35. Neiss, M. B., Sedikides, C., Shahinfar, A., & Kupersmidt, J. B. (2006). Self-evaluation in a naturalistic context: The case of juvenile offenders. *British Journal of Social Psychology*, *45*(3), 499–518. <https://doi.org/10.1348/014466605X64486>
36. Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(2), 241–268. <https://doi.org/10.3758/s13415-011-0083-5>
37. Odgers, C. L., Caspi, A., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Moffitt, T. E. (2007). Prediction of differential adult health burden by conduct problem subtypes in males. *Archives of General Psychiatry*, *64*(4), 476–484. <https://doi.org/10.1001/archpsyc.64.4.476>
38. Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, *20*(2), 673–716. <https://doi.org/10.1017/S0954579408000333>
39. Radke, S., Brazil, I., Schepel, I., Bulten, B., & De Bruijn, E. (2013). Unfair offers, unfair offenders? Fairness considerations in incarcerated individuals with and without psychopathy. *Frontiers in Human Neuroscience*, *7*. <https://doi.org/10.3389/fnhum.2013.00406>
40. Rodman, A. M., Powers, K. E., & Somerville, L. H. (2017). Development of self-protective biases in response to social evaluative feedback. *Proceedings of the National Academy of Sciences*, *114*(50), 13158–13163. <https://doi.org/10/gcqt6k>
41. Romeo, R., Knapp, M., & Scott, S. (2006). Economic cost of severe antisocial behaviour in children—And who pays it. *The British Journal of Psychiatry: The Journal of Mental Science*, *188*, 547–553. <https://doi.org/10.1192/bjpp.bp.104.007625>
42. Scott, S., Knapp, M., Henderson, J., & Maughan, B. (2001). Financial cost of social exclusion: Follow up study of antisocial children into adulthood. *BMJ*, *323*(7306), 191. <https://doi.org/10.1136/bmj.323.7306.191>
43. Seara-Cardoso, A., & Viding, E. (2015). Functional Neuroscience of Psychopathic Personality in Adults. *Journal of Personality*, *83*(6), 723–737. <https://doi.org/10.1111/jopy.12113>

44. Sheppes, G., & Gross, J. J. (2011). Is timing everything? Temporal considerations in emotion regulation. *Personality and Social Psychology Review*, *15*(4), 319–331. <https://doi.org/10.1177/1088868310395778>
45. Van Hazebroek, B. C. M., Blokland, A. A. J., Wermink, H. T., De Keijser, J. W., Popma, A., & Van Domburgh, L. (2019). Delinquent Development Among Early-Onset Offenders: Identifying and Characterizing Trajectories Based on Frequency Across Types of Offending. *Criminal Justice and Behavior*, *46*(11), 1542–1565. <https://doi.org/10.1177/0093854819876306>
46. Viding, E., & McCrory, E. (2019). Towards understanding atypical social affiliation in psychopathy. *The Lancet Psychiatry*, *6*(5), 437–444. [https://doi.org/10.1016/S2215-0366\(19\)30049-5](https://doi.org/10.1016/S2215-0366(19)30049-5)
47. Waller, R., & Wagner, N. (2019). The Sensitivity to Threat and Affiliative Reward (STAR) model and the development of callous-unemotional traits. *Neuroscience & Biobehavioral Reviews*, *107*, 656–671. <https://doi.org/10.1016/j.neubiorev.2019.10.005>
48. Zamani, A., Carhart-Harris, R., & Christoff, K. (2022). Prefrontal contributions to the stability and variability of thought and conscious experience. *Neuropsychopharmacology*, *47*(1), 329–348. <https://doi.org/10/gmw45f>



ADDENDUM

**NEDERLANDSTALIGE SAMENVATTING
(SUMMARY IN DUTCH)**

**DANKWOORD
(ACKNOWLEDGEMENTS IN DUTCH)**

CURRICULUM VITAE

PUBLICATIONS AND PHD ACTIVITIES

NEDERLANDSTALIGE SAMENVATTING (SUMMARY IN DUTCH)

Kinderen die al vroeg in hun leven met de politie in aanraking komen (voor hun twaalfde levensjaar) lopen het risico om later in het leven een heleboel problemen te ontwikkelen, zoals het aanhoudend vertonen van antisociaal gedrag, een slechte fysieke en mentale gezondheid, drugsproblemen en ze belanden vaker in de criminaliteit (Brazil et al., 2018; Shaw & Gross, 2008). Gelukkig weten we uit eerder onderzoek dat niet iedereen met dit verhoogde risico uiteindelijk op dit pad terecht komt: een groot deel stopt met antisociaal gedrag in de adolescentie (10-18 jaar) en de jongvolwassenheid (18-26 jaar), of laat zelfs helemaal geen antisociaal gedrag zien tijdens de ontwikkeling, en slechts een kleine groep vertoont daadwerkelijk zogenaamd 'persistent antisociaal gedrag' tijdens hun levensloop (Bersani and Doherty, 2018; Moffitt, 1993; 2018). Hoewel het dus duidelijk is *dat* deze verschillende ontwikkelingspatronen zich voordoen, is het niet heel duidelijk *waarom* en *hoe* mogelijke verschillen ontstaan, met name in de jongvolwassenheid. Een manier om hier meer duidelijkheid over te krijgen is door het bestuderen van mogelijke verschillen in gedrag, en welke neurobiologische mechanismen daaraan ten grondslag liggen. Het begrijpen van zulke gedragsmatige en neurale mechanismes kan uiteindelijk helpen om aanknopingspunten te geven voor effectieve preventie, interventie en behandeling (Gatzke & Raine, 2000) – wat erg belangrijk is omdat persistent antisociaal gedrag moeilijk te behandelen is, en niet alleen voor iemand zelf, maar ook voor slachtoffers en de maatschappij veel kosten met zich mee brengt (Foster, 2010; Knapp et al., 2002; Romeo et al., 2006; Scott et al., 2001).

In dit proefschrift heb ik mij gericht op gedragingen en hun neurale mechanismen die belangrijk zijn in de jongvolwassenheid, maar nog niet altijd goed begrepen worden in individuen die al vroeg in hun kindertijd antisociaal gedrag vertoonden. In de jongvolwassenheid vinden er veel veranderingen plaats in de sociale omgeving (Arnett 2000, Arnett et al. 2007), waarbij het belangrijk is om de sociale normen en kenmerken van de sociale situatie te begrijpen en tegelijkertijd je persoonlijke, individuele doelen niet uit het oog te verliezen. Het is daarvoor essentieel om jezelf goed te begrijpen en kennen, maar ook om de vaardigheden te hebben om te monitoren wat er om je heen gebeurt en te leren wat andere mensen van je vinden. Ook is het belangrijk om te leren hoe bepaalde gedragingen invloed hebben op jezelf en anderen binnen de sociale omgeving, en hoe je je gedrag eventueel kan aanpassen of reguleren als dat nodig is.

Uit eerder onderzoek onder typisch ontwikkelende jongvolwassenen weten we dat de structuur van het brein zich doorontwikkelt tot in de jongvolwassenheid, met name in de gebieden die belangrijk zijn voor het begrijpen van sociale informatie over jezelf en anderen, en het goed kunnen aanpassen van je gedrag in lijn met je eigen doelen en het welzijn van andere mensen (Taber-Thomas & Perez-Edgar, 2015). Een van die gebieden is de mediale prefrontale cortex (mPFC). Eerder onderzoek heeft laten zien dat de mPFC

betrokken is bij evaluaties die we over onszelf maken, en evaluaties die andere mensen over onszelf maken. Het is echter nooit in dezelfde individuen getest of het echt gaat om hetzelfde, overlappende deel van de mPFC. Dat is belangrijk om te weten, omdat dit mogelijk implicaties heeft voor de manier waarop je dit soort processen het beste kunt meten in moeilijk te bereiken en te includeren klinische groepen binnen onderzoek, zoals jongvolwassenen met een voorgeschiedenis van antisociaal gedrag. Het testen van deze hypothese was daarom het eerste doel van mijn proefschrift. Omdat ik voor het testen van deze hypothese taken heb gebruikt die al eerder gebruikt waren in typisch ontwikkelende jongvolwassenen, kon ik ook meteen kijken of ik de uitkomsten van die onderzoeken kon repliceren.

Het tweede en belangrijkste doel was om te onderzoeken hoe jongvolwassenen met en zonder voorgeschiedenis van antisociaal gedrag en variërende niveaus van psychopathische persoonlijkheidstrekken zichzelf evalueerden, *over* en *voor* zichzelf en anderen leerden, en reageren op sociale feedback informatie - en welke neurale mechanismen ten grondslag liggen aan deze sociaal-cognitieve processen en gedragingen. Daarvoor heb ik gebruik gemaakt van een combinatie van wetenschappelijke methoden, waaronder zelfrapportagevragenlijsten, diagnostische interviews, experimentele taken en het meten van functionele hersenactivatie met de MRI scanner (fMRI). Ik heb gebruik gemaakt van experimentele taken waarin onderscheid werd gemaakt tussen verschillende emotionele kenmerken, zoals valentie (positief, negatief of neutraal) (Smeijers et al., 2020) en andere verschillen in de sociale context, om beter te kunnen begrijpen of mogelijke verschillen zich enkel voordoen binnen bepaalde sociale omgevingen, of over alle sociale omgevingen heen (Nelson et al., 2008). Bovendien gebruikte ik twee complementaire fenotypische benaderingen om antisociaal gedrag in de vroege volwassenheid te karakteriseren: (1) een longitudinale, categorische benadering (door persistente, desistente en typische ontwikkelende groepen met elkaar te vergelijken, gebaseerd op psychiatrische diagnostische interviews) en een (2) cross-sectionele, dimensionale benadering (door continue niveaus van psychopathische persoonlijkheidstrekken, gemeten met een zelfrapportagevragenlijst, mee te nemen). Psychopathische persoonlijkheidstrekken vormen een belangrijke risicofactor voor het ontwikkelen van persistent gedrag (Cohn et al., 2015). Daarnaast kunnen zulke persoonlijkheidskenmerken ook gebruikt worden in onderzoek om beter te begrijpen waarom er zo veel individuele verschillen zijn tussen en binnen antisociale ontwikkelingstrajecten. Interessant genoeg heeft eerder onderzoek laten zien dat het hebben van psychopathische trekken in meerdere of mindere mate samenhangt met verschillen in gedrag en hersenactiviteit (Brazil et al., 2018). Psychopathie is een multidimensioneel construct, dat bestaat uit verschillende psychopathische dimensies (Andershed et al., 2002): kille-emotieloze trekken (gebrek aan empathie en spijt, vlakke emoties), grandioos-manipulatieve trekken (neiging tot liegen, manipulatie en narcisme) en impulsief-onverantwoordelijke trekken (neiging tot impulsief en onverantwoordelijk

gedrag). Eerder onderzoek combineert vaak deze dimensies in een totale schaal, of heeft zich vooral gericht op de kille-emotieloze trekken (Lilienfield, 2018). Het wordt echter steeds duidelijker dat de verschillende dimensies – hoewel ze sterk met elkaar samenhangen - verschillend gerelateerd kunnen zijn aan gedrag en hersenactiviteit, en dat het dus belangrijk is om ook naar de afzonderlijke dimensies te kijken (Cohn et al., 2015; Lilienfield, 2018).

Samenvatting van de belangrijkste bevindingen

Ik heb de verschillende doelen van dit proefschrift onderzocht in 4 empirische hoofdstukken. In **hoofdstuk 2** onderzocht ik of er een gemeenschappelijk neurale substraat in de mediale prefrontale cortex (mPFC) ten grondslag ligt aan zelfevaluatie vanuit het perspectief van het zelf (zelfbeeld beoordeling) en anderen (sociale evaluatie) bij 40 jongvolwassenen (18-30 jaar). De jongvolwassenen voerden twee fMRI-taken uit, de zelfbeeld taak en de Sociale Netwerk Agressie taak (SNAT). In de zelfbeeld taak (van der Cruijssen et al., 2018) evalueerden deelnemers kenmerken in twee condities. In de zelf-conditie evalueerden ze in hoeverre positieve en negatieve kenmerken binnen twee verschillende domeinen (prosociaal en fysiek uiterlijk) zichzelf beschreven op een vierpuntsschaal. In de controleconditie categoriseerden de deelnemers soortgelijke positieve en negatieve karaktertrekuitspraken in verschillende domeinen (prosociaal, fysieke verschijning, academisch of Ik weet het niet). In de SNAT (Achterberg et al., 2016) ontvingen deelnemers positieve, negatieve en neutrale sociale feedback van leeftijdsgenoten, en kregen vervolgens de kans om vergeldingsmaatregelen te nemen door een hard geluid richting de leeftijdsgenoot te sturen, waarbij een harder geluid meer agressie signaleerde.

Door eerder gebruikte taken te gebruiken kon ik ook de robuustheid en reproduceerbaarheid van eerdere gedrags- en neurobiologische bevindingen over zelfevaluatie beoordelen met behulp van de Self-Concept Task en SNAT (Achterberg et al., 2016; van der Cruijssen et al., 2018), wat belangrijk is in het licht van recente zorgen over de reproduceerbaarheid van fMRI-resultaten (Elliot et al., 2020). Ik was in staat om de eerder gevonden gedragspatronen te repliceren, waaruit bleek dat jongvolwassenen tijdens interne zelfevaluatie vaker vinden dat positieve kenmerken (vs. negatieve) goed bij zichzelf passen, en prosociale eigenschappen meer van toepassing zijn dan fysieke eigenschappen (van der Cruijssen et al., 2018). In de SNAT vond ik dat individuen meer agressie laten zien na negatieve feedback, minder na neutrale feedback en het minst na positieve feedback (Achterberg et al., 2016, 2017, 2018, 2020; Dobbelaar et al., 2021, 2022). Met behulp van Region-of-Interest (ROI) analyses onderzocht ik ook of ik neurale activiteitspatronen kon repliceren, wat voor het merendeel van de bevindingen het geval was. Zo vertoonden jongvolwassenen tijdens sociale feedbackverwerking de verwachte hogere activiteit in de anterieure insula en anterieure cingulate cortex (ACC) voor positieve en negatieve feedback ten opzichte van neutrale feedback (Achterberg et al., 2016), wat

past bij het idee dat deze hersengebieden saillante sociale informatie verwerken. Ik kon een eerdere bevinding dat tijdens zelfevaluaties sprake zou zijn van meer mPFC activiteit tijdens de zelfconditie in vergelijking met de controleconditie echter niet repliceren (van der Crujisen et al., 2018) - hoewel ik wel verschillen vond bij het vergelijken van positieve kenmerken met controlekenmerken, wat past bij het idee dat beter passende, positieve kenmerken tot meer mPFC activiteit kunnen leiden in vergelijking met minder toepasselijke negatieve traits (D'Argembeau, 2013).

Een derde, exploratief doel van de in hoofdstuk 2 beschreven studie was om te testen of de verschillende soorten sociale feedback (positief, negatief en neutraal) gerelateerd waren aan verschillende patronen van neurale activatie tijdens vergeldingsreacties. Ik vond dat het ontvangen van positieve feedback resulteerde in meer activiteit in de dorsolaterale PFC (dlPFC), vergeleken met negatieve en neutrale feedback, wat erop wijst dat dit hersengebied belangrijk zou kunnen zijn voor de regulatie van vergeldingsreacties, een bevinding die nu is bevestigd door latere studies met hetzelfde SNAT-paradigma (Dobbelaar et al., 2022, zie ook hoofdstuk 3).

Op groepsniveau vond ik overlap in de anterieure/rostrale mPFC voor positieve evaluaties vanuit het perspectief van zelf en anderen. Echter, op individueel niveau, wanneer ik keek naar correlaties tussen taken, vond ik geen bewijs voor overlap in de mPFC. In hoofdstuk 2 beargumenteerde ik dat deze discrepantie past bij ander onderzoek waaruit blijkt dat het onderzoeken van functionele hersenactivatie op betrouwbare wijze mechanistische inzichten kan onthullen op groepsniveau, maar niet op individueel niveau in de mate waarin deelnemers deze regio's betrekken bij verschillende taken, zoals bleek uit een relatief lage test-hertest betrouwbaarheid op individueel niveau (Elliott et al., 2020).

Vervolgens richtte ik me in **hoofdstuk 3** op hoe jongvolwassenen met en zonder een voorgeschiedenis van antisociaal gedrag - en met verschillende niveaus van psychopathische trekken - zichzelf evalueerden, en of ze vergelijkbare of verschillende patronen van hersenactivatie lieten zien tijdens zelfevaluatie over domeinen en contexten heen. Hiervoor voerde een subset van jongvolwassenen met een vroeg-beginnende persistente of desistente ontwikkeling van antisociaal gedrag ($n = 54$; leeftijd 18-30) uit de RESIST Cohort studie en de steekproef van typisch ontwikkelende jongvolwassenen ($n = 40$, zie ook hoofdstuk 2) de zelfbeeld taak beschreven in hoofdstuk 2 uit. Ik vond dit keer wel verhoogde activiteit in mPFC tijdens zelfevaluatie, consistent met eerdere studies (Denny et al., 2012; van der Crujisen et al., 2018) en liet zien dat dezelfde hersengebieden ten grondslag liggen aan zelfevaluaties over groepen met verschillende geschiedenissen van antisociaal gedrag heen. In de totale steekproef waren psychopathische trekken (gecombineerd totaal van kille-emotieloze trekken, grandioos-manipulatieve trekken en impulsief-onverantwoordelijke trekken) geassocieerd met meer negatieve en minder positieve zelfbeoordelingen in het prosociale domein, en niet in het fysieke verschijningsdomein. In termen van neurale activiteit waren kille-emotieloze geassocieerd

met minder anterieure mPFC activiteit tijdens algemene zelfevaluaties, wat kan wijzen op verschillen in hoe individuen met hoge niveaus van deze trekken abstracte informatie verwerken wanneer ze over zichzelf denken, mogelijk om meer stabiliteit in hun gedachten te behouden op basis van bestaande overtuigingen over zichzelf (Zamani et al., 2022). Al met al suggereren deze bevindingen dat het overkoepelende concept van psychopathie geassocieerd is met domeinspecifieke zelfbeoordelingen, terwijl specifieke subdimensies (bijv. kille-emotieloze trekken) samenhangen met afwijkend breinactivatiepatroon over domeinen heen - wat benadrukt dat het in aanmerking nemen van zowel de totale niveaus van psychopathische trekken als specifieke sub-dimensies in toekomstig onderzoek meer inzichten kan opleveren in de etiologie en complexe ontwikkelingspaden van antisociaal gedrag.

In **hoofdstuk 4** onderzocht ik hoe jongvolwassenen over zichzelf leerden en reageerden op sociale feedback van anderen. Meer specifiek bestudeerde ik neurale en gedragsmatige reacties op sociale feedback in dezelfde groepen jongvolwassenen met persistent antisociaal gedrag uit de kindertijd, desistent antisociaal gedrag uit de kindertijd en geen geschiedenis van antisociaal gedrag, en verschillende niveaus van psychopathische trekken zoals beschreven in hoofdstuk 2-3, met behulp van de SNAT zoals beschreven in hoofdstuk 2. Niet alleen persisters, maar beide groepen met een voorgeschiedenis van antisociaal gedrag (persisters en desisters) vertoonden verhoogde activiteit van de anterieure insula tijdens feedbackverwerking, ongeacht het type feedback (negatief, neutraal, positief), vergeleken met de typisch ontwikkelende controlegroep. Mogelijk weerspiegelt deze bevinding neurale overgevoeligheid voor saillante en zelf-relevante sociale feedbackinformatie, en meer in het algemeen, moeite in het onderscheiden van sociale feedbacksignalen (Kawamoto et al., 2015a, 2015b). Een andere bevinding was dat verhoogde activiteit in de dlPFC tijdens alle soorten feedbackverwerking specifiek was voor de desisterende groep. Deze verhoogde dlPFC activiteit weerspiegelt waarschijnlijk aandachtsprocessen in reactie context-afhankelijke veranderingen in feedback presentatie tussen trials (Bertsch et al., 2020; Niendam et al., 2012), die cognitieve en emotionele regulatie van daaropvolgend gedrag ondersteunen.

Tijdens de vergeldingsreacties vond ik niet dat jongvolwassenen met een persistent ontwikkelingstraject van antisociaal gedrag meer agressie vertoonden na negatieve feedback dan de andere twee groepen. Het was echter zo dat jongvolwassenen met een persistente antisociale ontwikkeling niet differentieerden in hun gedragsreacties, maar gelijksoortige agressieve reacties vertoonden voor alle soorten feedback, in tegenstelling tot de controlegroep en degenen met een desistent antisociaal traject. Bovendien vertoonden jongvolwassenen met een persistent antisociaal traject na het ontvangen van positieve feedback minder dlPFC-activiteit tijdens hun gedragsrespons, vergeleken met de andere twee groepen. Uit de bevindingen bleek verder dat individuen met een desistent antisociaal traject specifieke gedrags- en neurale mechanismen vertoonden die kunnen verklaren waarom zij erin slagen met succes af te zien van antisociaal gedrag - en

adaptief gedrag vertonen in veranderende sociale contexten. Zo vond ik een positieve associatie tussen agressieregulatie na positieve feedback en activiteit in de ACC en het dorsale striatum (caudatus en putamen) tijdens de gedragsrespons, die het sterkst was in het desistente antisociale ontwikkelingstraject. Wanneer ik individuele verschillen in psychopathische trekken bekeek, vond ik dat hogere niveaus van psychopathische karaktereigenschappen geassocieerd waren met meer agressie, maar ik vond geen bewijs dat psychopathie of de verschillende subdimensies geassocieerd waren met neurale activiteit tijdens feedbackverwerking of agressie.

Al met al suggereren deze bevindingen dat een aanhoudende antisociale ontwikkeling geassocieerd zou kunnen zijn met moeilijkheden bij het onderscheiden van sociale feedback, zowel tijdens het verwerken van deze feedback als tijdens de daaropvolgende gedragsreacties. Personen met een desistent ontwikkelingspatroon vertonen daarentegen eveneens een overgevoeligheid voor sociale feedbacksignalen, maar differentiëren wel in hun daaropvolgende vergeldingsreacties, zowel op gedrags- als op neuraal niveau.

In **hoofdstuk 5**, tenslotte, onderzocht ik of hoge niveaus van psychopathische trekken geassocieerd waren met verschillen in het leren voor jezelf en anderen in verschillende positieve (beloning-gerelateerde) en negatieve (verlies-gerelateerde) sociale contexten, en overeenkomstige veranderingen in neurale activiteit in het striatum, in dezelfde risicogroep van jongvolwassenen met een voorgeschiedenis van antisociaal gedrag ($n = 53$) als beschreven in hoofdstuk 3 en 4. Voor deze studie paste ik een sociale probabilistische fMRI leertaak aan (Christopoulos & King-Casas, 2015), waarin deelnemers moesten leren in verschillende contexten, gekenmerkt door verschillende uitkomstcontingenties, resulterend in beloningen of verliezen voor henzelf en een onbekende ander tegelijkertijd. In lijn met eerdere studies die dit soort sociale leertaken gebruiken vond ik dat jongvolwassenen sneller leerden wanneer het leren gevolgen had voor henzelf in vergelijking met anderen, en dat ze beter leerden in een beloningscontext dan in een verliescontext (Westhoff et al., 2021; Lockwood et al., 2016). Bovendien is deze studie een van de eerste studies die aantoonde dat hogere niveaus van psychopathische trekken geassocieerd waren met meer neurale differentiatie tussen zelf en ander in het striatum, wat erop zou kunnen wijzen dat individuen die hoog scoren op psychopathische persoonlijkheidskenmerken meer onderscheid maken in beloningswaardes vergeleken met individuen die lager scoren op psychopathische trekken (Overgaauw et al., 2020; Pujara et al., 2013; Hosking et al., 2017; Buckholz et al., 2010; Bjork et al., 2012; Lockwood et al., 2017). Samen passen deze bevindingen goed bij het idee dat leren afhangt van sociale contexten en individuele verschillen in psychopathische trekken, hoewel het belangrijk is om deze resultaten te repliceren.

Discussie en aanbevelingen voor vervolgonderzoek en de klinische praktijk

Samenvattend laten de studies in mijn proefschrift zien dat persistent antisociaal gedrag, en aanwezigheid van hoge niveaus van psychopathische persoonlijkheidstrekken in de jongvolwassenheid samenhangen met gedragsverschillen in zelf-relevante en feedback-gerelateerde informatieverwerking en agressieregulatie. De cognitieve en gedragsverschillen die ik waarnam waren geassocieerd met functionele veranderingen in verschillende frontolimbische hersengebieden, waaronder de (v)mPFC, dlPFC, ACC, Insula, en Striatum. De neurocognitieve problemen bij persistente antisociale personen lijken te worden gekenmerkt door een beperkt vermogen om onderscheid te maken tussen verschillende sociale informatiesignalen en om hun gedrag aan te passen aan specifieke en veranderende sociale contexten. Bevindingen uit het huidige proefschrift suggereren verder dat een deel van het succes van individuen met een desistent patroon van antisociaal gedrag om af te zien van agressie verklaard kan worden door hun vermogen om agressie te reguleren en sociale informatie te monitoren, ook al ervaren zij vergelijkbare neurale overgevoeligheid voor saillante, zelf-relevante sociale feedbackinformatie. Hoewel het zou kunnen dat de geobserveerde verschillen voortkomen uit een bepaald onvermogen om gedrag aan te passen bij individuen met een persistente antisociale ontwikkeling en hoge niveaus van psychopathische trekken, is het ook mogelijk dat ze simpelweg niet de motivatie hebben om hun gedrag aan te passen, een mogelijkheid die verder onderzocht moet worden in toekomstig onderzoek (Foulkes et al., 2014; Viding & McCrory, 2019; Drayton et al., 2018).

Op basis van de bevindingen van dit proefschrift doe ik een aantal eerste aanbevelingen voor de klinische praktijk. Zo zouden preventie en behandeling zich bijvoorbeeld kunnen richten op een combinatie van gedragsregulatiestrategieën en vaardigheden zoals het anders interpreteren van - en het vergroten van de gevoeligheid voor - context-specifieke sociale informatie, en op het leren en ondersteunen van gedragsregulatie (Bertsch et al, 2020), waarbij rekening moet worden gehouden met de motivatie van deelnemers (Foulkes et al., 2014; Viding & McCrory, 2019; Drayton et al., 2018; Gaule et al., 2021; Hawes, Price & Dadds, 2014). Voor goede implementatie van zulke fundamentele kennis is het belangrijk om goed te communiceren en samen te werken met verschillende belanghebbenden, zoals klinici, jongeren zelf, en beleidsmakers.

Uit het huidige proefschrift komt verder naar voren dat naast het onderzoeken van neurobiologische factoren, het onderzoeken van kenmerken van de sociale omgeving (zoals valentie) en individuele verschillen (zoals psychopathische trekken) eveneens belangrijk is om de ontwikkeling en heterogeniteit in uitingen van antisociaal gedrag goed te kunnen begrijpen. In toekomstig onderzoek is het belangrijk om van deze factoren ook de longitudinale ontwikkeling te bestuderen, idealiter in designs met grote samples waarin recht kan worden gedaan aan de complexe wisselwerking tussen sociale, individuele en neurobiologische kenmerken. Naast een longitudinale benadering kan er

juist ook worden gekozen voor het bestuderen van snelle, onmiddellijke korte termijn aanpassingen binnen experimentele taken en tussen trials (Flechtsenhar et al., 2022). Longitudinale benaderingen kunnen dus worden aangevuld met studies die specifiek gericht zijn op het identificeren of verduidelijken van belangrijke functionele, computationele en gedragsmechanismen – of specifieke situationele en persoonlijke kenmerken. Uiteindelijk is een combinatie van deze lange- en korte termijn ontwikkelingsbenaderingen, die uitgaat van hun gecombineerde sterke punten, het meest effectief om de ontwikkeling van antisociaal gedrag te begrijpen (Brazil et al., 2018).

CURRICULUM VITAE

Ilse van de Groep was born on April 28th, 1993 in Utrecht. After completing her secondary education in Harderwijk, at the Christelijk College Nassau Veluwe in 2011, she started studying psychology at Utrecht University. Upon completion of her bachelor's degree (2014; Cum Laude), which included the Social Sciences Honours programme von Humboldt College, she pursued the research master Social and Health Psychology (2014-2016). During her master's degree, Ilse participated in the graduate-level Young Innovators League Honours programme, which focused on designing and delivering innovative and sustainable solutions to real life societal challenges (2016). During and after her studies, Ilse worked as a research and teaching assistant in social and clinical psychology at Utrecht University. In 2018, Ilse started her PhD on the behavioral and neural development of antisocial behavior under supervision of Eveline Crone, Marieke Bos, Lucrez Nauta-Jansen and Arne Popma. In addition to her research, she worked as Outreach & Communication specialist at the Convergence Initiative Healthy Start, and made several graphic designs for NWO/NWA, NeurolabNL and SYNC lab. Ilse was also involved in the KNAW Honours programme Arts/Science for promising young artists and scientists, and participated in the NeurolabNL consortium during her PhD. In 2022, Ilse was awarded a PhD Excellence Award for Best Poster by the Erasmus Graduate School of Social Sciences and the Humanities, and in 2023 she received the Koningsheide Award for Best Article in Forensic Psychiatry 2020-2022. Ilse will continue her work as a postdoctoral researcher at the Erasmus SYNC lab and Leiden University.



PUBLICATIONS AND PHD ACTIVITIES

Achterberg, M., Becht, A., van der Crujisen, R., **van de Groep, I. H.**, Spaans, J. P., Klapwijk, E., & Crone, E. A. (2022). Longitudinal associations between social media use, mental well-being and structural brain development across adolescence. *Developmental Cognitive Neuroscience, 54*, 101088. doi: 10.1016/j.dcn.2022.101088

Crone, E.A., Green, K.G., **van de Groep, I.H.**, van der Crujisen, R. (2022). A Neurocognitive Model of Self Concept Development in Adolescence. *Annual Review of Developmental Psychology, 4*, 273-295. doi: 10.1146/annurev-devpsych-120920-023842

Green, K. H., **van de Groep, I. H.**, te Brinke, L. W., van der Crujisen, R., van Rossenberg, F., & ElMarroun, H. (2022). A perspective on enhancing representative samples in developmental human neuroscience: Connecting science to society. *Frontiers in integrative neuroscience, 16*, 981657. doi: 10.3389/fnint.2022.981657

van de Groep, I. H., Bos, M. G., Jansen, L. M.C., Kocevskaja, D., Bexkens, A., Cohn, M., ... & Crone, E. A. (2022). Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback. *Neuroimage: clinical, 34*, 102973. doi: 10.1016/j.nicl.2022.102973

van de Groep, I. H., Bos, M. G., Jansen, L. M.C., Achterberg, M., Popma, A., & Crone, E. A. (2021). Overlapping and distinct neural correlates of self-evaluations and self-regulation from the perspective of self and others. *Neuropsychologia, 161*, 108000. doi: 10.1016/j.neuropsychologia.2021.108000

Westhoff, B., Koele, I. J., & **van de Groep, I. H.** (2020). Social learning and the brain: How do we learn from and about other people? *Frontiers for Young Minds, 8(95)*, 1-9. doi: 10.3389/frym.2020.00095

van den Hout, M. A., van Dis, E. A., van Woudenberg, C., & **van de Groep, I. H.** (2019). OCD-like checking in the lab: A meta-analysis and improvement of an experimental paradigm. *Journal of Obsessive Compulsive and Related Disorders, 20*, 39-49. doi: 10.1016/j.jocrd.2017.11.006

van de Groep, I. H., de Haas, L. M., Schutte, I., & Bijleveld, E. (2017). Spontaneous eye blink rate (EBR) predicts poor performance in high-stakes situations. *International Journal of Psychophysiology, 119*, 50-57. doi: 10.1016/j.ijpsycho.2017.01.009

IN REVISION

van de Groep, I. H., Bos, M. G., Jansen, L. M.C., Popma, A., & Crone, E. A. Through the looking glass: the neural basis of self-concept in young adults with antisocial trajectories. doi: 10.31234/osf.io/b7f6k

PREPRINTS

van de Groep, I. H., Bos, M. G., Popma, A., Crone, E. A. & Jansen, L. M.C. (2022). A neurocognitive model of early onset persistent and desistant antisocial behavior in early adulthood. doi: 10.31234/osf.io/sw73u

ACADEMIC PRESENTATIONS, INVITED TALKS AND POSTER PRESENTATIONS

van de Groep, I. H., Bos, M.G., Jansen, L.M.C., Popma, A., & Crone, E.A. (2022). Understanding Early Adulthood Antisocial Behavior. Invited virtual talk at the Clinical, Educational, and Health Psychology Research Department, University College London.

van de Groep, I. H., Bos, M. G., Jansen, L. M.C., Popma, A., & Crone, E. A. (2022). Through the looking glass: the neural basis of self-concept in young adults with antisocial trajectories. Flux 2022 conference, Paris, France.

van de Groep, I.H. & Green, K.H. (2022). Mentale gezondheid van jongeren tijdens en na de coronacrisis. Jeugd in onderzoek conference. 'S-Hertogenbosch, the Netherlands.

van de Groep, I.H., Bos, M.G.N., Jansen, L.M.C., Popma, A., & Crone, E.A (2022). Social brain responses in antisocial behavior development. EFCAP 2022 virtual conference.

van de Groep, I.H., Bos, M.G.N., Jansen, L.M.C., Popma, A., & Crone, E.A (2022).). Through the looking glass: the neural basis of self-concept in young adults with antisocial trajectories. NVP 2022, Egmond aan Zee, the Netherlands.

van de Groep, I. H., Bos, M. G., Jansen, L. M., Kocevskaja, D., Bexkens, A., Cohn, M., ... & Crone, E. A. (2021). Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback. VNOP 2021 virtual conference.

van de Groep, I. H., Bos, M. G., Jansen, L. M., Kocevskaja, D., Bexkens, A., Cohn, M., ... & Crone, E. A. (2021). Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback. Flux 2021 virtual conference.

van de Groep, I.H., Bos, M.G.N., Jansen, L.M.C., Popma, A., & Crone, E.A (2019). Neurobiological correlates of social evaluation and self-concept: A Social information processing approach to understanding persistent antisocial behavior. Amsterdam Neuroscience, in Amsterdam, the Netherlands.

INVITED TALKS FOR PROFESSIONALS, STUDENTS AND THE GENERAL PUBLIC

- 2022 Round Table Citizen Science at the Nationale Wetenschapscommunicatiedag, together with dr. Suzanne van de Groep and dr. Karlijn Hermans.
- 2022 Round Table Juvenile Delinquency at the Rotterdammer of the Future event (Generation R Rotterdam), together with Cynthia Cristiani, policy maker at the Rotterdam Municipality.
- 2021 Provided an online lecture on “Recent developments in the study of antisocial behavior: Developmental trajectories, brain development and COVID-19” for the “Corpus Delicti” Leiden Criminology.
- 2021 Provided an online lecture on the impact of the covid crisis on the wellbeing of adolescents to ‘s Heeren Loo, together with dr. Suzanne van de Groep.
- 2020 Provided an online lecture on the impact of the covid crisis on the wellbeing of adolescents to BV Young (youth worker association), together with dr. Lysanne te Brinke.
- 2019 Provided a lecture and workshop on the development of the adolescent brain antisocial behavior for the Antwerp Police, together with dr. Eduard Klapwijk.
- 2019 Provided an interactive lecture and workshop ‘Rebel’ about risk-taking and antisocial behavior on the (Y)our Rights Festival, together with dr. Neeltje Blankenstein.

WORKSHOPS, TUTORIALS, EVENTS AND EXPLANATORY VIDEOS (SELECTION)

- 2022 Interactive workshop “Visualizing your research” at the Child & Adolescent psychiatry department – Amsterdam UMC and at Erasmus University – Graduate Research Day.
- 2022 RESIST Project Erasmus SYNC lab open science videos.
- 2021 Data Visualization skills exchange at the Flux 2021 virtual conference, together with Lina van Drunen.

2021 Social SYNC Lab, Expeditie Next Franeker (~6000) children).

2019 'Kijk in je brein' experience at Leiden Museum night.

AWARDS & HONOURS

2023 Koningsheide Award for Best Article in Forensic Psychiatry / Psychology 2020-2022 (€2000) for "Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback".

2021 EGHS PhD Excellence Award for Best Poster 2021 "Resisting aggression in social contexts: The influence of life-course persistent antisocial behavior on behavioral and neural responses to social feedback", presented at Flux 2021 virtual conference.

2019 Arts/science: Academy Honours Programme for Young Artists and Scientists.

GRAPHIC DESIGN & SCIENCE VISUALIZATIONS

- Erasmus SYNC-lab (2022). E-magazine: Bouncing back after the Corona Crisis.
- Healthy Start (2022). Healthy Start Strategic Plan.
- NeurolabNL (2022). Education and Safety for Youth: A co-creation between Science and Society.
- Manifesto NWA youth challenge (2021). Advice for policymakers, scientists and administrators about adolescents and corona.
- Erasmus SYNC-lab (2021). E-magazine: The adolescent brain in the corona crisis: From vulnerabilities to opportunities.

COMMUNICATION STRATEGIES, WEBSITES & SOCIAL MEDIA

2022 Social media (Twitter; Linked-In), website and newsletter management & development Healthy Start
Development and implementation of the Healthy start communication strategy

2020-2022 Social media (Twitter; Linked-In) and website management & development Erasmus SYNC lab

MEDIA APPEARANCES

- Wegwijzer jeugd en veiligheid (2022). Interview over promotieonderzoek: 'Criminele jongeren verwerken feedback anders'
- AD (2022). Interview over promotieonderzoek: 'Zo ontwikkelt je brein als je als kind iets gejat of in brand gestoken hebt'

- Druk (2022). Interview over zelfbeeld: "Wow, wat ben je mooi/slim/lelijk/dik/dom!" in het boek "Druk: Alles wat je moet weten als je hoofd overloopt"
- Erasmus Magazine (2022). Interview over onderzoek en inspiratiebronnen: 'Een Familiestamboom leert me over het alledaagse leven van vroeger en nu'
- Radio 1 (2020). Radio-interview 'De invloed van de coronacrisis op jeugdcriminaliteit'
- Erasmus Magazine (2020). Interview over inzet voor diversiteit en inclusie: 'Hoe een masterprogramma en onderzoeksinstituten racisme en witte superioriteit tegengaan'

POPULAR SCIENCE WRITING & BLOGS (SELECTION)

- **Ilse H. van de Groep** (2022). "Educating the researcher of the future: What can we learn from large transdisciplinary research projects?" (Erasmus SYNC Blog)
- **Ilse H. van de Groep** & Merel Spaander (2020). "Antisociaal gedrag in crisistijd: waarom jongeren de kroon spannen" (NeurolabNL Blog)
- **Ilse H. van de Groep** (2019). "Why are we so obsessed by true crime stories?" (Leiden Psychology Blog)
- **Ilse H. van de Groep** (2018). "Does social media use prevent us from remembering the moments we try to preserve?" (Leiden Psychology Blog)

TEACHING

2018-2022 Bachelor and Master thesis and internship supervisor (15 students across various disciplines: psychology, criminology, neuroscience, medicine), research master thesis supervisor (2 students), co-supervision of 5 research assistants, workgroup teacher Bachelor courses School Psychology and Clinical Child and Adolescent psychology, replacing coordinator School Psychology (covering during pregnancy leave).

COURSES

Leiden University & LIBC

- Introduction to teaching for PhD Candidates
- Effective communication: How to Build Effective Work relationships
- Data- and project management for PhD's
- Negotiations for PhD's
- Networking Skills for Young Scientists
- Strengthening your Academic Resilience
- Scientific Conduct
- fMRI Data Analysis and Statistics
- LIBC Scan license
- Practical Linux

REVIEWING

Ad hoc reviewer for Nature Mental Health, JAMA Pediatrics, Psychological Medicine, Social Cognitive and Affective Neuroscience, Human Brain Mapping, Developmental Cognitive Neuroscience, Journal of Early Adolescence, Journal of Obsessive-Compulsive and Related Disorders.

