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

Ilse H. Van de Groep a,b,d,*, Marieke G.N. Bos b,c, Lucres M.C. Jansen d, Michelle Achterberg a,b,c, Arne Popma d, Eveline A. Crone a,b

a Erasmus School of Social and Behavioral Sciences, Erasmus University Rotterdam, the Netherlands
b Leiden Institute for Brain and Cognition, Leiden University, the Netherlands
c Department of Developmental and Educational Psychology, Institute of Psychology, Leiden University, the Netherlands
d Department of Child and Adolescent Psychiatry, Amsterdam University Medical Center, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Self-evaluation
Peer feedback
Medial prefrontal cortex
Dorsolateral prefrontal cortex
Aggression

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.

1. 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 et al., 2017). The medial prefrontal cortex (PFC) has been associated with both evaluations of one’s internal self-concept (i.e., internal self-evaluation, Denny et al., 2012) as well as evaluations of self by others (external self-evaluation) (Flanagan and Beer, 2013; Somerville et al., 2006), suggesting that it may have an intertwined function in monitoring self in relation to others (Crone and 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 et al., 2018). While many studies have shown evidence for the association between rejection and aggression and the neural correlates that precede such retaliatory behavior (Bertich, Florange & Herpertz, 2020), less is known about the
brain areas involved in aggressive behavior following different types of social feedback (Riva et al., 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 and Beer, 2013; van der Crujsen et al., 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 et al., 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 and Peake, 2012; Veroude et al., 2014), there is evidence that anterior medial PFC activation is stronger for traits that are evaluated positively (van der Crujsen 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, 2015) 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 et al., 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 et al., 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 and Van der Molen, 2010). Subsequent studies revealed that medial PFC may be particularly important to protecting self-views (Hughes and 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 et al., 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 and 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 et al., 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 et al., 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 and 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 and 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 Lijzen, Bakermans-Kranenburg and Crone, 2020). In addition, these latter studies have shown that the dorsolateral PFC (dIPFC) is involved in individual differences in aggression regulation. In particular, those participants who showed stronger activity during evaluation in the dIPFC, 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 dIPFC, 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 Crujsen 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 dIPFC) 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 dIPFC 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 a brief version of the self-concept fMRI task (van der Cruijsen, Peters, van der Aar and Crone, 2018) (see Fig. 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 Fig. 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 4400 ms.

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 Fig. 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 longer noise blast duration (i.e., a loud noise). Noise blast duration was visualized by a volume bar (see Fig. 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 et al., 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 et al. (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 (500 ms), followed by social feedback (2500 ms), another fixation screen (jittered between 3000 and 5000 ms), and the volume bar screen (5000 ms). Upon response, the volume bar fills up, with a new colored block appearing every 350 ms, until the button is released or the maximum volume is reached (at 3500 ms). Subsequently, the volume bar screen remained visible in that state for the remainder of the 5000 ms. Before subsequent new trials, another fixation cross was presented (jittered between 0 and 11550 ms). 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 & Vollm, 2011) 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 s, 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 vol, 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 s (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 (Welcome Department of Cognitive Neurology) to allow for a direct comparison with previously published articles (Achterberg et al., 2016; van der Cruijsen 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 et al., 1997). Functional images were visually checked before preprocessing, and following each pre-processing step to ensure quality control.

2.5. 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 modeled as a series of zero duration events convolved with the hemodynamic response function (HRF). First, we modeled the onset of social feedback with a zero duration and with separate regressors for the positive, negative and neutral peer feedback and Second, we modeled the start of the noise burst for the length of the noise burst duration (i.e. length of button press) and with separate regressors for noise burst 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.

2.6. 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 Cruijsen 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 burst 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.

2.7. Neuroimaging methods: region-of-interest analyses

In order to examine whether neural activity findings for self-evaluations (van der Cruijsen et al., 2017, 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 et al., 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 overlap medial PFC region from the both task analyses, we correlated these values with each other, and with the behavioral ratings of both tasks.

2.8. 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 and 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) × 2 (valence: positive vs. negative) repeated-measures ANOVA. As can be seen in Fig. 2A and Supplementary Table S2, there was a main effect of valence, F(1, 38) = 412.94, p < .001, ηp² = 0.92, 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, ηp² = 0.30, 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 × valence interaction effect, F(1,38) = 3.28, p = .06.

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 Fig. 2B and Supplementary Table S3, This analysis showed that participants significantly differentiated between valences, F(2, 76) = 41.15, p < .001, ηp² = 0.52 (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 < 0.001).

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.

![Fig. 2.](image-url) (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.)
reversed contrast, Self-Negative contrast showed significant activation in several regions including the medial PFC (see Fig. 3A and Table 2). The reversed contrast, Negative Self > Control 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 Fig. 3A and Table 2). The reversed contrast, Negative Self > Control 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.

### 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 Fig. 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.

#### 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

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| 1. Positive self-evaluation | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 | -25 |
| 2. Negative self-evaluation | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 | -39 |
| 3. Noise blast duration | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 | -0.05 |
| 4. Noise blast feedback | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 |
| 5. Noise blast feedback | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| 6. Mean Noise blast duration | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 |
| 7. Noise blast duration | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |
| 8. Self-evaluation | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 | -0.40 |
| 9. ROI- | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 | -0.47 |
| 10. mPFC ROI | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 | -0.44 |
| 11. mPFC ROI | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 | -0.41 |
| 12. mPFC ROI | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 | -0.37 |
| 13. mPFC ROI | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 | -0.34 |
| 14. mPFC ROI | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 | -0.31 |
| 15. mPFC ROI | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 | -0.29 |

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, 2013). * indicates p < .05. ** indicates p < .01.

### 3.3. Meta-analysis results

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, ηp² = 0.20. 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, ηp² = 0.35 (see Fig. 3B and supplementary table S4-5).

ROI- To test whether neural activity for self-evaluations (van der Cruijsen et al., 2017, 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, ηp² = 0.20. 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, ηp² = 0.35 (see Fig. 3B and supplementary table S4-5).
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 Fig. 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 Fig. 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 394921–394924.

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 Fig. 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, $\eta^2_p = .118$, right insula F(1.76, 65.06) = 4.38, p = .024, $\eta^2_p = .097$ (all GG-corrected).

As can be seen in Fig. 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

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 Fig. 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 Fig. 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 394921–394924.

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 Fig. 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, $\eta^2_p = .118$, right insula F(1.76, 65.06) = 4.38, p = .024, $\eta^2_p = .097$ (all GG-corrected).

As can be seen in Fig. 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

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.

Table 2

| Area of activation | MNI Coordinates | Test statistic | Cluster size |
|-------------------|----------------|---------------|-------------|
|                   | x   | y   | z   | T    |   |
| Control > Positive Self | Left Lingual | -9  | 82  | -14 | 6.14 | 143 |
|                   | Left Frontal Inferior | -48  | 35  | 22  | 6.10 | 118 |
|                   | Triangularis | 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/FOUSRROB for a full, unthresholded overview of activation.

Table 3

| Area of activation | MNI Coordinates | Test statistic | Cluster size |
|-------------------|----------------|---------------|-------------|
|                   | x   | y   | z   | T    |   |
| Positive > Negative feedback | Left Calcarine | 3  | -82  | -8  | 6.32 | 571 |
| Positive > Neutral feedback | Right Fusiform | 24  | -73  | -11 | 4.86 | 873 |
| Positive > Neutral feedback | 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 |
| Negative > Neutral feedback | Right Inferior Temporal | 51  | -64  | 5  | 5.04 | 798 |
| Positive > Neutral feedback | Right Middle Occipital | -45  | 82  | 7  | 6.04 | 418 |
| Negative > Neutral feedback | 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/FOUSRROB for a full, unthresholded overview of activation.
Post-hoc tests yielded a significantly higher activation following negative feedback compared to activation following neutral feedback in the left insula, \( p = .004 \). The other differences between conditions in the ACC and left/right insula were not significant, all other \( p ' s > 0.076 \). The repeated measures ANOVA for the medial PFC showed no significant differences between conditions, \( F(2, 68) = 1.56, p = .22 \) (see Fig. 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 (Fig. 5). The parameter estimate difference scores for negative > positive feedback and negative > neutral feedback were correlated with the noise blast differences scores. The correlates for the negative > positive feedback resulted in a significant negative correlation, \( r = -0.45, p = .004 \), 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 (Fig. 5). The correlation for the negative > neutral difference scores was not significant.

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 Fig. 6 for a visual presentation of the mPFC overlap).

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 Fig. 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 S10295-S10301.

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.

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 Cruijsen et al., 2017, 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, 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 Cruijsen 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 Cruijsen et al., 2018). However, we did observe this activity at the ROI level for positive self-trait relative to the control condition (Denny et al., 2012; van der Cruijsen 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 Cruijsen et al., 2017). A similar anterior medial PFC region is often implicated for self-processing specifically (Koski et al., 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

---

### 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 | Cluster size |
|--------------------|-----------------|---------------|-------------|
| 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 | $48$ $32$ $22$ | $4.58$ | $269$ |
| Triangularis |                  |               |             |
| Right Supramarginal Gyrus | $51$ $-46$ $52$ | $4.04$ | $154$ |
| Positive > Neutral feedback |                |               |             |
| Left Frontal Inferior | $-48$ $14$ $31$ | $5.52$ | $612$ |

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/FOUSBR0B for a full, unthresholded overview of activation.

---

### Fig. 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.

### Fig. 7

Whole brain effects of the contrasts SNAT Positive > Neutral and Positive > Neutral Self and their overlap during the noise blast response.
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 (Fig. 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 Crijnsen et al., 2019). Moreover, we confirmed the well-established association between proactive and reactive aggression (Card and 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 et al., 2010; Euler et al., 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 dPFC 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, 2021; Hoppenbrouwers et al., 2013). However, due to the explorative nature of these analyses, replication is warranted to confirm our conclusions.

4.1. 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 Crijnsen 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 Crijnsen 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 et al., 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 (Wijers 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).

5. 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., 2018). Neural mechanisms of the reaction-attention link. Soc. Cognit. Affect. Neurosci. 13 (5), 501–512. https://doi.org/10.1093/socan/mys025.

Cocconi, C.A., Kooliokian, V., W., K.R.S.-K., Fike, G.B., Evans, A.C., 1997. BrainWeb: online interface to a 3D MRI simulated brain database. Neuroimage 5, 425.

Crone, E.A., Fulligni, A.J., 2020. Self and others in adolescence. Annu. Rev. Psychol. 71, 447–468. https://doi.org/10.1146/annurev-psych-011419-050937.

Cumming, G., 2013. The new statistics: why and how. Psychol. Sci. https://doi.org/10.1177/0956797613504966.

D’Argeambeau, A., 2013. On the role of the ventromedial prefrontal cortex in self- processing: the valuation hypothesis. Front. Hum. Neurosci. 7, 372. https://doi.org/10.3389/fnhum.2013.00372.

Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109. https://doi.org/10.1002/(SICI)1097-0193(19992)8:3<109::AID-HBM1>3.0.CO;2-N.

Davis, M.H., 1983. Measuring individual differences in empathy: evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44 (1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113.

de la Vega, A., Chang, I.L.J., B., T.M., W., T.D., Yarkoni, T., 2016. Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. J. Neurosci. 36 (24), 6553–6562. https://doi.org/10.1523/JNEUROSCI.4402-15.2016.

Denny, B.T., Kober, H., W., T.D., Ochun, K., 2012. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. J. Cognit. Neurosci. 24 (8), 1742–1752. https://doi.org/10.1162/jocn.2012.00233.

Eisenberg, N., Eggum, N.D., Gionta, L.D., 2010. Empathy-related responding: associations with prosocial behavior, aggression, and Intergroup relations. Soc.Issues Pol. Rev. 4 (1), 1–180. https://doi.org/10.1751-2490.2010.01020.x.

Elliott, M.L., Knodt, A.R., Ireland, D., Morris, M.L., Poulton, R., Ramakshma, S., et al., 2020. What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychol. Sci. 9567976201916786.

Euler, Felix, Stefin, Celia, Stadler, Christina, 2017. Distinct profiles of reactive and proactive aggression in adolescents: associations with cognitive and affective empathy. Child Adolesc. Psychiatr. Ment. Health 11 (1), 1–14. https://doi.org/10.1186/s13034-016-0141-4.

Feng, C., Yan, F., 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.

Flagan, T., Beers, J.S., 2013. Three ways in which midline regions contribute to self- evaluation. Front. Hum. Neurosci. 7, 450. https://doi.org/10.3389/fnhum.2013.00450.

Gilbert, S.J., Spengler, S., Simons, J.S., Steel, J.D., Lawrie, S.M., Frith, C.D., Burgess, P.W., 2006. Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. J. Cognit. Neurosci. 18 (6), 932–948. https://doi.org/10.1162/jocn.2006.18.6.932.

Gorgolewski, K.J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S.S., Mument, C., Sochat, V.V., Nichols, T.E., Fuldcrad, 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. Front. Neuroinf. 9 https://doi.org/10.3389/fnhum.2015.00008.

Gunther, Moor, B., Van Leijenhorst, L., Rombouts, S.A., Crone, E.A., Van der Molen, M.W., 2016. Control your anger! the neural basis of aggression regulation in response to positive social feedback. Soc. Cognit. Affect. Neurosci. 11 (5), 712–720. https://doi.org/10.1093/socan/mvs154.

Harter, S., 2012. The construction of the self. In: Developmental and Sociocultural Foundations, second ed. The Guilford Press, New York.

Hoppénbrouwers, 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), 1377–1387. https://doi.org/10.1016/j.cortex.2012.06.003.

Hughes, B.L., Beers, J.S., 2013. Protecting the self: the effect of social-evaluative threat on neural representations of self. J. Cognit. Neurosci. 25 (4), 613–622. https://doi.org/10.1162/jocn.2013.00042.

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. https://doi.org/10.1037/0012-1649.43.4.1038.

Korn, C.W., Prehn, K., Park, S.O., Walter, H., Heekeren, H.R., 2012. Positively biased processing of self-relevant social feedback. J. Neurosci. 32 (47), 16832–16844. https://doi.org/10.1525/JNEUROSCI.3016-12.2012.

Koski, J.E., McHaney, J.R., Rigney, A.E., Beer, J.S., 2020. Reconsidering longstanding assumptions about the role of medial prefrontal cortex (MPPC) in social evaluation. Neuroimage 214, 116752. https://doi.org/10.1016/j.neuroimage.2020.116752.

I.H. Van de Groep et al.

Neuropsychologia 161 (2021) 108000
Lieberman, M.D., Straccia, M.A., Mayer, 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. https://doi.org/10.1016/j.neubiorev.2018.12.021.

Moffitt, T.E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R.J., Harrington, H., et al., 2011. A gradient of childhood self-control predicts health, wealth, and public safety. Proc. Natl. Acad. Sci. U. S. A. 108 (7), 2693–2698. https://doi.org/10.1073/pnas.1007067108.

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. Soc. Neurosci. 8 (2), 178–188. https://doi.org/10.1080/17470919.2012.720602.

Pfeifer, J.H., Kahn, L.E., Merchant, J.S., Peake, S.J., Veroude, K., Masten, C.L., et al., 2013. Longitudinal change in the neural bases of adolescent social self-evaluations: effects of age and pubertal development. J. Neurosci. 33 (17), 7415–7419. https://doi.org/10.1523/JNEUROSCI.4074-12.2013.

Raine, A., Dodge, K., Locner, R., Gatzke-Kopp, L., Reynolds, C., Liu, J., 2006. Proactive aggression questionnaire: differential correlates of reactive and proactive aggression in adolescent boys. Aggress. Behav. 32 (2), 159–171. https://doi.org/10.1002/ab.20115.

Reniers, R.L., Corcoran, R., Drake, R., Shryane, N.M., Vollm, B.A., 2011. The QCAE: a questionnaire of cognitive and affective empathy. J. Pers. Assess. 93 (1), 84–95. https://doi.org/10.1080/00223891.2010.528494.

Riva, P., Romero Lauro, L., DeWall, C.N., Chester, D.S., Bushman, B.J., 2015. Reducing aggressive responses to social exclusion using transcranial direct current stimulation. Dev. Cognit. Neurosci. 2 (1), 55–69. https://doi.org/10.1016/j.dcn.2011.07.012.

Riva, P., Kahn, L.E., Merchant, J.S., Peake, S.J., Veroude, K., Masten, C.L., et al., 2013. Longitudinal change in the neural bases of adolescent social self-evaluations: effects of age and pubertal development. J. Neurosci. 33 (17), 7415–7419. https://doi.org/10.1523/JNEUROSCI.4074-12.2013.

Somerville, L.H., Kerley, W.M., Heatherton, T.F., Kelley, W.M., 2006. Anterior cingulate cortex responds differentially to expectancy violation and social rejection. Nat. Neurosci. 9 (8), 1007–1008. https://doi.org/10.1038/nn1728.

Van der Crijns, R., Peters, S., Crane, E.A., 2017. Neural correlates of evaluating self and close-others in physical, academic and prosocial domains. Brain Cognit. 118, 45–53. https://doi.org/10.1016/j.bandc.2017.07.008.

Van der Crijns, R., Peters, S., van der Aar, L.P.E., Crane, E.A., 2018. The neural signature of self-concept development in adolescence: the role of domain and valence distinctions. Dev. Cognit. Neurosci. 30, 1–12. https://doi.org/10.1016/j.dcn.2017.11.005.

Veroude, K., Jolles, J., Croiset, G., Krabbendam, L., 2014. Sex differences in the neural bases of social appraisals. Soc. Cognit. Affect Neurosci. 9 (4), 513–519. https://doi.org/10.1093/scan/nst015.

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. Nat. Rev. Neurosci. 15 (3), 193–201. https://doi.org/10.1038/nrn3679.

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 https://doi.org/10.1038/s41467-018-05533-2.

Zheng, W., Li, Y., Ye, H., Luo, J., 2021. Effect of modulating DLPFC activity on antisocial and prosocial behavior: evidence from a tDCS study. Front. Psychol. 11, 579792 https://doi.org/10.3389/fpsyg.2020.57979.