# Like me, or else: Nature, nurture and neural mechanisms of social emotion regulation in childhood

Achterberg, M.

# CHAPTER FOUR

# Heritability of aggression following social evaluation in middle childhood: An fMRI study

This chapter is published as: Achterberg M., Van Duijvenvoorde A.C.K., Van der Meulen M., Bakermans M.J. & Crone E.A.M. (2018), Heritability of aggression following social evaluation in middle childhood: An fMRI study, *Human Brain Mapping 39*(7): 2828-2841.

# Abstract

Middle childhood marks an important phase for developing and maintaining social relations. At the same time this phase is marked by a gap in our knowledge of the genetic and environmental influences on brain responses to social feedback and their relation to behavioral aggression. In a large developmental twin sample (509 7-9-year-olds) the heritability and neural underpinnings of behavioral aggression following social evaluation were investigated, using the Social Network Aggression Task (SNAT). Participants viewed pictures of peers that gave positive, neutral or negative feedback to the participant's profile. Next, participants could blast a loud noise towards the peer as an index of aggression. Genetic modeling revealed that aggression following negative feedback was influenced by both genetics and environmental (shared as well as unique environment). On a neural level (n=385), the anterior insula and anterior cingulate cortex gyrus responded to both positive and negative feedback, suggesting they signal for social salience cues. The medial prefrontal cortex and inferior frontal gyrus were specifically activated during negative feedback, whereas positive feedback resulted in increased activation in caudate, supplementary motor cortex (SMA) and dorsolateral prefrontal cortex (DLPFC). Decreased SMA and DLPFC activation during negative feedback was associated with more aggressive behavior after negative feedback. Moreover, genetic modeling showed that 13-14% of the variance in dorsolateral PFC activity was explained by genetics. Our results suggest that the processing of social feedback is partly explained by genetic factors, whereas shared environmental influences play a role in behavioral aggression following feedback.

*Keywords: Behavioral genetics; Dorsolateral prefrontal cortex; Peer feedback; Twin study*

# Introduction

Dealing with social evaluations and regulating emotions in the case of negative social feedback are important prerequisites for developing social relations. Several prior studies have shown that negative social feedback can lead to aggressive behavior (Chester *et al.*, 2014; Achterberg *et al.*, 2016b; Achterberg *et al.*, 2017). This type of retaliation may be associated with emotional responses to negative feedback and a lack of impulse control. The capacity to regulate impulsive behavior increases from childhood to adulthood, which has been linked to the increased regulatory control of the prefrontal cortex (PFC) (Somerville *et al.*, 2010; Casey, 2015). Indeed, prior studies in adults showed that stronger brain connectivity between nucleus accumbens and the lateral PFC was related to lower retaliatory aggression (Chester and DeWall, 2016). Moreover, increased dorsolateral PFC (DLPFC) activity after negative social feedback has been associated to less subsequent aggression (Riva *et al.*, 2015; Achterberg *et al.*, 2016b). Therefore, the prefrontal cortex may be important for regulation of neural responses to social emotions and may signal which children are better able to regulate emotions than others. Middle childhood, ranging from approximately 7/8 years until the start of puberty, marks an important phase for regulating (social) emotions and developing social relations. Previous studies have mainly focused on the developmental trajectories of social rejection and acceptance (Guyer *et al.*, 2008; Gunther Moor *et al.*, 2010b; Silk *et al.*, 2014; Guyer *et al.*, 2016). At the same time there is a gap in our understanding of the genetic and environmental influences of brain responses to social feedback and regulatory responses. In this study, we therefore investigated the neural underpinnings and heritability of social feedback processing and subsequent aggression in middle childhood.

The way children respond to social feedback and show aggression in response to negative feedback has only recently been examined using experimental designs. Studies including children, adolescents and adults have used social feedback tasks in chat room settings to unravel neural responses to social feedback, namely social acceptance and rejection (Guyer *et al.*, 2016). These studies point to the anterior cingulate cortex  gyrus (ACCg), the medial prefrontal cortex (mPFC), and the anterior insula as important brain regions related to social evaluation and social motivation (Cacioppo *et al.*, 2013; Rotge *et al.*, 2015; Apps *et al.*, 2016). The dorsal ACC / ACCg was found to be activated in response to unexpected social feedback, irrespective of whether it was positive or negative (Somerville *et al.*, 2006). Recently, we developed a social network aggression task (SNAT) to study neural responses to social feedback, both in adults and 7-10-year-old children (Achterberg *et al.*, 2016b; Achterberg *et al.*, 2017). Consistent with prior studies, the ACCg and the anterior insula were active during both positive and negative feedback in adults, indicating that these regions signal social salient cues (Achterberg *et al.*, 2016b). These effects were

also present in middle childhood, but less pronounced (Achterberg *et al.*, 2017). However, prior studies in children used relatively small samples, which might have been underpowered, specifically since neuroimaging data in developmental samples are more prone to data loss and artifacts due to movement (O'Shaughnessy *et al.*, 2008). The current study therefore set out to include over 500 participants, thereby asserting sufficient sample size and statistical power, even after data loss due to excessive motion (Euser *et al.*, 2016).

Prior studies in adults showed that the DLPFC was negatively related to aggression following social evaluation, suggesting that this region is important for regulating aggression (Achterberg *et al.* (2016b), see also Riva *et al.* (2015)). Since the PFC gradually develops until early adulthood (Lenroot and Giedd, 2006; van Duijvenvoorde *et al.*, 2016a), there is ample opportunity for environmental influences. An important question therefore concerns to what extent behavioral and neural responses to social feedback, and subsequent aggression, are influenced by genetic and/or shared environmental factors. Twin models have been particularly important in unraveling to what extent genetic and environmental factors account for the variance in aggression. These studies have shown that trait aggression has both genetic and environmental components (Porsch *et al.*, 2016). Heritability estimates for behavioral aggression are high for both children and adults, explaining up to 48% of the variance (for meta-analyses, see Rhee and Waldman (2002); Ferguson (2010); Tuvblad and Baker (2011)). We aimed to explore whether neural reactions to social feedback that could elicit aggression show similar heritability estimates. Studies of the genetics of functional neuroimaging are currently limited to studies using resting state fMRI (Richmond *et al.*, 2016) or cognitive working memory tasks (Jansen *et al.*, 2015). These studies mostly point to (moderate) genetic influences, with few studies showing significant shared environmental components. It should be noted that these findings are largely based on adult twin studies, whereas previous research showed that heritability estimates of brain measures are stronger in adulthood than in childhood (Lenroot *et al.*, 2009; Lenroot and Giedd, 2011; van den Heuvel *et al.*, 2013). In this study we therefore used a large developmental twin sample (*N*=509 7-9-year-olds), to investigate i) the heritability of behavioral aggression following social evaluation; ii) the neural underpinnings of social evaluation and their relation to behavioral aggression; and iii) the heritability of these neural underpinnings.

We hypothesized that negative social feedback would result in behavioral aggression (Chester *et al.*, 2014; Achterberg *et al.*, 2016b; Achterberg *et al.*, 2017). Prior studies have shown that trait aggression has a relatively strong genetic component (Porsch *et al.*, 2016), however the influences of genetics and environment on state aggression such as measured with the SNAT are not yet known. On a neural level, we predicted to find a network of regions that process social feedback irrespective of valence, as prior research showed in adults (Achterberg *et al.*, 2016b), including the ACCg and the (anterior) insula. In

addition, we will investigate possible brain-behavior relations between activation of these regions and the aggression measure. Based on prior studies (Riva *et al.*, 2015; Achterberg *et al.*, 2016b), we predicted that the lateral prefrontal cortex would be most strongly correlated to aggression regulation. Since the literature on the heritability of task-based fMRI is limited, and the current study is the first to study such heritability in middle childhood, no a priori hypotheses were formed for the exploratory analyses on heritability of neural activation.

# Methods

## Participants

Participants in this study took part in the longitudinal twin study of the Leiden Consortium on Individual Development (L-CID). The Dutch Central Committee Human Research (CCMO) approved the study and its procedures. Families with a twin born between 2006 – 2009, living within two hours travel time from Leiden, were recruited through municipal registries and received an invitation to participate by post. Parents could show their interest in participation using a reply card. 512 children (256 families) between the ages 7 and 9 were included in the L-CID study. Written informed consent was obtained from both parents. All twin-pairs had a shared home environment, were fluent in Dutch, and had normal or corrected-to-normal vision. The majority of the sample was Caucasian (91%) and right-handed (87%). Since the sample represents a population sample, we did not exclude children with a psychiatric disorder. Ten participants (2%) were diagnosed with an Axis-I disorder: eight with attention deficit hyperactivity disorder (ADHD); one with generalized anxiety disorder (GAD), and one with pervasive developmental disorder- not otherwise specific (PDD-NOS). Three participants did not have data from the SNAT due to technical problems. Therefore, our final behavioral sample consisted of 509 participants with a mean age of 7.95 ± 0.67 (age range: 7.02-9.68, 49% boys, see Table 1), with 253 complete twin pairs (55% MZ; based on DNA, see section 2.5). Data from 30 twin pairs were previously reported (Achterberg et al., 2017).

Twenty-seven participants did not perform the SNAT in the MRI scanner: 13 due to anxiety, 6 due to MRI contra-indications, 4 participants did not have parental consent for MRI participation, and 4 participants could not be scanned due to technical system failure. For all participants who underwent the MRI scan, anatomical MRI scans were reviewed and cleared by a radiologist from the radiology department of the Leiden University Medical Center (LUMC). Four anomalous findings were reported. To prevent registration errors due to anomalous brain anatomy, these participants were excluded. An additional 89 participants were excluded due to excessive head motion, which was defined as >3 mm motion (1 voxel) in any direction (x, y, z) in more than 2 blocks of the SNAT task (3 blocks in total). Finally, four participants were excluded due to

preprocessing errors. Our final MRI sample consisted of 385 participants with a mean age of 7.99 ± 0.68 (age range: 7.02-9.68, 47% boys, see Table 1), with 158 complete twin pairs (55% MZ; based on DNA, see section 2.5). Participants' intelligence (IQ) was estimated with the subsets 'similarities' and 'block design' of the Wechsler Intelligence Scale for Children, third edition (WISC-III; Wechsler, 1997). Estimated IQs were in the normal range (72.50 - 137.50), with an average IQ of 104 (see Table 1). There were no significant differences in IQ between children in the final sample (n=385) and those who could not be included in the MRI analyses (n=124) ($t(507)=1.36$, $p=.175$), nor were there significant gender differences ($\chi(1, N=512)=2.80$, $p=.092$). Children that could not be included in the MRI analyses were, however, significantly younger ($M=7.80$, $SD=0.64$) than children in the final sample ($M=7.99$, $SD=0.67$, $t(507)=2.72$, $p=.007$), but this effect was small ($d=0.29$).

**Table 1.** Demographic characteristics.

| | **Behavioral sample** | **MRI sample** |
|---|---|---|
| N | 509 | 385 |
| Boys | 49% | 47% |
| Left handed | 13.0% | 12.0% |
| Caucasian | 91.0% | 93.0% |
| AXIS-I disorder | 10 (2%)[1] | 8 (2%)[2] |
| Age (SD) | 7.94 (.67) | 7.99 (.68) |
| Range | 7.02 - 9.68 | 7.02 - 9.68 |
| Mean IQ (SD) | 103.62 (11.77) | 104.03 (11.84) |
| IQ range | 72.50 - 137.50 | 72.50 - 137.50 |
| | | |
| Complete twin pairs | 253 | 158 |
| Monozygotic | 138 (55%) | 87 (55%) |
| Caucasian | 230 (91%) | 150 (95)% |

[1] 8 ADHD; 1 PDD-NOS; 1 Generalized Anxiety Disorder

[2] 6 ADHD; 1 PDD-NOS; 1 Generalized Anxiety Disorder

## Social Network Aggression Task

### Experimental design

The Social Network Aggression Task (SNAT) as described in Achterberg et al. (2016b; 2017) was used to measure (imagined) aggression after social evaluation. Prior to the fMRI session, the children filled in a personal profile at home, which was handed in at least one week before the actual fMRI session. The profile page

consisted of questions such as: 'What is your favorite movie?', 'What is your favorite sport?', and 'What is your biggest wish?'. Children were informed that their profiles were reviewed by other, unfamiliar, children. During the SNAT the children were presented with pictures and feedback from same-aged peers in response to their personal profile. Every trial consisted of feedback from a new unfamiliar child. This feedback could either be positive ('I like your profile', visualized by a green thumb up); negative ('I do not like your profile', red thumb down) or neutral ('I don't know what to think of your profile', grey circle). Following each peer feedback, the children were instructed to imagine that they could send a loud noise blast to this peer. We specifically instructed the children to imagine this to reduce deception, because it has been shown that imagined play also leads to aggression (Konijn et al., 2007). The longer they pressed the button the more intense the noise would be, which was visually represented by a volume bar (Figure 1). To keep task demands as similar as possible between the conditions, participants were instructed to always press the button, but they could choose whether they wanted a short noise at low intensity or a long noise at high intensity. Unbeknownst to the participants, others did not judge the profile, and the photos were created by morphing two children of an existing data base (matching the age range) into a new, non-existing child. Peer pictures were randomly coupled to feedback, ensuring equal gender proportions for each type of feedback.



Participants were familiarized with the MRI scanner during a practice session in a mock scanner. Then participants received instructions on how to perform the SNAT and the children were exposed to the noise blast twice during a practice session: once with stepwise build-up of intensity and once at maximum intensity. Participants did not hear the noise during the fMRI session, to prevent that they would punish themselves by pressing the button. To familiarize participants with the task, participants performed six practice trials. After the mock scanner session, one of the twins continued with the actual scan, while the other twin performed the WISC-III and other behavioral tasks. First-born and second-born children were randomly assigned to the scan session or behavioral tasks as their first task. When the first child completed the scan, he/she continued with the WISC-III and behavioral tasks while the other child participated in the scanning session.

The SNAT consisted of 60 trials, three blocks of 20 trials for each social feedback condition (positive, neutral, negative), that were presented semi-randomized to ensure that no condition was presented more than three times in a row. The optimal jitter timing and order of events were calculated with Optseq 2 (Dale, 1999). Each trial started with a fixation screen (500 ms), followed by social feedback (2500 ms). After another jittered fixation screen (3000-5000 ms), the noise screen with the volume bar appeared, which was presented for a total of 5000 ms. Children were instructed to deliver the noise blast by pressing one of the buttons on the button box attached to their legs, with their right index

finger. As soon as the participant started the button press, the volume bar started to fill up with a new colored block appearing every 350 ms. After releasing the button, or at maximum intensity (after 3500 ms), the volume bar stopped increasing and stayed on the screen for the remainder of the 5000 ms. Before the start of the next trial, another jittered fixation cross was presented (0 -11550 ms) (Figure 1). The length of the noise blast duration (i.e., length of button press) in milliseconds was used as a measure of imagined aggression.

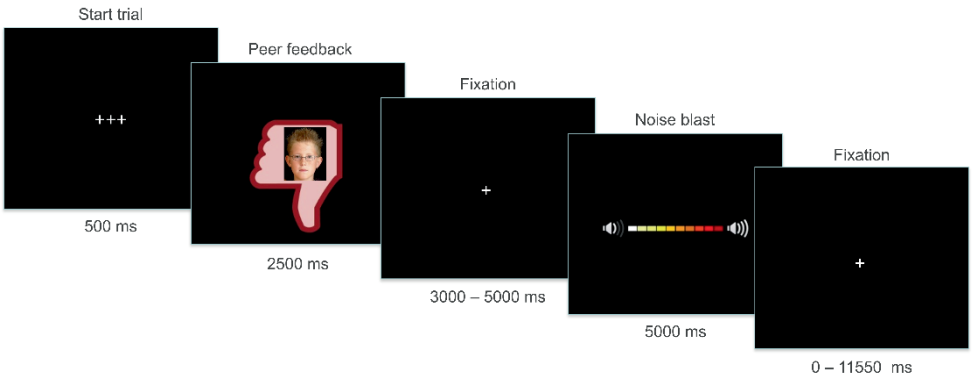


***Figure 1**. Example of one trial of the social network aggression task.*

## Social feedback manipulation check

The social feedback manipulation was checked using an exit interview with questions on how much they liked the feedback (*'How much did you like reactions with a thumb up?'*, *'How much did you like reactions with a circle?'*, and *'How much did you like reactions with a thumb down?')*. Participants rated the reactions on a 6-point scale, with 1 representing *very little* and 6 representing *very much*. In addition, we asked two open questions: *'what did you think of the game?'*, and '*what did you think of the noises that you could deliver'*. None of the participants expressed doubts about the cover story.

To verify whether children differentially evaluated the social feedback conditions (positive, negative, neutral), we analyzed answers to the exit questions with a repeated measures ANOVA. Data from the exit questions were missing for 5 participants. Results (Greenhouse-Geisser corrected) showed a significant main effect of type of feedback on the subjective evaluation of social feedback with a large effect size ($F(2, 1002)= 19.16$, $p<.001$, $\omega^2=0.62$). Pairwise comparisons showed that participants liked negative feedback ($M=2.27$, $SD=1.18$) significantly less than neutral feedback ($M=4.14$, $SD=0.87$, $p<.001$, $d= 1.80$) and positive feedback ($M=5.33$, $SD=0.88$, $p<.001$, $d= 2.94$). Participants also liked neutral feedback significantly less than positive feedback ($p<.001$, $d= 1.37$).

## MRI data acquisition

MRI scans were acquired with a standard whole-head coil on a Philips Ingenia 3.0 Tesla MR system. To prevent head motion, foam inserts surrounded the children's heads. The total scan protocol lasted 56 minutes, including two fMRI tasks, high resolution T2 and T1 scans, diffusion tensor imaging scans and a resting state fMRI scan. The order of the scans was the same for all participants and always started with the SNAT. The SNAT was projected on a screen that was viewed through a mirror on the head coil. Functional scans were collected during three runs T2*-weighted echo planar images (EPI). The first two volumes were discarded to allow for equilibration of T1 saturation effect. Volumes covered the whole brain with a field of view (FOV) = 220 (ap) x 220 (rl) x 111.65 (fh) mm; repetition time (TR) of 2.2 seconds; echo time (TE) = 30 ms; flip angle (FA) = 80°; sequential acquisition, 37 slices; and voxel size = 2.75 x 2.75 x 2.75 mm. Subsequently, a high-resolution 3D T1scan was obtained as anatomical reference (FOV= 224 (ap) x 177 (rl) x 168 (fh); TR = 9.72 ms; TE = 4.95 ms; FA = 8°; 140 slices; voxel size 0.875 x 0.875 x 0.875 mm).



## MRI data analyses

### Preprocessing

MRI data were analyzed with SPM 8 (Wellcome Trust Centre for Neuroimaging, London). Images were corrected for slice timing acquisition and rigid body motion. Functional scans were spatially normalized to T1 templates. Due to T1 misregistrations, five participants were normalized to an EPI template. Volumes of all participants were resampled to 3x3x3 mm voxels. Data were spatially smoothed with a 6 mm full width at half maximum (FWHM) isotropic Gaussian kernel. Translational movement parameters were calculated for all participants. Participants that had at least two blocks of fMRI data with <3 mm (1 voxel) motion in any direction were included (*N*=385).

### First-level analyses

Statistical analyses were performed on individual subjects' data using a general linear model. The fMRI time series were modeled as a series of two events convolved with the hemodynamic response function (HRF). The onset of social feedback was modeled as the first event, with a zero duration and with separate regressors for the positive, negative, and neutral peer feedback. The start of the noise blast (second event) was modeled 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 judgments. Trials on which the participants failed to respond in time were modeled separately as covariate of no interest and were excluded from further analyses. All participants had at least 10 trials for each feedback type. To account for possible motion induced error that had not been

solved by realignment, we included six additional motion regressors (corresponding to the three translational and rotational directions) as covariates of no interest. The least squares parameter estimates of height of the best-fitting canonical HRF for each condition were used in pairwise contrasts. The pairwise comparisons resulted in subject-specific contrast images.

**Higher-level group analyses**

Subject-specific contrast images were used for the group analyses. A full factorial ANOVA with three levels (positive, negative and neutral judgment) was used to investigate the neural response to the social feedback event. To investigate regions that were activated during both negative and positive feedback, we conducted a conjunction analysis to explore the general valence effects of social evaluation (conjunction negative > neutral and positive > neutral). Based on Nichols *et al.* (2005), we used the 'logical AND' strategy. The 'logical AND' strategy requires that all the comparisons in the conjunction are individually significant (Nichols *et al.*, 2005). Next, we calculated the contrasts negative > positive and positive > negative to investigate brain regions that were specifically activated for social rejection or social acceptance. All results were family wise error (FWE) voxel level corrected, with $p_{FWE}<.05$. Coordinates for local maxima are reported in MNI space.

**Region of Interest analyses**

SPM8's MarsBaR toolbox (Brett *et al.*, 2002) was used to extract patterns of activation from the whole brain group analyses in order to investigate possible brain-behavior associations and as input for the genetic modeling. Parameter estimates (PE, average Beta values) were extracted from regions that were significantly activated in the whole brain analyses. Specifically, the following ten regions were extracted: the left and right insula and ACCg (from the conjunction contrast); the mPFC and left and right IFG (contrast negative>positive); and the left and right DLPFC, SMA, and caudate (contrast positive>negative). For the brain-behavior relations we focused on associations with noise-blast difference scores following negative social feedback (negative-positive and negative-neutral, corrected for age and IQ).

## Genetic modeling

Zygosity was determined using DNA analyses. DNA was tested with buccal cell samples collected via a mouth swab (Whatman Sterile Omni Swab). Buccal samples were collected directly after the MRI session, thereby ensuring that the children did not have anything to eat or drink for at least one hour prior to DNA collection. The results of the DNA analyses indicated that 55% of the twin pairs was MZ.

Phenotypic similarities among twin pairs can be divided into similarities due to shared genetic factors (A) and shared environmental factors (C), while dissimilarities are ascribed to unique environmental influences and measurement error (E). We used behavioral genetic modeling with the OpenMX package (Neale *et al.*, 2016) in R (R Core Team, 2015) to get an estimate of these A, C, and E components. Comparisons of the ACE model with more parsimonious models (AE model; CE model; or E model) are described in the Supplementary Materials. When ACE models show the best fit, both heritability, shared and unique environment are important contributors to explain the variance in the outcome variable. AE models indicate that genetic and unique environmental factors play a role; whilst CE models indicate influences of the shared environment and unique environment. If the E model has no worse fit than AE or CE models, variance in the outcome variable is accounted for by unique environmental factors and measurement error.



## Statistical Analysis

In order to detect outliers in the data, we transformed the raw data to z-values. Based on the Z-distribution, 99.9% of z-scores lie between -3.29 and +3.29. Z-values outside this range (<-3.29 or >3.29) were defined as outliers. Outlying scores were winsorized (Tabachnick and Fidell, 2013). To assess effects of condition (positive, neutral, negative) on noise blast duration (in ms) we used a linear mixed-effect model approach using the lme4 package in R (Bates *et al.*, 2015) in R (R Core Team, 2015). Data was fitted on the average response times after positive, neutral and negative trials. Random intercepts per participants and per family allows to account for the nesting of condition within participant (ChildID) and the nesting of twin-pairs within families (FamilyID). Additionally, a random slope of condition was included per participant. Fixed effects included condition (factor with 3 levels), as well as participant's age and IQ as covariates, which were grand mean centered. All main effects and two-way interactions between age $^{x}$ condition and age $^{x}$ IQ were included. P-values were determined using Kenward-Rogers approximation as implemented in the mixed function in the afex package (Singmann, 2013). The fitted mixed-effect model is specified in R as:

$$noiseblast \sim condition * age\_meancentered + condition * IQ\_meancentered + (condition|childID) + (1|familyID).$$

To derive a measure of individual differences in aggression we calculated the differences in noise blast duration between conditions (negative-positive; negative-neutral; neutral-positive). Brain-behavior associations were investigated by least square regressions with ROI activation predicting noise blast difference scores. Due to the nested nature of twin data, the data violates the assumption

of homoscedasticity. Although the estimator of the regression parameters is not influenced when this assumption is violated, the estimator of the covariance matrix can be biased, resulting in too liberal or too conservative significance tests (Hayes & Chai, 2007). Therefore, we used heteroscedasticity-consistent standard error (HCSE) estimators, by using the HCSE macro of Hayes and Cai (2007). As recommended by Long and Ervin (2000), we used the HC3 method. Moreover, we performed genetic modeling of behavioral responses (noise blast difference scores) and neural responses (ROI activation) to social feedback using the OpenMX package (Neale *et al.*, 2016) in R (R Core Team, 2015).

# Results

## Behavioral analyses

### Social feedback retaliation

The linear mixed-effect model showed a significant main effect of type of social feedback on noise blast duration, $F(2, 505) = 300.8754$, $p<.001$. Pairwise comparisons revealed that noise blast duration after negative feedback ($M$=2688, $SD$=736) was significantly longer than noise blast duration after neutral feedback ($M$=1906, $SD$=648, $p<.001$), and after positive feedback ($M$=1459, $SD$=852, $p<.001$). Noise blast duration was significantly longer after neutral feedback than after positive feedback ($p<.001$). There were also significant noise blast $^{x}$ age $F(2, 505) = 10.57$, $p<.001$) and noise blast $^{x}$ IQ interaction effects $F(2, 505) = 12.27$, $p<.001$), showing larger condition effects for older children and for children with higher IQ. To control for possible confounding effects of age and IQ, we included these variables as regressors in further models. There were no significant gender differences in noise blast duration after positive, neutral or negative feedback (independent sample T-tests, all $p$'s>.05). Results did not change after exclusion of children with an Axis-I disorder.

### Twin analyses

To investigate twin-effects in (imagined) aggression after social feedback we calculated the differences in noise blast duration between negative versus positive feedback, negative versus neutral feedback; and neutral versus positive feedback. Next, we performed Pearson's correlations between these differences scores within MZ ($n$=138) and DZ ($n$=115) twin pairs (Table 2). Behavioral genetic analyses revealed that aggression following negative relative to positive social feedback was moderately influenced by genetics (A= 20%, 95% CI: 0-37%), and to a lesser extent influenced by shared environment (C=6%, 95% CI: 0-34%). Unique environment and measurement error explained the largest part of the variance in aggression after negative feedback (E=74%, 95% CI:0.63-0.90), see Table 2. The best fitting model was an ACE-model, see Table S1. Aggression following negative

relative to neutral feedback showed similar influences of shared environment (C=8%) and relatively less influence of genetics (A=10%, Table 2), and was best described by a CE-model (Table S1). Aggression following neutral relative to positive social feedback showed no influence of shared environment (C=0%) and was most influenced by unique environment (90%, see Table 2 and Table S1).

**Table 2.** Noise blast twin analyses. Pearson's correlations and ACE models for noise blast difference scores.

| **Noise blast difference** | | **MZ** | **DZ** | | **A²** | **C²** | **E²** |
|---|---|---|---|---|---|---|---|
| Negative - Positive | *r* | .21 | .24 | *ACE* | 0.20 | 0.06 | 0.74 |
| | *p* | .016 | .010 | *95% CI* | 0.00 - 0.37 | 0.00 - 0.34 | 0.63 - 0.90 |
| Negative - Neutral | *r* | .19 | .25 | *ACE* | 0.10 | 0.08 | 0.82 |
| | *p* | .025 | .007 | *95% CI* | 0.00 - 0.40 | 0.00 - 0.32 | 0.60 - 0.98 |
| Neutral - Positive | *r* | .10 | .04 | *ACE* | 0.10 | 0.00 | 0.90 |
| | *p* | .260 | .67 | *95% CI* | 0.00 - 0.26 | 0.00 - 0.13 | 0.74 - 1.00 |

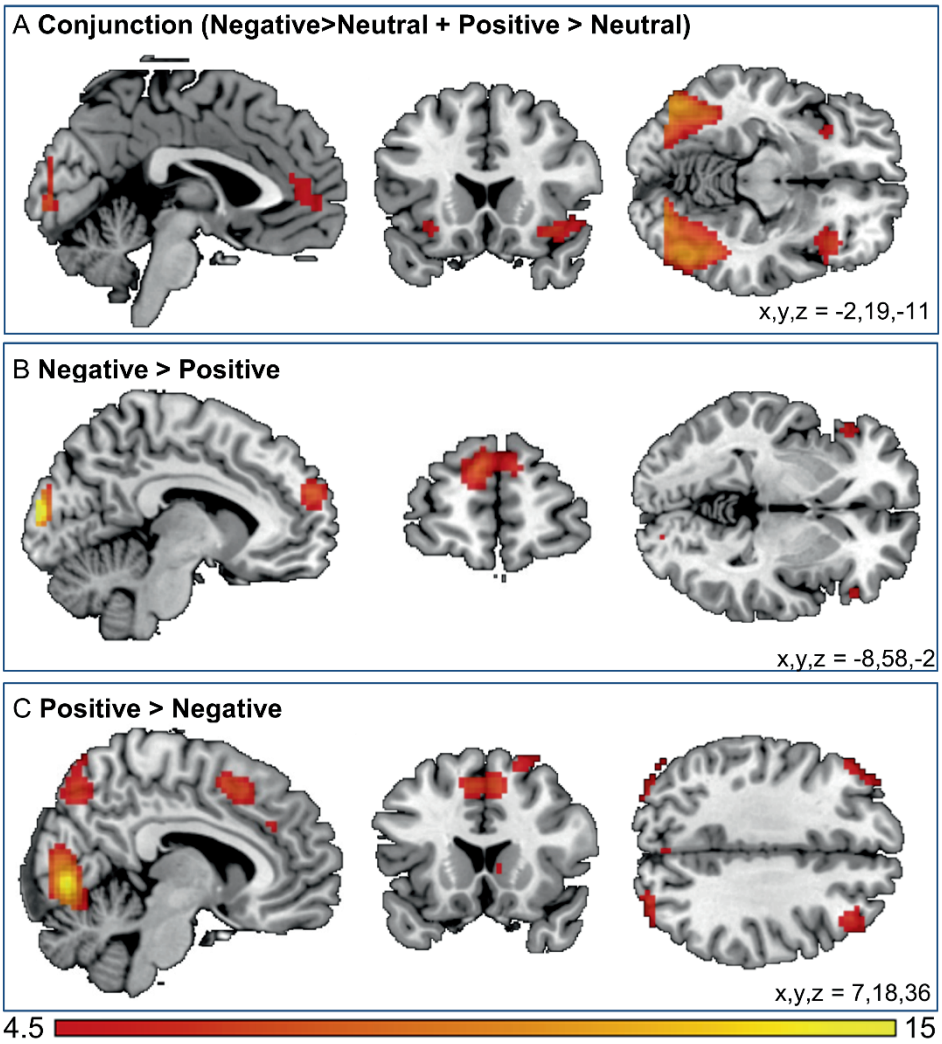


***Figure 2.*** *Whole brain results for A) the conjunction negative>neutral and positive>neutral; B) the contrast negative>positive; and C) the contrast positive>negative. Results were family wise error corrected ($p_{FWE}$<.05).*

## Neural analyses

### Whole brain analyses

To investigate the general valence effects of social feedback, we examined neural activity for positive versus neutral *and* negative versus neutral feedback using a conjunction analysis. We found common activation across positive and negative feedback in a wide network of regions including left and right insula, the ACCg, and the lateral occipital cortex (Figure 2a and Table 3).

To investigate effects of negative versus positive social feedback, we investigated the contrasts negative>positive and positive>negative. The contrast negative>positive feedback resulted in activation with local maxima in the medial PFC, the left and right inferior frontal gyrus (IFG), and the occipital pole (Figure 2b and Table 3). The reversed contrast positive>negative resulted in increased activation in the left and right orbitofrontal cortex (OFC), the precuneus, the supplementary motor cortex (SMA), the right caudate, the left and right DLPFC, and the lingual gyrus (Figure 2c and Table 3). Results did not change after exclusion of children with an Axis-I disorder (Table S3).



## Brain-behavior analyses

To investigate possible brain-behavior associations in the clusters from the whole brain contrasts 10 ROIs were selected based on a priori hypotheses to predict behavioral aggression using least square regressions with HCSE. We chose 3 ROIs from the conjunction (the ACCg, the left insula, and the right insula), 3 ROIs from the contrast negative>positive (the mPFC, the left IFG, and the right IFG) and 4 ROIs from the contrast positive>negative (the SMA, the right caudate, the left DLPFC, and the right DLPFC) (Table 3). We focused on associations with noise-blast difference scores following negative social feedback (negative-positive and negative-neutral, corrected for age and IQ). We observed a significant association between noise blast differences and activity in left DLPFC, right DLPFC activation, and SMA activation (Table 4, Figure 3). These associations showed that greater activation during positive (versus negative) social evaluation was associated with more aggression after negative social feedback, see Figure 3a-d. To visualize this effect in more detail, we plotted the PE's of the right DLPFC for participants with low aggression after negative feedback and participants with high aggression after negative feedback (Figure 3e). Groups were based on a median split of the noise-blast difference scores following negative social feedback (negative-positive, corrected for age and IQ). Participants who differentiated more in aggression (larger noise blast difference positive versus negative feedback) also differentiated more on a neural level (brain activation after positive versus negative feedback), see Figure 3e. In other words, participants who showed less DLPFC activity during negative feedback relative to positive feedback, were more aggressive after negative feedback. These associations did, however, not survive Bonferroni correction ($p's > 0.025$).

All other ROIs showed no behavioral-brain associations (all *p*'s>.05, see Table 4). Results did not change after exclusion of children with an Axis-I disorder (Table S4). Note that we did not observe any significant clusters of activation scaling with behavior when we performed exploratory whole brain regression analyses with the consecutive noise blast difference scores as covariates of interest (on the contrasts positive>negative, negative>positive, positive>neutral and negative>neutral).

## Twin analyses

To investigate twin-effects we calculated Pearson's correlations for neural activation during social feedback in the 10 ROIs for MZ (*n*=87) and DZ (*n*=71) twins, see Table 5. Behavioral genetic analyses revealed that only variance in activation in regions following positive feedback was influenced by genetic factors. Specifically, genetics accounted for 13% (95% CI: 0-32%) of the variance in left DLPFC activation and for 14% (95% CI: 0 - 34%) of the variance in right DLPFC (Table 5). Ten percent of the variance in SMA (95% CI: 0-31%) and right caudate (95% CI: 0-29%) activation was explained by genetics (see Table 5). Estimates for the shared environment were zero, and all of the residuary variance was explained by E (unique environment and measurement error). Genetic modeling for neural activation in the other ROIs revealed minimal or no influence of either genetics or shared environment (estimates 0-4%), and were best explained by unique environment and/or measurement error (Table 5). Variance in neural activation in all ROIs was best explained by an E-model (Table S2).

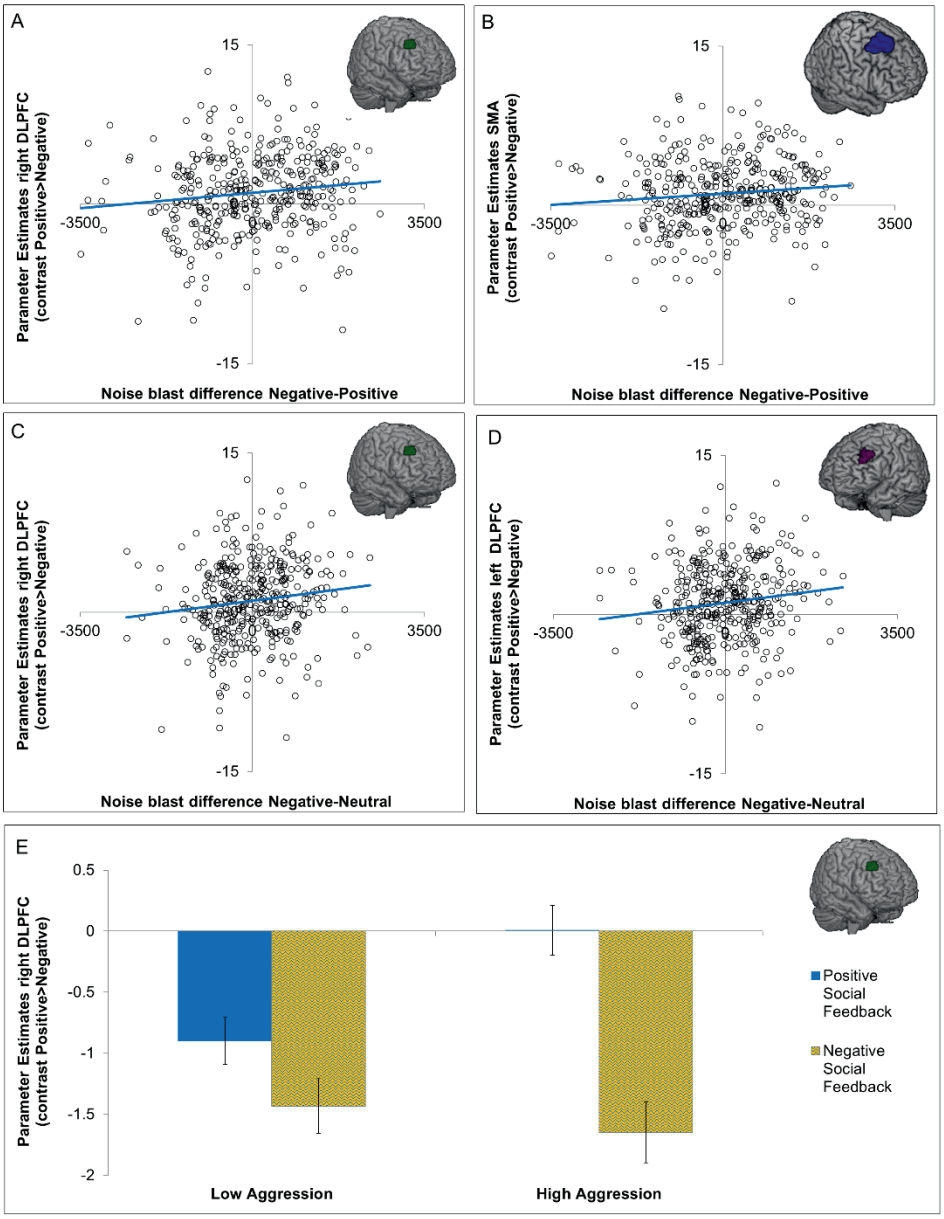


***Figure 3.*** *Visual representation of the brain-behavior associations. (A) right dorsolateral prefrontal cortex (DLPFC) and noise blast difference negative-positive; (B) supplementary motor cortex (SMA) and noise blast difference negative-positive; (C) right DLPFC and noise blast difference negative-neutral; (D) left DLPFC and noise blast difference negative-neutral; and (E) right DLPFC activity after positive and negative social feedback for children with low and high aggression.*

**Table 3.** MNI coordinates for local maxima activated for the whole brain contrasts.

| | Anatomical Region | Voxels | pFWE | T | x | y | z |
|---|---|---|---|---|---|---|---|
| *Conjunction Negative>Neutral and Positive>Neutral* | | | | | | | |
| | Lateral Occipital Cortex | 3550 | <.001 | 14.03 | -42 | -85 | 4 |
| | | | <.001 | 13.79 | -48 | -76 | -5 |
| | | | <.001 | 12.72 | 48 | -70 | -5 |
| | Lateral Occipital Cortex | 124 | <.001 | 6.74 | -24 | -64 | 61 |
| * | Right insula | 101 | <.001 | 6.35 | 39 | 23 | -11 |
| * | Left insula | 30 | .001 | 5.26 | -33 | 26 | -5 |
| | | | .005 | 4.98 | -30 | 20 | -11 |
| | | | .026 | 4.61 | -30 | 11 | -17 |
| * | Rostral ACC | 108 | .002 | 5.19 | 0 | 47 | 10 |
| | | | .002 | 5.18 | -6 | 53 | 1 |
| | | | .004 | 5.02 | 12 | 47 | 13 |
| | Left insula (posterior) | 4 | .007 | 4.93 | -45 | 14 | -5 |
| | Right IFG | 7 | .010 | 4.84 | 51 | 23 | 13 |
| | Supplementary Motor Cortex | 4 | .035 | 4.54 | 6 | 11 | 64 |
| *Negative > Positive* | | | | | | | |
| | Occipital pole | 118 | <.001 | 13.45 | -9 | -97 | 13 |
| * | Medial PFC | 153 | <.001 | 7.16 | -9 | 59 | 25 |
| | | | <.001 | 5.54 | 9 | 59 | 25 |
| | Occipital pole | 51 | <.001 | 6.25 | 27 | -91 | 16 |
| | | | 0.003 | 5.10 | 18 | -94 | 13 |
| * | Left IFG | 66 | <.001 | 6.11 | -54 | 29 | 4 |
| | | | .001 | 5.28 | -45 | 26 | -8 |
| * | Right IFG | 19 | .002 | 5.23 | 51 | 32 | -2 |
| | | | .018 | 4.70 | 57 | 32 | 7 |
| | Left Central Opercular Cortex | 3 | .017 | 4.72 | -36 | -16 | 25 |
| | Left vlPFC | 1 | .042 | 4.49 | -21 | 50 | 7 |
| | Right vlPFC | 1 | .048 | 4.45 | 30 | 50 | -2 |

**Table 3.** (continued)

| | Anatomical Region | Voxels | pFWE | T | x | y | z |
|---|---|---|---|---|---|---|---|
| *Positive > Negative* | | | | | | | |
| | Lingual gyrus | 762 | <.001 | 13.76 | 3 | -76 | -2 |
| | | | <.001 | 11.43 | -18 | -85 | -8 |
| | | | <.001 | 9.63 | -24 | -79 | -11 |
| | Right OFC | 52 | <.001 | 7.58 | 42 | 59 | -8 |
| | | | <.001 | 5.68 | 36 | 56 | -14 |
| * | Supplementary Motor Cortex | 463 | <.001 | 7.43 | -6 | 14 | 49 |
| | | | <.001 | 7.40 | 24 | 5 | 55 |
| | | | <.001 | 6.80 | 6 | 14 | 49 |
| | Precuneous | 174 | <.001 | 6.19 | 6 | -70 | 49 |
| | | | .001 | 5.27 | 9 | -73 | 64 |
| | Left OFC | 26 | <.001 | 6.16 | -45 | 56 | 1 |
| | | | .002 | 5.19 | -48 | 50 | -5 |
| | | | .023 | 4.65 | -36 | 62 | -2 |
| | Left superior frontal gyrus | 125 | <.001 | 6.04 | -24 | 5 | 64 |
| | Lateral Occipital Cortex | 193 | <.001 | 6.01 | 42 | -76 | 46 |
| | | | <.001 | 5.72 | 27 | -82 | 31 |
| | | | <.001 | 5.54 | 39 | -85 | 34 |
| * | Right dorsolateral PFC | 90 | <.001 | 5.87 | 39 | 32 | 37 |
| | Lateral Occipital Cortex | 91 | <.001 | 5.83 | -42 | -82 | 40 |
| | | | <.001 | 5.50 | -33 | -67 | 64 |
| | | | <.001 | 5.49 | -51 | -70 | 46 |
| * | Left dorsolateral PFC | 88 | <.001 | 5.58 | -45 | 41 | 34 |
| | | | .001 | 5.32 | -48 | 32 | 37 |
| | | | .006 | 4.95 | -39 | 38 | 43 |
| | Left middle OFC | 5 | .003 | 5.10 | -18 | 56 | -17 |
| * | Right Caudate | 12 | .004 | 5.07 | 12 | 20 | 4 |
| | Left Supermarginal gyrus | 9 | .004 | 5.07 | -57 | -46 | 55 |
| | Dorsal ACC | 5 | .008 | 4.89 | 6 | 35 | 31 |
| | Right Middle temporal gyrus | 3 | .015 | 4.74 | 63 | -22 | -17 |
| | Left OFC | 1 | .022 | 4.66 | -42 | 53 | -11 |

* Cluster used as region of interest in subsequent analyses

**Table 4.** Brain-behavior associations. Least square regressions with heteroskedasticity corrected standard error estimations with brain activation in the regions of interest predicting behavioral aggression.

| | | Conjunction | | | Negative>Positive | | | Positive>Neutral | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Noise blast difference | | ACC gyrus | left insula | right insula | medial PFC | left IFG | right IFG | SMA | right caudate | left DLPFC | right DLPFC |
| Negative - Positive | *r* | .07 | .08 | .07 | .02 | -0.01 | .05 | .11 | -.04 | .10 | .13 |
| | *p* | .152 | .105 | .169 | .674 | .845 | .401 | .027 | .460 | .074 | .017 |
| Negative - Neutral | *r* | .06 | .09 | .04 | .02 | .02 | .05 | .09 | -.00 | .13 | .13 |
| | *p* | .256 | .081 | .441 | .711 | .675 | .349 | .087 | .936 | .009 | .013 |

ACC: Anterior Cingulate Cortex; DLPFC: dorsolateral prefrontal cortex IFG: inferior frontal gyrus; PFC: prefrontal cortex; SMA: supplementary motor area

**Table 5.** Region Of Interest twin analyses. Pearson's correlations and ACE models for brain activation in the regions of interest (ACC: Anterior Cingulate Cortex; AI: Anterior Insula; IFG: inferior frontal gyrus; SMA: supplementary motor area; DLPFC: dorsolateral prefrontal cortex).

| ROI | | MZ | DZ | | $A^2$ | $C^2$ | $E^2$ |
|---|---|---|---|---|---|---|---|
| ***Conjunction Negative>Neutral and Positive>Neutral*** | | | | | | | |
| ACC gyrus | *r* | -0.04 | 0.14 | *ACE* | 0 | 0.06 | 0.94 |
| | *p* | 0.739 | 0.249 | *95% CI* | 0.00 - 0.20 | 0.00 - 0.21 | 0.80-1.00 |
| Left AI | *r* | -0.07 | -0.14 | *ACE* | 0 | 0 | 1 |
| | *p* | 0.493 | 0.252 | *95% CI* | 0.00 - 0.11 | 0.00 - 0.09 | 0.89 - 1.00 |
| Right AI | *r* | 0.06 | -0.11 | *ACE* | 0 | 0 | 1 |
| | *p* | 0.611 | 0.377 | *95% CI* | 0.00 - 0.19 | 0.00 - 0.12 | 0.81 - 1.00 |
| ***Negative > Positive*** | | | | | | | |
| Medial PFC | *r* | 0.12 | -0.2 | *ACE* | 0.01 | 0 | 0.99 |
| | *p* | 0.274 | 0.091 | *95% CI* | 0.00 - 0.21 | 0.00 - 0.12 | 0.79 - 1.00 |
| Left IFG | *r* | 0 | -0.06 | *ACE* | 0 | 0 | 1 |
| | *p* | 0.987 | 0.607 | *95% CI* | 0.00 - 0.19 | 0.00 - 0.13 | 0.81- 1.00 |
| Right IFG | *r* | 0.02 | 0.06 | *ACE* | 0 | 0.04 | 0.96 |
| | *p* | 0.853 | 0.628 | *95% CI* | 0.00 - 0.22 | 0.00 - 0.19 | 0.81 - 1.00 |
| ***Positive > Negative*** | | | | | | | |
| SMA | *r* | 0.23 | -0.21 | *ACE* | 0.1 | 0 | 0.9 |
| | *p* | 0.031 | 0.087 | *95% CI* | 0.00 - 0.31 | 0.00 - 0.14 | 0.69 - 1.00 |
| Right caudate | *r* | 0.12 | 0.02 | *ACE* | 0.1 | 0 | 0.9 |
| | *p* | 0.289 | 0.855 | *95% CI* | 0.00 - 0.29 | 0.00 - 0.22 | 0.71 - 1.00 |
| Left DLPFC | *r* | 0.18 | -0.05 | *ACE* | 0.13 | 0 | 0.87 |
| | *p* | 0.09 | 0.652 | *95% CI* | 0.00 - 0.32 | 0.00 - 0.20 | 0.68 - 1.00 |
| Right DLPFC | *r* | 0.27 | -0.22 | *ACE* | 0.14 | 0 | 0.86 |
| | *p* | 0.01 | 0.06 | *95% CI* | 0.00-0.34 | 0.00 - 0.14 | 0.66-1.00 |

# Discussion

This study aimed to investigate genetic and shared environmental influences on neural activity and aggression following social feedback in children. Consistent with prior studies, negative social feedback resulted in behavioral aggression (Achterberg *et al.*, 2016b; Achterberg *et al.*, 2017). Behavioral genetic modeling revealed that aggression following negative feedback (negative-positive and negative-neutral) was influenced by genetic as well as shared and unique environmental influences. Genetic influences ranged from 10-20%, whereas approximately 7% of the variance was explained by shared environmental influences. Although previous studies have also found influences of shared environment, with similar (Ferguson, 2010) or higher estimates (Rhee and Waldman, 2002; Porsch *et al.*, 2016), most studies have suggested stronger genetic influences (around 50%) on behavioral aggression (Rhee and Waldman, 2002; Ferguson, 2010). These differences can be partly attributed to the way the aggression was assessed. Indeed, a review of Tuvblad and Baker (2011) showed that twin correlations of aggression based on parent/ teacher reports were twice as high as twin correlations of observed aggressive behavior. Using single raters for multiple children might result in inflated genetic influences (Tuvblad and Baker, 2011), and an experimental design can overcome such rater bias. This study is the first to use an experimental task to test genetic influences on reactive social aggression in a developmental twin-sample. It shows that environmental factors are important predictors of reactive aggressive behaviors. In line with our results, longitudinal stability in reactive aggression has been shown to be influenced by environmental effects (Tuvblad *et al.*, 2009).

Our analyses of neural responses to negative, positive, and neutral social feedback showed that brain activation in the ACCg and anterior insula was related to general valiance/ social saliency. The ACCg has been suggested to be sensitive to determining others' motivation (Apps *et al.*, 2016), which is important in the processing of social feedback, irrespective of whether it is positive or negative. Moreover, the ACCg has been shown to have strong structural and functional connectivity with the anterior insula (Apps *et al.*, 2016), and together these regions have been indicated as the salience network (Damoiseaux *et al.*, 2006; van Duijvenvoorde *et al.*, 2016a). Our results show that activation of regions coding social saliency is present already in childhood, indicating this might be a core social motivational mechanism in humans. Previous social evaluation studies did not report heightened activation that was specific for negative social feedback (Gunther Moor *et al.*, 2010b; Guyer *et al.*, 2012; Achterberg *et al.*, 2016b; Achterberg *et al.*, 2017), which might be due to the smaller samples in previous studies ($n$=30-60) as compared to the current study (N=385). In the current study, medial PFC and IFG were activated during negative feedback. Interestingly, the ACCg is connected to the portions of the mPFC that signal other-oriented

information (Apps *et al.*, 2016; Lee and Seo, 2016), and both play important roles in social cognition and behavior (Blakemore, 2008). Our results suggest that whereas the ACCg signals for social salient cues, the mPFC might signal for social threatening cues. Positive feedback, on the other hand, resulted in heightened activation in the caudate, SMA and bilateral DLPFC, which is consistent with previous social evaluation paradigms that reported increased activation in striatum (Davey *et al.*, 2010; Gunther Moor *et al.*, 2010b; Guyer *et al.*, 2012), superior frontal gyrus/SMA (Gunther Moor *et al.*, 2010b; Guyer *et al.*, 2012), and middle frontal gyrus /DLPFC (Gunther Moor *et al.*, 2010b).

Interestingly, SMA and DLPFC activity were also associated with aggressive behavior on the task. SMA and DLPFC activations were related to aggression after negative (relative to neutral and/or positive) feedback. Post hoc visualization of PE values showed that children who were *more* aggressive after negative feedback showed relatively *less* activation of the DLPFC during negative feedback compared to positive social feedback. This is in line with prior studies in adults which showed that *more* DLPFC activity after negative social feedback was related to *less* subsequent aggression (Riva *et al.*, 2015; Achterberg *et al.*, 2016b). It should be noted, however, that we did not observe brain-behavior associations when we performed whole brain regression analyses, in contrast to earlier studies in adults (Achterberg *et al.*, 2016b). Moreover, our brain-behavior associations on ROIs did not survive Bonferroni correction. The DLPFC is one of the brain regions that take longest to mature (Sowell *et al.*, 2001; Gogtay *et al.*, 2004), leaving ample room for individual, developmental differences. Although our sample size was fairly large compared to previous fMRI studies, individual developmental differences are best captured with longitudinal designs, due to individual variation in the timing of brain maturation.

We did not find significant brain-behavior associations in other ROIs (caudate, IFG, insula, mPFC, ACCg) that responded to social peer feedback. The lack of brain-behavior associations might indicate that these regions signal for social cues, but are not sensitive to retaliation behaviors. Indeed previous studies have indicated the IFG, insula, mPFC and ACCg as important regions of the "social brain" (for reviews, see Blakemore (2008) and Adolphs (2009)). The social brain is defined as a network of brain regions that is activated when we evaluate others and think about others' intentions and feelings (Brothers, 1990; Blakemore, 2008). Activation in these regions during peer feedback evaluation could indicate that children evaluate the intentions of the peers, but might not be specifically related to the actions they intent towards that peer. Regions that did show a relation with aggression, namely the SMA and DLPFC, have indeed been shown to be associated with behavioral motor planning (SMA) and behavioral control (DLPFC) in previous research (Riva *et al.*, 2015; Achterberg *et al.*, 2016b).

Genetic modeling showed that genetics played a role in activation in the DLPFC, the SMA and the right caudate, with 10-14% of the variance explained by genetics. Previous heritability studies on structural brain measures have focused

on rather large anatomical regions (i.e., the whole frontal cortex) and also report genetic influences (Jansen *et al.*, 2015). One developmental study that specifically investigated heritability of the DLPFC showed heritability estimates of around 40% for cortical thickness (age range 5-19, Lenroot *et al.* (2009)). Only a handful of studies have addressed heritability in task-based fMRI (for an overview, see Jansen *et al.* (2015)). Blokland and colleagues (2011) investigated brain activation during a working memory task in young adults (aged 20-30) and showed heritability of brain function in (amongst others) DLPFC, ranging from 20-65%. To our best knowledge, our study is the first to investigate the heritability of task-based fMRI in middle childhood, so direct comparisons to previous studies cannot be made. However, test-retest reliability studies on task-based fMRI in developmental samples have shown higher interclass correlation coefficients (ICCs) for lateral PFC regions than for subcortical regions (van den Bulk *et al.*, 2013; Peters *et al.*, 2016), indicating that the DLPFC might indeed reflect trait-like genetic influences. An important next step would be to reveal *which* environmental and genetic factors play a role in explaining the variance in brain activation and aggression following social evaluation, and test whether specific environmental influences (e.g. supportive parenting) might moderate the influence of specific genetic factors (for example, see the study protocol of Euser *et al.* (2016).

Several limitations of the current study may be addressed in future research. First, the cover story of the SNAT task explicitly stated that the peers would not hear the noise blast. This decision was based on previous studies using a similar design (Konijn *et al.*, 2007). Therefore the aggression measure reflects imagined aggression. Future studies may separate real aggression from imagined aggression to test any neural differences between these two types of aggression. Second, although our sample size can be considered large with regards to fMRI, it is rather small for behavioral genetic modeling. The statistical power of genetic studies is influenced by, amongst others, the sample size and the ratio MZ:DZ (Visscher, 2004; Verhulst, 2017). Our genetic analyses of neural responses resulted in high estimates for the E component (and specifically E- models, see supplementary materials), reflecting influences from the unique environment and measurement error. However, our sample size may have been insufficient to detect significant contributions of A (genetics) and C (shared environment). Fortunately, our sample did have an approximately equal numbers of MZ and DZ twins, which is considered optimal (Visscher, 2004). Moreover, prior studies have showed that the E component was also the primary determinant of variance in structural brain measures (Lenroot *et al.*, 2009), highlighting the urgent need to disentangle unique environmental influences from measurement error. Last, we used several ROIs to investigate brain-behavior associations and twin correlations. Significant results did not survive Bonferroni correction for multiple

testing, and therefore need to be interpreted with caution. Nevertheless, our results provide important hypotheses which can be further examined in future (meta-) analyses.

# Conclusion

Taken together, our results suggest that the processing of social feedback is partly explained by genetic factors, and the level of behavioral aggression following these evaluations are related to genetics and shared environmental influences. The regulatory role of DLPFC in aggression regulation fits with prior research in adults (Riva *et al.*, 2015; Chester and DeWall, 2016) and may be sensitive to developmental changes (Somerville *et al.*, 2010; Casey, 2015). Our findings underscore that the way children react to positive and negative social feedback is influenced by environmental factors. This stresses the important role of environmental inputs on observed behavior, such as parents and teachers, and point to an important role for parenting programs and interventions.

# Acknowledgments

The Leiden Consortium on Individual Development is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). The authors declare no conflict of interests.

# Supplementary Materials

**Genetic modeling - comparison of parsimonious models**

Similarities among twin pairs are divided into similarities due to shared genetic factors (A) and shared environmental factors (C), while dissimilarities are ascribed to unique environmental influences and measurement error (E). Behavioral genetic modeling with the OpenMX package (Neale *et al.*, 2016) in R (R Core Team, 2015) provides estimates of these A, C, and E components. To investigate whether the more parsimonious AE model (with C fixed to zero), CE model (with A fixed to zero) or E model (with both A and C fixed to zero) showed a better fit to the data, we subtracted the log-likelihood of the AE and CE models from the log-likelihood of the ACE model and the fit of the E model from the fit of the AE or CE models to get an estimate of the Log-likelihood Ratio Test (LRT). In most circumstances LRT follows the $X^2$ distribution, with 3.84 as a critical value at $p$=.05, thus a LRT>3.84 indicates a significantly worse fit of the data. In addition, we used the Akaike Information Criterion (AIC; Akaike (1974)) a standardized model-fit metric, to compare the different models. Lower AIC values indicate a better model fit. When ACE models show the best fit, both heritability, shared and unique environment are important contributors to explain the variance in the outcome variable. AE models indicate that genetic and unique environmental factors play a role; whilst CE models indicate influences of the shared environment and unique environment. If the E model has no worse fit than AE or CE models, variance in the outcome variable is accounted for by unique environmental factors and measurement error.

**Table S1.** Twin analyses on noise blast difference scores. ACE models compared to parsimonious AE, CE and E models.

| Noise blast difference | model | A[2] | C[2] | E[2] | LTR | AIC |
|---|---|---|---|---|---|---|
| Negative - Positive | * ACE | 0.20 | 0.06 | 0.74 | | 7542.16 |
| | AE | 0.24 | - | 0.76 | 4.17 | 7544.33 |
| | CE | - | 0.14 | 0.86 | 38.67 | 7578.84 |
| | E | - | - | 1.00 | >22.18 | 7599.02 |
| Negative - Neutral | ACE | 0.10 | 0.08 | 0.82 | | 7173.47 |
| | AE | 0.09 | - | 0.91 | -0.33 | 7171.13 |
| | * CE | - | 0.20 | 0.80 | -.5.58 | 7165.88 |
| | E | - | - | 1.00 | >23.81 | 7192.95 |
| Neutral - Positive | ACE | 0.10 | 0.00 | 0.90 | | 6888.43 |
| | AE | 0.10 | - | 0.90 | <.001 | 6886.43 |
| | CE | - | 0.07 | 0.93 | 0.19 | 6886.63 |
| | * E | - | - | 1.00 | <1.39 | 6885.83 |

[1] LTR < 3.85 equals a significant better fit of the model ($p$<.05)

[2] Lower AIC values indicate a better model fit

* asterics indicate the best model fit

**Table S2.** Twin analyses on brain activation in the regions of interest (ACC: Anterior Cingulate Cortex; PFC: prefrontal cortex; IFG: inferior frontal gyrus; SMA: supplementary motor area; DLPFC: dorsolateral prefrontal cortex). ACE models compared to parsimonious AE, CE and E models

| ROI | model | A[2] | C[2] | E[2] | LTR[1] | AIC[2] |
|---|---|---|---|---|---|---|
| ***Conjunction Negative>Neutral and Positive>Neutral*** | | | | | | |
| ACC gyrus | ACE | 0.00 | 0.04 | 0.96 | | 944.02 |
| | AE | 0.02 | - | 0.98 | 0.38 | 942.41 |
| | CE | - | 0.04 | 0.96 | <0.001 | 942.02 |
| | * E | - | - | 1.00 | <0.50 | 940.53 |
| Left Insula | ACE | 0.00 | 0.00 | 1.00 | | 1130.48 |
| | AE | 0.00 | - | 1.00 | <0.001 | 1128.48 |
| | CE | - | 0.00 | 1.00 | <0.001 | 1128.48 |
| | * E | - | - | 1.00 | <0.001 | 1126.48 |
| Right Insula | ACE | 0.01 | 0.00 | 0.99 | | 1072.13 |
| | AE | 0.01 | - | 0.99 | <0.001 | 1070.13 |
| | CE | - | 0.00 | 1.00 | <0.001 | 1070.13 |
| | * E | - | - | 1.00 | <0.001 | 1068.13 |
| ***Negative > Positive*** | | | | | | |
| Medial PFC | ACE | 0.01 | 0.00 | 0.99 | | 950.65 |
| | AE | 0.01 | - | 0.99 | <0.001 | 948.65 |
| | CE | - | 0.00 | 1.00 | 0.01 | 948.66 |
| | * E | - | - | 1.00 | <0.01 | 946.66 |
| Left IFG | ACE | 0.00 | 0.00 | 1.00 | | 1141.15 |
| | AE | 0.00 | - | 1.00 | <0.001 | 1139.15 |
| | CE | - | 0.00 | 1.00 | <0.001 | 1139.15 |
| | * E | - | - | 1.00 | <0.001 | 1137.15 |
| Right IFG | ACE | 0.00 | 0.04 | 0.96 | | 1160.12 |
| | AE | 0.04 | - | 0.96 | 0.07 | 1158.19 |
| | CE | - | 0.04 | 0.96 | <0.001 | 1158.12 |
| | * E | - | - | 1.00 | <0.021 | 1156.32 |

[1] LTR < 3.85 equals a significant better fit of the model ($p$<.05)

[2] Lower AIC values indicate a better model fit

* asterics indicate the best model fit

**Table S2.** (continued)

| ROI | model | A² | C² | E² | LTR¹ | AIC² |
|---|---|---|---|---|---|---|
| ***Positive > Negative*** | | | | | | |
| SMA | ACE | 0.10 | 0.00 | 0.90 | | 1003.64 |
| | AE | 0.10 | - | 0.90 | <0.001 | 1001.64 |
| | CE | - | 0.00 | 1.00 | 0.87 | 1002.52 |
| | * E | - | - | 1.00 | <0,87 | 1000.52 |
| Right caudate | ACE | 0.10 | 0.00 | 0.90 | | 1308.21 |
| | AE | 0.10 | - | 0.90 | <0.001 | 1306.21 |
| | CE | - | 0.08 | 0.92 | 0.24 | 1306.45 |
| | * E | - | - | 1.00 | <1.48 | 1305.36 |
| Left DLPFC | ACE | 0.13 | 0.00 | 0.87 | | 1064.97 |
| | AE | 0.13 | - | 0.87 | <0.001 | 1062.97 |
| | CE | - | 0.07 | 0.93 | 0.96 | 1063.93 |
| | * E | - | - | 1.00 | <1,64 | 1062.61 |
| Right DLPFC | ACE | 0.14 | 0.00 | 0.86 | | 1108.45 |
| | AE | 0.14 | - | 0.86 | <0.001 | 1106.45 |
| | CE | - | 0.03 | 0.97 | 1.83 | 1108.29 |
| | * E | - | - | 1.00 | <1.97 | 1106.42 |

¹ LTR < 3.85 equals a significant better fit of the model ($p$<.05)

² Lower AIC values indicate a better model fit

* asterics indicate the best model fit

**Table S3.** MNI coordinates for local maxima activated for the whole brain contrasts without participants with pathology (N=377). ACC: Anterior Cingulate Cortex; IFG: Inferior Frontal Gyrus; SMA: Supplementary motor cortex; OFC: Orbitofrontal Cortex; PFC: Prefrontal Cortex

| Anatomical Region | Voxels | pFWE | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Conjunction Negative>Neutral and Positive>Neutral*** | | | | | | |
| Lateral Occipital Cortex | 3379 | <.001 | 13.74 | -45 | -82 | 1 |
| | | | 13.57 | -48 | -76 | -5 |
| | | | 12.52 | 48 | -70 | -5 |
| Occipital Cortex | 113 | <.001 | 6.81 | -24 | -64 | 61 |
| Right insula | 80 | <.001 | 6.31 | 39 | 23 | -11 |
| | | | 6.07 | 33 | 17 | -14 |
| Left insula | 28 | .001 | 5.15 | -33 | 26 | -5 |
| | | | 4.95 | -30 | 20 | -11 |
| Medial PFC | 5 | .013 | 5.03 | -6 | 53 | -2 |
| Right IFG | 7 | .009 | 4.93 | 51 | 23 | 13 |
| Rostral ACC | 31 | <.001 | 4.91 | 12 | 47 | 13 |
| | | | 4.85 | 3 | 56 | 19 |
| | | | 4.81 | 0 | 47 | 10 |
| Left insula | 2 | .024 | 4.67 | -45 | 14 | -5 |
| SMA | 1 | .032 | 4.61 | 6 | 5 | 67 |
| SMA | 1 | .032 | 4.57 | 6 | 11 | 64 |
| ACC | 1 | .032 | 4.52 | 0 | 47 | 1 |
| ***Negative > Positive*** | | | | | | |
| Occipital pole | 132 | <.001 | 16.55 | -9 | -97 | 13 |
| Occipital pole | 118 | <.001 | 8.39 | 27 | -91 | 13 |
| | | | 8.19 | 18 | -94 | 13 |
| Medial PFC | 138 | <.001 | 6.95 | -9 | 56 | 25 |
| | | | 5.46 | 9 | 62 | 25 |
| Left IFG | 57 | <.001 | 6.35 | -54 | 29 | 4 |
| | | | 5.24 | -45 | 26 | -8 |
| Right IFG | 16 | .003 | 5.15 | 51 | 32 | -2 |
| | | | 4.86 | 57 | 32 | 7 |
| Right Occipital Fusiform Gyrus | 3 | .021 | 4.83 | 18 | -85 | -5 |

**Table S3.** (continued)

| Anatomical Region | Voxels | pFWE | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Negative > Positive*** | | | | | | |
| Left Lateral Occipital Cortex | 9 | .008 | 4.72 | -48 | -82 | 1 |
| Left Central Opercular Cortex | 1 | .033 | 4.63 | -36 | -16 | 25 |
| ***Positive > Negative*** | | | | | | |
| Lingual gyrus | 844 | <.001 | 14.75 | 6 | -76 | -2 |
| | | | 13.96 | -18 | -85 | -8 |
| | | | 10.93 | 18 | -73 | -11 |
| Right superior frontal gyrus | 353 | <.001 | 7.27 | 24 | 5 | 55 |
| | | | 7.07 | -6 | 14 | 49 |
| | | | 6.41 | 9 | 11 | 52 |
| Right Lateral Occipital Cortex | 133 | <.001 | 6.90 | 30 | -82 | 31 |
| | | | 5.74 | 42 | -76 | 46 |
| | | | 5.62 | 39 | -73 | 55 |
| Precuneous | 151 | <.001 | 6.14 | 0 | -70 | 49 |
| | | | 5.20 | 9 | -73 | 64 |
| Left Superior Frontal Gyrus | 98 | <.001 | 6.05 | -24 | 2 | 58 |
| Right OFC | 32 | .001 | 6.03 | 42 | 59 | -8 |
| | | | 5.62 | 48 | 53 | -2 |
| | | | 4.89 | 36 | 56 | -14 |
| Left Lateral Occipital Cortex | 58 | <.001 | 5.69 | -36 | -85 | 40 |
| | | | 5.36 | -39 | -70 | 58 |
| | | | 5.23 | -51 | -67 | 49 |
| Left OFC | 15 | .004 | 5.68 | -45 | 56 | 4 |
| Right dorsolateral PFC | 47 | <.001 | 5.51 | 39 | 32 | 37 |
| | | | 4.89 | 39 | 32 | 46 |

**Table S3.** (continued)

| Anatomical Region | Voxels | pFWE | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Positive > Negative*** | | | | | | |
| Left dorsolateral PFC | 41 | <.001 | 5.43 | -45 | 41 | 34 |
| | | | 5.06 | -48 | 32 | 37 |
| | | | 4.82 | -36 | 47 | 40 |
| Right Caudate | 6 | .012 | 4.95 | 9 | 20 | 4 |
| Left middle OFC | 2 | .026 | 4.88 | -18 | 56 | -17 |
| Right Supermarginal gyrus | 13 | .005 | 4.82 | 60 | -43 | 49 |
| | | | 4.62 | 57 | -40 | 58 |
| Left Supermarginal gyrus | 2 | .026 | 4.73 | -48 | -58 | 58 |
| Dorsal ACC | 3 | .021 | 4.73 | 6 | 35 | 31 |
| Left OFC | 2 | .026 | 4.69 | -48 | 50 | -5 |
| Left Supermarginal gyrus | 1 | .033 | 4.54 | -57 | -46 | 55 |

**Table S4.** Brain-behavior associations without participants with pathology (N=377). Least square regressions with heteroskedasticity corrected standard error estimations with brain activation in the regions of interest predicting behavioral aggression.

| | | Conjunction | | | Negative>Positive | | | Positive>Neutral | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Noise blast difference | | ACC gyrus | left insula | right insula | medial PFC | left IFG | right IFG | SMA | right caudate | left DLPFC | right DLPFC |
| Negative - Positive | *r* | .08 | .08 | .07 | .01 | .03 | .05 | .11 | .04 | .09 | .13 |
| | *p* | .146 | .125 | .186 | .832 | .548 | .407 | .032 | .433 | .083 | .019 |
| Negative - Neutral | *r* | .06 | .09 | .03 | .01 | .00 | .05 | .09 | .00 | .13 | .13 |
| | *p* | .263 | .096 | .504 | .817 | .896 | .346 | .089 | .914 | .011 | .014 |

*ACC: Anterior Cingulate Cortex; DLPFC: dorsolateral Prefrontal Cortex IFG: Inferior Frontal Gyrus; OFC: Orbitofrontal Cortex; PFC: Prefrontal Cortex; SMA: Supplementary Motor Cortex*