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

Achterberg, M.

# CHAPTER FIVE

# Longitudinal changes in DLPFC activation within childhood are related to decreased aggression following social rejection

This chapter is based on: Achterberg M., Van Duijvenvoorde A.C.K., IJzendoorn, M.H., Bakermans M.J. & Crone E.A.M. Longitudinal changes in DLPFC activation within childhood are related to decreased aggression following social rejection (in revision, 2019)

# Abstract

Regulating aggression in the case of negative social feedback is an important prerequisite for developing and maintaining social relations. Prior studies in adults highlighted the role of the dorsolateral prefrontal cortex (DLPFC) as a regulating mechanism for behavioral control. Despite the fact that middle-to-late childhood is an important period for both brain maturation and social relations, no prior study examined development of aggression regulation following social feedback within childhood. The current study investigated this using a longitudinal fMRI study, with 456 same-sex twins undergoing two fMRI sessions, across the transition from middle childhood (7-9 years) to late childhood (9-11 years). Aggression regulation was studied using the Social Network Aggression Task: 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. Confirmatory analyses revealed that behavioral aggression after social evaluation decreased over time, whereas neural activation in anterior insula, medial PFC and DLPFC increased over time. Exploratory whole brain-behavior analyses in late childhood showed a negative association between aggression and bilateral DLPFC, with increased DLPFC activity resulting in decreased aggression. Change analyses further revealed that children who showed larger increases in DLPFC activity from middle to late childhood showed stronger decreases in aggression over time. These findings highlight the importance of the development of social emotion regulation mechanisms within childhood.

*Keywords: Social evaluation processing; Social emotion regulation; Dorsolateral prefrontal cortex; development; childhood;*

# Introduction

Regulating emotions in social interactions is one of the most important requirements for developing social relations in childhood. With increasing age, children become better at regulating their emotions (Silvers *et al.*, 2012), which has been suggested to be related to the development of cognitive and behavioral control functions between early childhood and adolescence (Diamond, 2013; Casey, 2015). Few studies have investigated the development of social emotion regulation within childhood, despite empirical findings showing that middle-to-late childhood marks the most rapid changes in cognitive control (Luna *et al.*, 2004; Zelazo and Carlson, 2012; Peters *et al.*, 2016). Although neuroimaging studies have shed light on the underlying neurobiological changes that sub serve childhood development in cognitive control, most studies have relied on cross-sectional comparisons which hinders the possibility to examine within-person change. The current study builds upon new insights in the neural processing of social emotion regulation by examining within childhood change in neural and behavioral social control in a longitudinal fMRI study.

Emotion regulation is of upmost importance when social interactions result in rejection. It is well documented that social rejection can lead to aggression and retaliation (Dodge *et al.*, 2003; Nesdale and Lambert, 2007; Chester *et al.*, 2014; Novin *et al.*, 2018). Social evaluation, including social acceptance and rejection, has previously been studied using ecologically valid social judgment paradigms, in which participants' profiles are evaluated by same-aged peers (Somerville *et al.*, 2006; Gunther Moor *et al.*, 2010b; Hughes and Beer, 2013; Silk *et al.*, 2014). Developmental neuroimaging studies including adolescent participants showed that receiving positive (acceptance) relative to negative (rejection) social feedback was associated with increased neural activity in the ventral medial prefrontal cortex (MPFC), the anterior insula (AI), and the anterior cingulate cortex (ACC) (Gunther Moor *et al.*, 2010a; Guyer *et al.*, 2016). The Social Network Aggression Task is an extended social evaluation paradigm that includes also a neutral feedback condition, and that provides participants with the opportunity to blast a loud noise towards the peer that evaluated them (Achterberg *et al.*, 2016b; Achterberg *et al.*, 2017; Achterberg *et al.*, 2018b). Consistent with prior studies (Dalgleish *et al.*, 2017), it was found that both adults and children showed stronger ACC and AI activity in this task after receiving both positive and negative feedback (relative to neutral feedback), indicating that these regions signal social salient cues (Achterberg *et al.*, 2018b). How neural responses to social evaluation feedback influence behavioral aggression in childhood, and how these neural regions change over time, remains currently unknown.

Controlling emotions elicited by social evaluation feedback relies on cognitive control, that is: individuals with better cognitive control functions show less subsequent aggression following rejection (Chester *et al.*, 2014). Moreover,

increased activation in the dACC and AI was related to less aggression in adults with high executive functioning, whereas adults with low executive functioning showed increased aggression with increasing neural activation (Chester *et al.*, 2014). Prior studies in adults further showed that the dorsolateral prefrontal cortex (DLPFC) might serve as a regulating mechanism for aggression after social evaluation, such that increased DLPFC activity after social rejection was related to less behavioral aggression (Riva *et al.*, 2015; Achterberg *et al.*, 2016b). Moreover, stronger functional connectivity between the lateral PFC and limbic regions was related to less retaliatory aggression (Chester and DeWall, 2016). Interestingly, prior theoretical perspectives have suggested that DLPFC maturation is an important underlying mechanism for developing a variety of control functions in childhood (Bunge and Zelazo, 2006; Diamond, 2013). Prior research revealed that in 7-8 year old children there were indications for associations between DLPFC and behavioral aggression(Achterberg *et al.*, 2018b), although these were less pronounced than in adults. Taken together, studies in adults showed a link between cognitive control and regulation of emotions after rejection in the ACC/insula (Chester *et al.*, 2014) and DLPFC (Achterberg *et al.*, 2016b), but no study to date examined longitudinal developmental changes in these brain regions in childhood in the context of social evaluation. These prior studies led us to hypothesize that within-person maturation of the ACC/AI and DLPFC may be associated with better aggression regulation in childhood.

The current study makes use of a unique developmental twin sample of the Leiden Consortium for Individual Development (L-CID; Euser *et al.* (2016)). The design is based on recent insights showing that home environment is an important factor that impacts children's behavioral control (Sektnan *et al.*, 2010; Vrijhof *et al.*, 2018). The L-CID study makes use of the video feedback intervention to promote positive parenting and sensitive discipline (VIPP-SD), an attachment based intervention that aims to enhance parental sensitivity and sensitive discipline (Juffer *et al.*, 2017a). The VIPP-SD has proven to diminish externalizing behavior problems such as aggression in younger age groups (0-6 years (Van Zeijl *et al.*, 2006; Juffer *et al.*, 2017b)). The L-CID study tests whether the VIPP-SD is also effective in parents with older children and possibly likewise beneficial for behavioral outcomes of older children. Therefore, this study design allows us to not only examine the development of aggression regulation within individuals over time, but also the effect of genetics and variations in the social environment.

Using this unique study design, we address the following research questions: i) How does aggression regulation following social evaluation changes longitudinally within childhood? And ii) to what extent are these changes dependent on heritability and changes in the social environment? In doing so, 492 same-sex twins (246 families) underwent two fMRI sessions across the transition from middle childhood (7-9 years) to late childhood (9-11 years). In between fMRI sessions, families received either the VIPP-Twins or a dummy

intervention (Euser *et al.*, 2016). Using linear mixed effects modeling, we first investigated how behavioral aggression after positive, negative and neutral social feedback changed over time, and whether variation in the environment influenced these changes. Next, we investigated changes in brain responses related to positive, negative and neutral social feedback longitudinally within childhood and examined brain-behavior associations. Based on previous studies, we selected the AI, the IFG, the MPFC, and DLPFC as regions of interest (Gunther Moor *et al.*, 2010b; Vijayakumar *et al.*, 2017; Achterberg *et al.*, 2018b). To test individual differences in aggression regulation we additionally performed exploratory whole brain-behavior MRI analyses to test for relations between prefrontal cortex activation and aggression regulation.

# Methods

## Participants



Participants in this study took part in the longitudinal twin study of the Leiden Consortium on Individual Development (L-CID (Euser *et al.*, 2016)). The procedures were approved by the Dutch Central Committee Human Research (CCMO) and written informed consent was obtained from both parents. 512 children (256 families) between the ages 7 and 9 were included at the first wave (previously described in Achterberg *et al.* (2018b), van der Meulen *et al.* (2018)), with a mean age of 7.94 ± 0.67 (49% boys, 55% monozygotic). The majority of the sample was Caucasian (91%) and right-handed (87%). 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). Intelligence (IQ) was estimated at W1 with the subtests '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). 456 children participated in a second lab two years later (for details regarding participant dropout see **Figure S1** and *supplementary materials*). **Table 1** provides an overview of demographic characteristics of the sample at wave 1 (W1) and wave 2 (W2). Participants underwent an MRI scan as part of the lab visits. At W1, 385 participants were included in the MRI analyses (mean age 7.99 ± 0.68, 47% boys, see also Achterberg *et al.* (2018b)). At W2 360 participants were included in the MRI analyses (mean age 10.01 ± 0.67, 48% boys). A total of 293 participants were included on the MRI analyses at both waves (mean age W1: 7.99 ± 0.66, 47% boys).

## Parenting intervention

Families were contacted 1.5 year after W1 to inform them on a parenting support program for parents of twins (VIPP-Twins (Euser *et al.*, 2016)). We then explained that we were unable to personally visit all families within the L-CID to offer the training. Therefore, families would be randomly assigned to either receiving the training in person, through six home visits (see Juffer and Bakermans-Kranenburg (2018)), or to alternatively discuss the development of your twin through six phone meetings (dummy intervention - control group). Detailed sample selection is described in the *supplementary materials*. The VIPP-Twins group consisted of n=164 children, of which 133 had sufficient quality MRI data (**Figure S1**). The control group consisted of n=244 children, of which n=186 had sufficient quality MRI data (**Figure S1**). Twenty-seven families (n=54 children) did not comply with random assignment to one of the conditions. These families received the (non-randomly assigned) dummy intervention in order to keep this group comparable to the control group for future analyses within the longitudinal L-CID study. Given that the participants in the non-randomly assigned control group could not be included in the analyses, these participants' MRI data were used as a reference group, and used to create task-relevant independent regions of interest (ROI) (see **section 2.4.4**).

## Social Network Aggression Task

The Social Network Aggression Task (SNAT) as described in Achterberg et al. (2016b; 2017; 2018b) was used to measure aggression after social feedback. 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. To keep task demands as similar as possible between the conditions, participants were instructed to always press the button. The longer they pressed the button the more intense the noise would be, which was visually represented by a volume bar. Participants received instructions on how to perform the SNAT and the children were exposed to the noise blast during a practice session. Thereafter, participants practiced six trials of the task. The time line of a SNAT trial was as follows: start screen (500 ms), social feedback (2500 ms), fixation screen (3000-5000 ms), noise screen (5000 ms), intra-trial interval fixation screen (0-11550 ms), see **Figure 1a**. The optimal jitter timing and order of events were calculated with Optseq 2 (Dale, 1999). The SNAT consisted of 60 trials, three runs of 20 trials for each feedback condition (positive, neutral, negative). Intra class coefficient (ICC) analyses (modeled with a two-way mixed model using the consistency definition) showed poor (ICC<0.40, (Cicchetti, 1994)) consistency in noise blast duration after positive (ICC=0.32 [95%CI= 0.24-0.41]), neutral (ICC=0.26 [ 95%CI=0.17-0.35]) and negative feedback (ICC=0.17 [95%CI=0.08 - 0.26)] between W1 and W2.

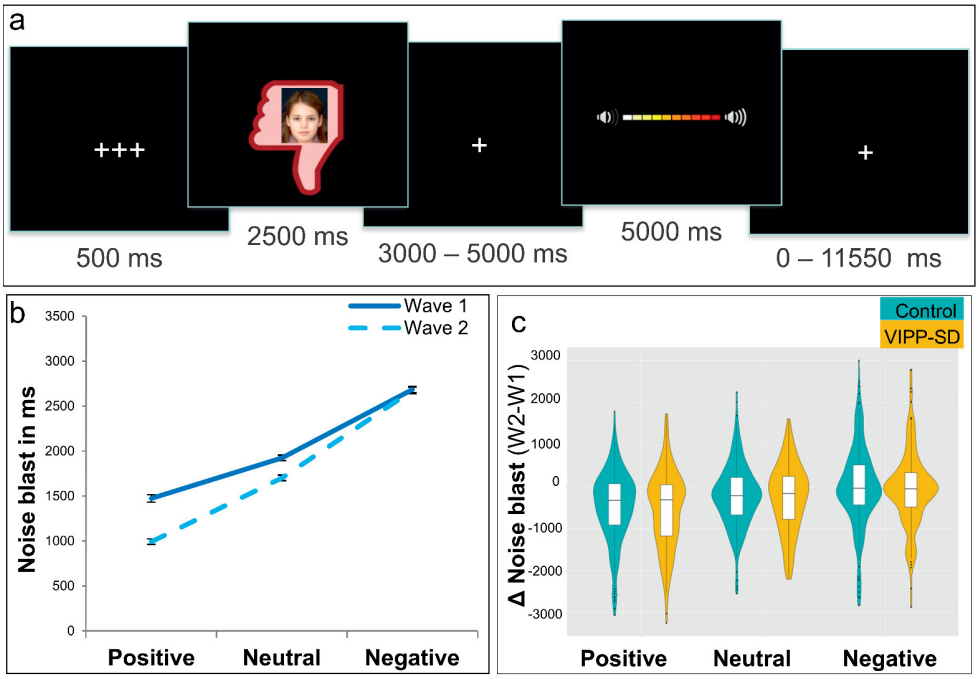


***Figure 1.*** *Social Network Aggression Task (SNAT). a) Visualization of one trial with negative social feedback. b) Noise blast duration is influenced by condition, wave, and condition × wave. c) Individual differences in change for noise blast after positive, neutral and negative social feedback did not differ between the VIPP-Twin and control group.*

## 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 (see also Achterberg and van der Meulen (2019)). 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).

**Preprocessing**
MRI data were analyzed with SPM8 (Wellcome Trust Centre for Neuroimaging, London). The exact same preprocessing steps were used in preprocessing MRI data from W1 and W2. Images were corrected for slice timing acquisition and rigid body motion. Functional scans were spatially normalized to T1 templates. Some participants did not finish the T1 scan and were normalized to an EPI template (W1: n=5 at W1; n=10 at W2). 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 out of three runs of fMRI data with <3 mm (1 voxel) motion in all directions were included in subject-specific analyses (W1: n=385; W2: n=358).

**Subject-specific analyses**
Statistical analyses were performed on individual subjects' data using a general linear model, previously described in Achterberg *et al.* (2018b). 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 was modeled as the second event, with the HRF modeled for the length of the noise blast 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. Additionally, six motion regressors (corresponding to the three translational and rotational directions) were included 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.

## Confirmatory ROI analyses

**ROI selection**
Regions of interest were based on higher-level group analyses of W2 in an independent reference group (the non-randomized dummy control group, n=41, **Table S1**). The advantage of this approach is that the participants were in exactly the same study protocol, but were not included in the subsequent analyses, leading to an independent selection of ROIs (Poldrack, 2007). Using comparable sample sizes, we previously reported replicable results of main effects of the social network aggression task (Achterberg *et al.*, 2017). We first investigated social feedback (positive, neutral, negative) versus fixation (see *supplementary materials*, **Figure S2a** and **Table S1**). SPM8's MarsBaR toolbox (Brett *et al.*, 2002)

was used to construct ROIs based on the whole brain contrast by masking significant activation with regions from the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer *et al.*, 2002). Based on a-priori hypotheses, we selected the bilateral anterior insula, inferior frontal gyrus (IFG), and medial prefrontal cortex (MPFC) from the *social feedback vs fixation* contrast, see **Figure 2**. In addition to the *social feedback vs fixation* contrast, we also investigated the specific conditions. From the contrast *positive vs negative social feedback* (see **Figure S2b** and **Table S1**), we selected the left dorsolateral prefrontal cortex (DLPFC) as additional ROI (**Figure 2**). The contrasts *negative vs positive social feedback* did not result in clusters of significant activation. The contrasts *positive vs neutral social feedback*; and *negative vs neutral social feedback* resulted in increased activation in occipital (visual) cortex (**Table S1**), but given that this was not an a priori hypothesized area, this region was not included in ROI selection.

Thus, in total, four ROIs were used in further analyses: the bilateral AI, bilateral IFG, MPFC, and the left DLPFC (see **Figure 2**). Parameter estimates (PE, average Beta values) were extracted from the subject-specific contrasts (*positive vs fixation*, *neutral vs fixation*, and *negative vs fixation*) for the entire sample minus the reference group with available MRI data on W1 (n=343) and W2 (n=317). ICC analyses (two-way mixed model using consistency) showed low consistency (ICC's<0.40, (Cicchetti, 1994)) in brain activation for the contrasts negative>neutral, negative>positive, and positive>neutral feedback between W1 and W2 (see **Table S2**).



**Linear mixed effects models**

To test time-related changes in participant's behavior (noise blast length) and ROI brain activation (parameter estimates) we used linear mixed effects models using the lme4 package (Bates *et al.*, 2015) in R (R Core Team, 2015). For these analyses we included the whole sample minus the reference group (n=458). Data was fitted on the average response times (for behavior) and average parameter estimates (for ROIs) after positive, neutral and negative social feedback. Two random effects were included to account for the nesting of condition and waves within participant (ChildID) and the nesting of twin-pairs within families (FamilyID). Fixed effects included feedback condition (3 levels: positive, neutral, and negative), wave (2 levels: wave 1 and wave 2), and intervention group (2 levels: VIPP-SD and control) and all 2-way and 3-way interactions. Participant's gender and estimated IQ (grand mean centered) were included as covariates and all main effects and two-way interactions between covariates and condition were included (gender $^{x}$ condition and condition $^{x}$ IQ). The fitted mixed-effect model was specified in R as:

$$Noise/ROI \sim condition * wave * intervention + condition * gender + condition * IQ + (1|childID) + (1|familyID).$$

In addition, we examined associations between brain and behavioral responses, in which we were specifically interest to what extent behavior was associated with neural activation. To this end, we added noise blast to the model including all 2 and 3-way interactions with condition and wave. Results were inspected with type III ANOVA's using Satterthwaite's method. Significant main effects of condition were further inspected using least-square means, with Kenward-Roger corrected degrees of freedom and Bonferroni adjusted p-values.

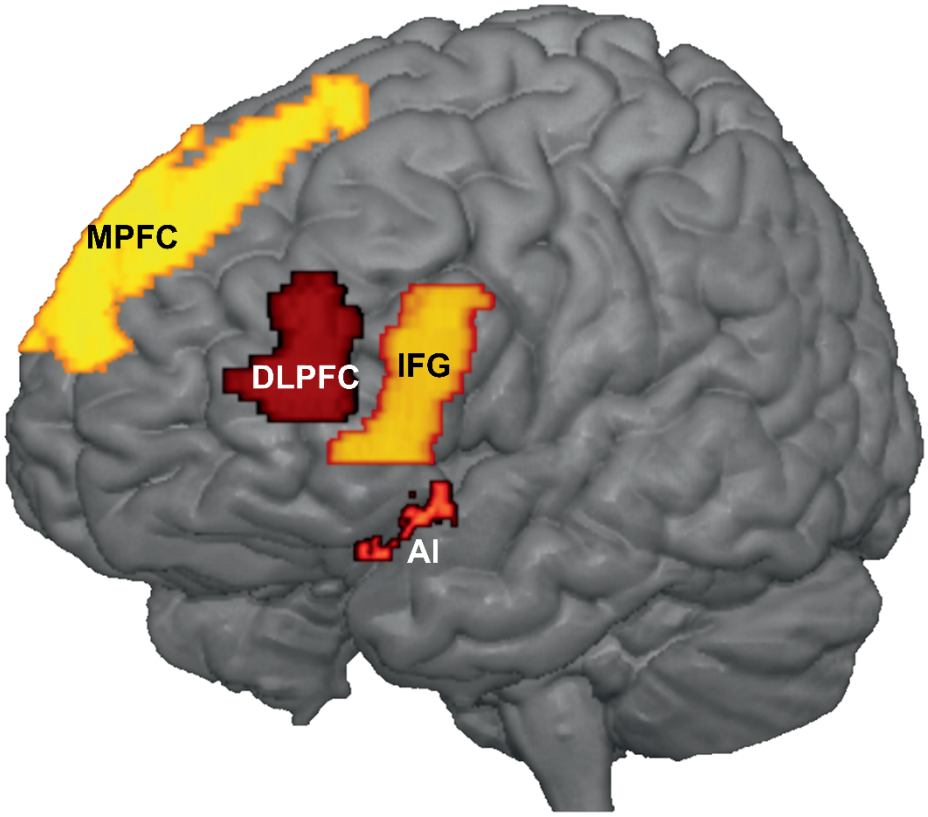


***Figure 2.*** *Regions of interest in the left hemisphere. mPFC= medial prefrontal cortex, dlPFC = dorsolateral prefrontal cortex, IFG = inferior frontal gyrus, AI= anterior insula. IFG and AI ROIs are bilateral.*

## Exploratory analyses

### Whole brain analyses at wave 2

In order to prevent that specific effects were overlooked due to a smaller sample size in the reference group, we performed exploratory whole brain analyses at

wave 2 including the VIPP-SD group, the control group and the reference group (n=360). Results were False Discovery Rate (FDR) cluster corrected ($p$FDRcc<0.05), with a primary voxel-wise threshold of $p$<.005 (uncorrected) (Woo *et al.*, 2014). We computed a full factorial ANOVA with three levels (positive, negative and neutral feedback) to investigate the neural response to the social feedback. Similarly to the whole brain analyses at wave 1 (reported in Achterberg *et al.* (2018b)), we first explored the general valence effects of social evaluation, by calculating a conjunction (using the "logical AND" strategy, see Nichols *et al.* (2005)) of *positive vs neutral* and *negative vs neutral* social feedback. Next, we calculated the contrasts *negative vs positive* and *positive vs negative* to investigate brain regions that were specifically activated for social rejection or social acceptance.

**Brain-behavior analyses**

In addition to neural responses to social feedback, we also examined whole brain-behavior relations in late childhood (wave 2). Similar to previous brain-behavior analyses in adults (Achterberg *et al.*, 2016b) we conducted a whole brain regression analysis at the moment of receiving negative social feedback (*negative vs neutral*), with the difference in noise blast duration after negative and neutral feedback as a regressor. In this way, we tested how initial neural responses to feedback were related to subsequent aggression. The difference in noise blast was computed by:

$$\Delta NegNeut\ W2 = \ Negative\ noise\ blast\ W2 - \ Neutral\ noise\ blast\ W2.$$

To investigate brain-behavior associations across time, we computed the difference over time in noise blasts duration for the contrast negative-neutral and for brain activation in this contrast. A total of 293 participants had behavioral and brain data available at two waves and were included in the analyses regarding brain-behavior associations over time. Difference scores over time for behavior and brain were computed as follows:

$$\begin{aligned}\Delta NegNeut\ behavior \\ &= (Noise\ blast\ negative\ W2\ - Noise\ blast\ neutral\ W2) \\ &- \ (Noise\ blast\ negative\ W1\ - Noise\ blast\ neutral\ W1)\end{aligned}$$

$$\begin{aligned}\Delta NegNeut\ brain\ &= (Neural\ activity\ negative\ W2\ - Neural\ activity\ neutral\ W2) \\ &- (Neural\ activity\ negative\ W1\ - Neural\ activity\ neutral\ W1)\end{aligned}$$

## Behavioral genetic analyses

To examine genetic and environmental influences on brain and behavior, we calculated Pearson within-twin correlations for mono- and dizygotic twin pairs. Similarities among twin pairs can be due to additive genetic variance (A) and

common (shared) environmental factors (C), while dissimilarities are ascribed to unique environmental influences and measurement error (E) (see **Figure S3**). We used behavioral genetic modeling with the OpenMX package (Neale *et al.*, 2016) in R (R Core Team, 2015) to calculate these A, C, and E estimates (see *supplementary materials*).

# Results

## Behavioral aggression following social evaluation

To test whether behavioral aggression decreased with increasing age, we performed a linear mixed-effect model on noise blast duration after feedback across two waves. The linear mixed effect model for noise blast duration showed the expected main effect of type of social feedback (**Table S3**). Noise blast duration was longer after negative feedback compared to neutral feedback, and shortest after positive feedback (all pairwise comparisons $p<.001$). We also found the expected main effect of wave (**Table S3**), with shorter noise blast durations at wave 2 compared to wave 1, indicating a decrease of behavioral aggression over time. Moreover, there was a significant condition × wave interaction effect (**Table S3**). As can be seen in **Figure 1b**, noise blast duration decreased more strongly between wave 1 and 2 after positive feedback than after negative feedback ($F=23.75$, $p<.001$) and more after positive feedback than after neutral feedback ($F=16.27$, $p<.001$). The same result was observed for neutral feedback: noise blast duration decreased more strongly between wave 1 and 2 after neutral feedback than after negative feedback ($F=5.00$, $p=.025$). That is, over time children showed a decrease in behavioral aggression, and this effect was most pronounced for aggression following positive feedback, see **Figure 1b**. We did not find any main or interaction effects of the parental intervention on behavioral aggression (**Table S3**) and visualization of the data showed large individual differences in aggression regulation in both groups (**Figure 1c**).

## Confirmatory ROI analyses

Confirmatory ROI analyses were performed in two steps: First, we examined neural responses patterns after social feedback across two time points. Second, we examined relations between changes in neural activity and noise blast durations.

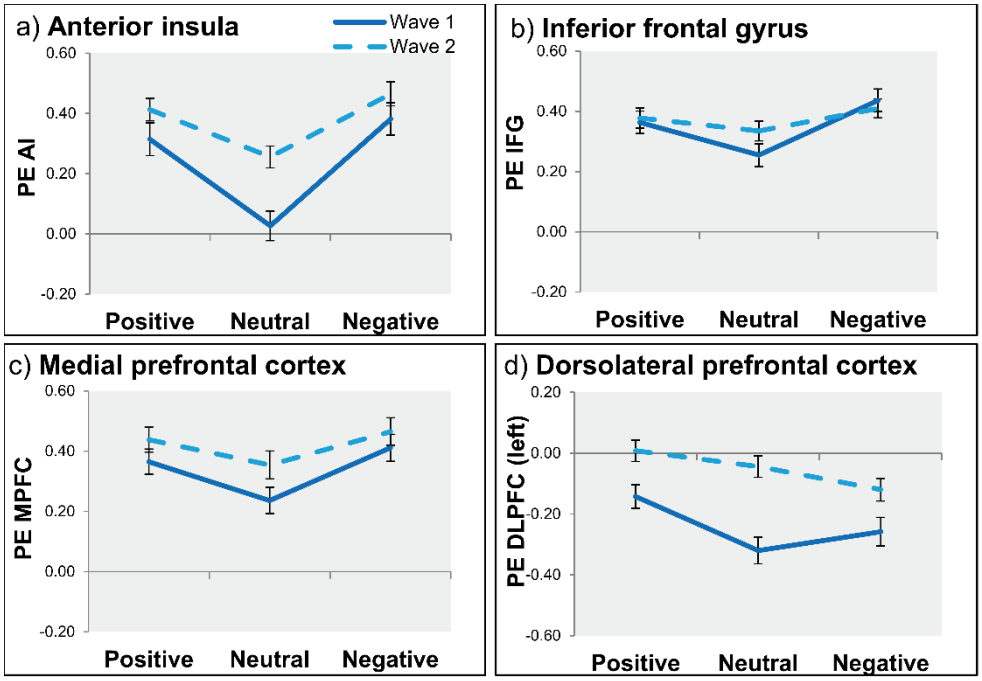


***Figure 3.*** *Neural activation after positive, neutral and negative social feedback at wave 1 (solid lines) and wave 2 (dotted lines) for the anterior insula (a), the inferior frontal gyrus (b), the medial prefrontal cortex (c) and the dorsolateral prefrontal cortex (DLPFC). PE = parameter estimates.*

**Neural responses following social evaluation**

To test for developmental changes in neural responses to social feedback, we performed linear mixed effect models on four ROIs (AI, IFG, MPFC and dlPFC). As expected, we observed significant main effects of type of social feedback on neural activation in all ROIs (**Table S4**). Patterns of activity differed between the ROIs. For the AI, IFG and MPFC there was significantly more neural activation after negative and positive feedback, relative to neutral feedback (**Figure 3a**, **3b** and **3c**), but the differences between positive and negative social feedback were not significant. For the DLPFC, in contrast, there was more activation after positive social feedback compared to both neutral and negative feedback, but no significant difference between neutral and negative social feedback, see **Figure 3d**. Next, we addressed whether these activity patterns changed over time, by testing for main effects and interactions with wave. We observed a significant effect of wave in the AI, the MPFC and the DLPFC, with generally stronger neural activation at wave 2 compared to wave 1 (**Figure 3a**, **3c** , **3d** and **Table S4**). There were no main or interaction effects of the parental intervention (**Table S4**).

**Brain-Behavior associations**

To investigate brain-behavior associations we added noise blast duration as a factor to the previously tested models. We found a significant main effect of noise blast duration on AI and DLPFC activation (**Table S5**). These findings indicated that increased AI activation was associated with longer noise blast (B=1.11e-04), whereas increased DLPFC activation was associated with shorter noise blast (B= -3.57e-05). The IFG and MPFC did not show significant brain-behavior associations. The condition × noise blast interaction effects on brain activation in the ROIs were not significant (see **Table S5**).

## Exploratory analyses

**Whole brain analyses on social evaluation processing**

To prevent that specific effects were overlooked by due to a relatively small sample size in the reference group, we performed exploratory whole brain analyses at wave 2 including the VIPP-SD group, the control group and the reference group (n=360). Results from the whole brain contrasts for wave 2 (children ages 9-11-years see **Figure S3, Table S6**) resulted in similar patterns of neural activation as was previously observed at wave 1 (children aged 7-9 years, Achterberg et al., 2018,) and in a different sample of adults (Achterberg et al., 2016). These results are described in more detail in the *supplement materials*.

**Brain-behavior analyses on aggression following negative feedback**

We conducted a whole brain regression analysis at wave 2 for receiving negative feedback (contrast Negative vs Neutral), with the difference in noise blast duration after negative and neutral feedback as a regressor (ΔNegNeut W2, see section 2.6.2.). Consistent with our hypothesis, we observed a negative association between behavioral aggression and activation in the bilateral DLPFC (**Figure 4a, Table 2**). Visualization of the effect (**Figure 4b**) showed that an increase in DLPFC activation after negative feedback (relative to neutral feedback) resulted in less subsequent behavioral aggression.

To test whether children who showed larger increases in DLPFC activity over time also showed less behavioral aggression over time, we included the data points at wave 1 to the analysis. Note that for this analysis we only included participants who had behavioral and brain data available at two waves (n=293). For these participants, we calculated the relation between the change in DLPFC activation (ΔNegNeut brain, see section 2.6.3.) in whole-brain DLPFC ROI (**Figure 4a**) and the change in noise blast duration (ΔNegNeut behavior, see section 2.6.3). We found a significant negative association ($r$=-.16, $p$=.005), indicating that children who showed the largest increase in DLPFC activation across childhood also showed the largest decrease in behavioral aggression across childhood (**Figure 4c**).

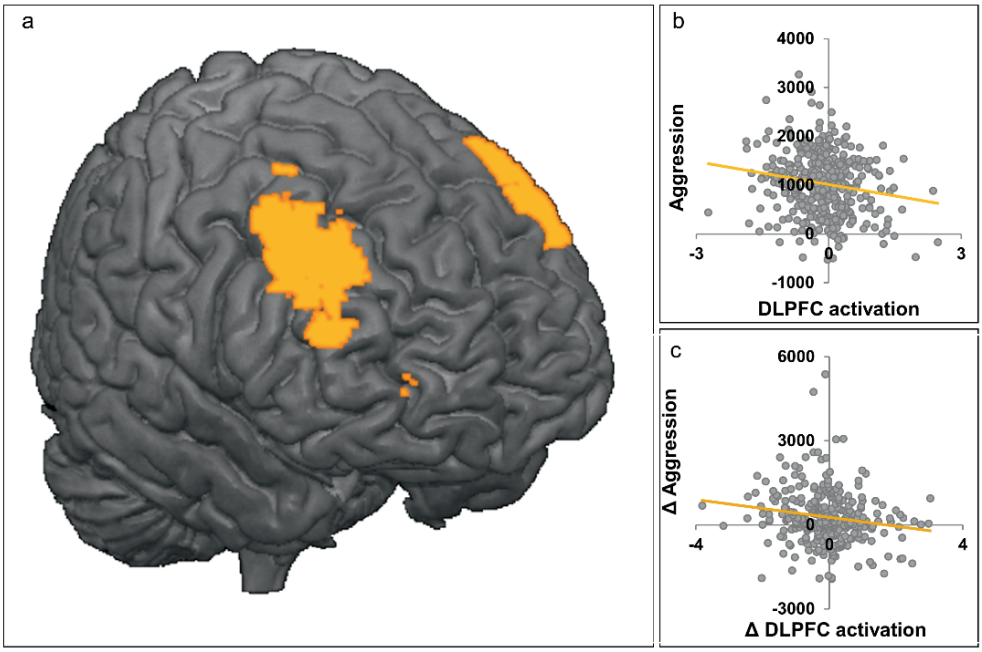


***Figure 4.*** *Whole brain-behavior analyses with all available MRI data at wave 2 (N=360). a) Significant cluster of activation in bilateral DLPFC for negative> neutral social feedback with noise blast (Δ negative-neutral) as regressor. b) Visualization of brain-behavior association at wave 2: increased DLPFC activity after negative feedback is related to decreased aggression. c) Brain-behavior association over time: the change in DLPFC activation is negatively correlated to the change in aggression, with larger increases in DLPFC activity over time being related to larger decreases in aggression.*

## Genetic and environmental influences

Given that our sample consists of both mono- and dizygotic twins, we were able to test for effects of genetics, shared environment and unique environments. As can been seen in **Table 3**, behavioral aggression was driven by a combination of genetic, shared and unique environmental factors. Variation in neural activity in the salience ROIs (AI, IFG, MPFC) showed little to no genetic influence, but did show moderate effects of shared environmental effects. Most variation was explained by the unique, non-shared environment (including measurement error). For DLPFC activation, results were inconclusive. There were some indications of heritability (i.e., on individual differences in positive-neutral), whereas individual differences were partly explained by shared environment (negative-neutral). Again, most individual differences were explained by unique non-shared environment (including measurement error).

# Discussion

There is a great need to have a better understanding of the mechanisms that drive changes in emotion regulation during social interactions across childhood. The current study tested the neural signature of aggression regulation in childhood in the context of social evaluations, specifically social acceptance and social rejection. For this purpose, we made use of the unique longitudinal L-CID cohort, which allowed us to examine the development of aggression regulation within individuals over time and take into account possible effects of genetics and environmental variations. By using longitudinal behavioral-neural comparisons, we were able to address the question how change in neural activity relates to change in behavioral development. The current study revealed three main findings: 1) behavioral aggression after social evaluation decreased over time, and this decrease was most pronounced for aggression after positive and neutral social feedback; 2) confirmatory ROI analyses showed that increased activity in AI was related to more aggression, whereas increased activity in DLPFC was correlated with less aggression; and 3) bilateral DLPFC was correlated to less subsequent aggression following negative social feedback. Longitudinal comparisons confirmed that a larger increase in DLPFC activity across childhood was related to a larger decrease in behavioral aggression after negative social feedback.

The behavioral results confirmed our initial hypothesis that behavioral aggression decreases over time, consistent with prior reports on age related increases in behavioral control (Diamond, 2013; Casey, 2015). Interestingly, however, these reductions in aggression were most pronounced following positive and neural feedback, suggesting that participants were more motivated to refrain from aggression towards liked others. These findings fit well with research showing that the importance of being liked and accepted by others increases over the course of childhood and into adolescence (Rodman *et al.*, 2017; Sherman *et al.*, 2018a). Thus, with increasing age, children become more focused on refraining punishment towards people with whom they socially connect and they differentiate more between liked (individuals signaling social acceptance) and disliked (individuals signaling social rejection) others (see also Guroglu *et al.* (2014)).

By using functional neuroimaging we were able to address the neural correlates following social evaluation feedback across two time points. Consistent with prior reports (Achterberg *et al.*, 2018b), children activate the same network across two waves, with stronger activity in ACC, AI and IFG after both positive and negative social feedback (relative to neutral feedback). These findings fit well with results from the adult literature, showing that neural activation in ACC, AI, and IFG, is associated with social rejection (Eisenberger *et al.*, 2003; Cacioppo *et al.*, 2013)) and signaling social salient events (Dalgleish *et al.*, 2017). The DLPFC, in contrast, was more active for positive than

negative/neutral feedback, comparable to the behavioral results showing a stronger reduction over time in aggression following positive feedback. Interestingly, AI and DLPFC also showed opposite relations to aggression. Even though both regions increased in activation over time, stronger AI activity was associated with more behavioral aggression and stronger DLFPC activity was associated with less behavioral aggression. The AI results are comparable to a previous finding in adults with low executive control functions, showing that for individuals with low executive control AI activity and aggression were positively correlated (Chester *et al.*, 2014). Even though we did not observe changes in AI activity over time, an interesting direction for future research will be to examine whether this relation is stronger in childhood than adolescence and adulthood, when executive control functions increase.

The positive relation between DLPFC activity and aggression regulation was confirmed in several analyses. First, bilateral DLPFC activity was the only neural predictor in a whole brain regression analysis for aggression control following negative relative to neural feedback. These findings fit well with two decades of research pinpointing the DLPFC as an important regions for cognitive control development (Luna *et al.*, 2004; Luna *et al.*, 2010; Crone and Steinbeis, 2017). The current study extends this finding to the novel domain of social interactions, and demonstrates that the same 'cold' regulatory control functions are also important for regulation 'hot' emotions in social evaluation contexts (Zelazo and Carlson, 2012; Welsh and Peterson, 2014). Moreover, DLPFC activity also explains individual differences in emotion regulation following rejection. A change-change analysis confirmed that those children who showed the largest increase in DLPFC activity after negative social feedback, also showed the largest reductions in behavioral aggression following negative feedback. This study was performed in a relatively small age range, from 7-9-year old to 9-11-year old, to provide a detailed analysis of changes in childhood. The results provide a window for understanding individual differences in these developmental trajectories, showing that some children develop stronger regulation skills already in childhood. Future research should examine these questions in a longer developmental time window (including more time points) using large samples, which allows disentangling general developmental patterns from individual differences in trajectories.

An intriguing question for future research is whether and how social influences impact individual differences in developmental trajectories. In this study, we addressed this question by examining the effects of a randomized control parenting intervention. Behavioral genetic analyses revealed mostly environmental influences on both behavior and brain (moderate effects of shared environment). Therefore, it was unexpected that we did not find effects of the parental intervention on brain and behavioral outcomes. Although previous studies using VIPP-SD in younger children reported transfer effects (i.e., less externalizing problems in children (Juffer *et al.*, 2017a)) the current study did not

reveal effects for the VIPP-Twins on behavioral emotion regulation or neural activity. One possible explanation is that participants were tested in a relatively short period after the parenting interventions was completed (approximately one month), and effects on the child may only be visible after a longer time period (Bakermans-Kranenburg *et al.*, 2008). Alternatively, during the transition from middle childhood to early adolescence, peers become more important (Berndt, 2004). An interesting future direction for interventions is therefore to target the peer-environment. One particularly ecological valid way to study the peer environment is to focus on social media use (Giglietto *et al.*, 2012). Despite the fact that social media are everywhere around us and used by almost everyone on a daily basis, little scientific research has been conducted on the effects of social media on the developing brain (Crone and Konijn, 2018). Social judgment paradigms as the SNAT mimic social rejection and acceptation by peers in a way that is comparable to social media environments where individuals connect based on first impression. Future research could take into account variations of the social environment by additionally monitoring real life social media use (for example using a smartphone app, see Montag *et al.* (2017)).

# Conclusion

This study set out to test longitudinal changes in neural systems underlying social evaluation and aggression regulation, and its relation to behavioral outcomes. We found an increase in behavioral control across childhood, as behavioral aggression decreased over time and DLPFC activation was related to decreased behavioral aggression. Notably, children that showed larger increases in DLPFC activity within childhood also displayed the largest longitudinal decrease in behavioral aggression. These results gain in our understanding on how the developing brain processes social feedback and suggest that the DLPFC might serve as emotion regulation mechanisms in terms of negative social feedback. However, it remains unknown how these results relate to actual, real-life social interactions such as social media use. Novel approaches are needed to bring together both real-life social media monitoring, as well as innovative experimental neuroimaging as this will provide cutting edge research and can provide insights through a neuro-mechanistic approach.



# Acknowledgements

We are grateful to Dr. Mara van der Meulen for her collaboration on the longitudinal MRI data collection. 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).

# Supplementary Materials

## Participants and sample selection

Of the initial 256 families, 10 families (3.8%) dropped out of the study directly after W1, whereas one family (n=2) was included in the L-CID study after W1. An additional 19 families (7.4%) dropped out before W2, after randomization of the parental intervention (see **Figure S1**). The remaining 456 children participated in a second lab visit at W2 (time between waves 2.06±0.10, time range: 1.86-2.53). Participants underwent an MRI scan as part of the lab visits. All 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. At W1, 27 participants did not start the scan due to anxiety (n=13), contraindications (n=6), or lack of parental consent for MRI participation (n=4), or technical issues with the MR system (n=4) (Achterberg and van der Meulen, 2019). Eighty-nine participants were excluded at W1 due to excessive head motion, which was defined as >3 mm motion (1 voxel) in any direction (x, y, z) in more than 2 runs of the SNAT task (3 runs in total). An additional seven participants were excluded due to data export failures. At W1, 385 participants were included in the MRI analyses (mean age 7.99 ± 0.68, 47% boys, see also Achterberg *et al.* (2018b)). At W2 48 participants did not start the scan due to anxiety (n=26), contraindications (n=10), or due to lack of parental consent for MRI participation (n=10). 46 participants were excluded at W2 due to excessive head motion and two participants were excluded due to data export failures. At W2 360 participants were included in the MRI analyses (mean age 10.01 ± 0.67, 48% boys).

Of the initial sample that participated at W1, 246 families were contacted 1.5 year after W1 to inform them on a parenting support program for parents of twins (VIPP-Twins (Euser *et al.*, 2016)). 91 families (37%) were assigned to the parental intervention group and received the VIPP-Twins, of which 9 families (9.9%) dropped out before the second MRI visit (final VIPP-Twins group: n=164, of which n=133 with sufficient quality MRI (**Figure S1**)). 129 families (52%) were assigned to the control group and received the dummy intervention, of which 7 families (5.5%) dropped out before the second MRI visit (final control group: n=244, of which n=186 with sufficient quality MRI (**Figure S1**)). Twenty-seven (11%) families did not want to be randomly assigned to one of the conditions. These families received the (non-randomly assigned) dummy intervention in order to keep this group comparable to the control group for future analyses within the longitudinal L-CID study. Given that the participants in the non-randomly assigned control group could not be included in the analyses, these participants were used as a reference group for regions of interest (ROI) selection (see section 2.4.4). Of the 27 families in the reference group, 3 dropped out before

W2. Of the remaining 48 children (**Figure S1**), 43 participated in the MRI session. Two participants were excluded due to excessive head motion. The final reference group therefore consisted of 41 participants, with a mean age of 10.13 ± 0.71 (age range: 9.09-11.28, 63% boys).

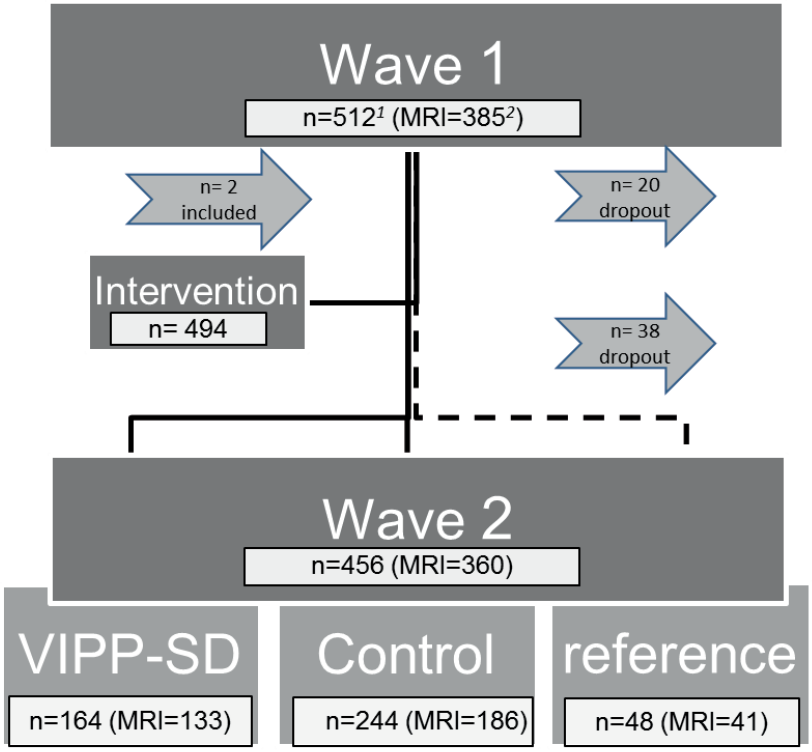


***Figure S1.*** *Participant flowchart. minus reference group: [1] n=458, [2] n=343*

## Whole brain analyses reference group

Regions of interest were based on higher level group analyses of W2 in an independent reference group (the non-randomized dummy control group, n=41, **Table S1**). A full-factorial ANOVA with three levels (positive, negative and neutral feedback) was used to investigate the neural response to the social feedback event in the reference group. Results were False Discovery Rate (FDR) cluster corrected ($p$FDRcc<0.05), with a primary voxel-wise threshold of $p$<.005 (uncorrected) (Woo *et al.*, 2014). We first investigated social feedback (positive, neutral, negative) versus fixation. This contrast resulted in activation in amongst others the fusiform gyrus, the inferior frontal gyrus, and the superior frontal gyrus (see **Figure S2a** and **Table S1**). In addition to the *social feedback vs fixation* contrast, we also investigated the specific conditions. The contrast *Positive vs*

*Negative feedback* resulted in activation in the right lingual gyrus, the left middle frontal gyrus, and the right inferior parietal lobule (see **Table S1**, **Figure S2**). The contrasts *positive vs neutral social feedback*; and *negative vs neutral social feedback* resulted in increased activation in occipital (visual) cortex (**Table S1**).

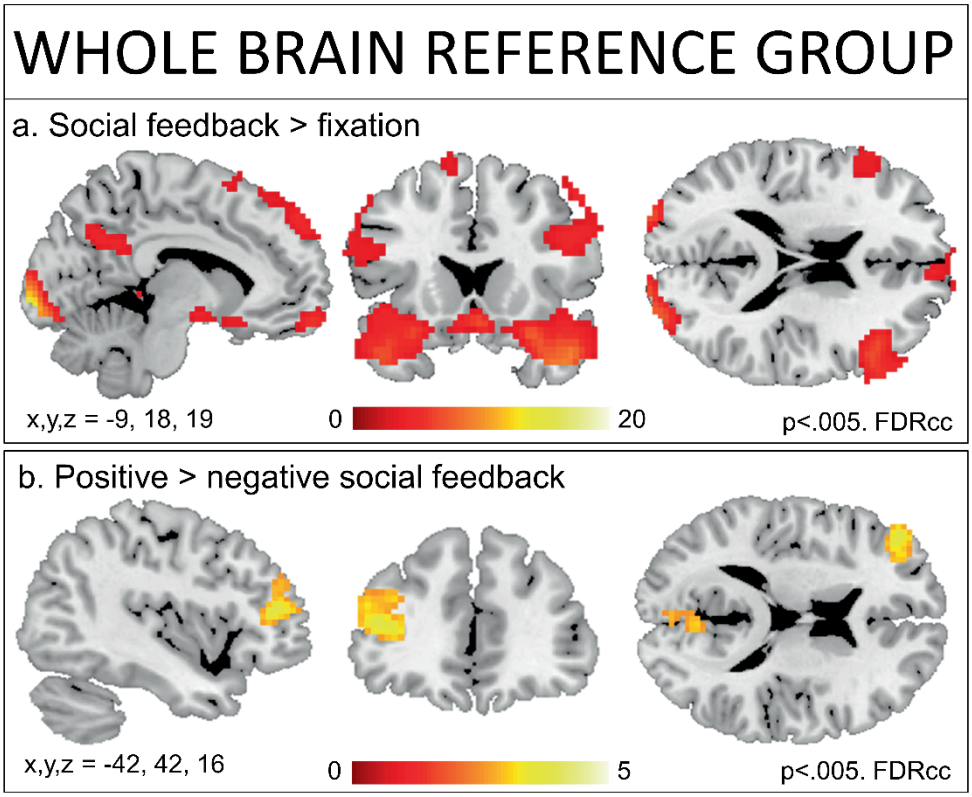


***Figure S2.*** *Whole brain analyses for reference group (n=41).*

## Behavioral genetic analyses

Similarities among twin pairs are divided into similarities due to additive genetic factors (A) and common (shared) environmental factors (C), while dissimilarities are ascribed to unique non-shared environmental influences and measurement error (E). Behavioral genetic modeling with the OpenMX package (Neale *et al.*, 2016) in R (R Core Team, 2015) was used to provide estimates of these A, C, and E components. The correlation of the shared environment (factor C) was set to 1 for both MZ and DZ twins, while the correlation of the genetic factor (A) was set to 1 for monozygotic twins and to 0.5 for dizygotic twins. The last factor, unique environmental influences and measurement error, was freely estimated (**Figure S4**). We calculated the ACE models for noise blast duration and brain activation

in the contrasts negative-neutral, negative-positive and positive-neutral. High estimates of A indicate that genetic factors play an important role, whilst C estimates indicate influences of the shared environment. If the E estimate is the highest, variance in motion is mostly accounted for by unique environmental factors and measurement error.

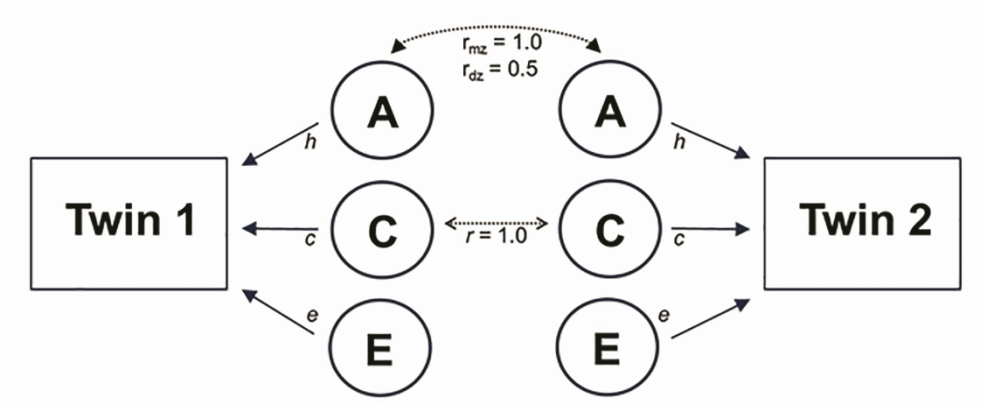


***Figure S3.*** *ACE model. The correlation between the additive genetic factor (A) of twin 1 and 2 is set to 1.0 for monozygotic (MZ) twins and to 0.5 for dizygotic (DZ) twins. The correlation between common, shared environmental factors (C) is set to 1.0 for both MZ and DZ twins. The correlation between the unique, non-shared environmental factors (including measurement error, (E)) is freely estimated within the model.*



## Exploratory whole brain analyses

To prevent that specific developmental effects were overlooked, we performed exploratory whole brain analyses at wave 2 including the VIPP-SD group, the control group and the reference group (n=360). We first investigated the general valence effects of social evaluation, that is to say, regions in the brain that were active after positive and negative feedback, relative to neutral social feedback. In doing so, we calculated a conjunction of *positive vs neutral* and *negative vs neutral* social feedback. We found common activation across positive and negative feedback in three clusters of activation: in the left AI; in the right AI extending into the right IFG; and in the occipital lobe, extending into the fusiform gyrus (**Figure S4a, Table S6**). To test for specific effects of positive versus negative social feedback, we examined pair-wise contrasts on social rejection and social acceptance. The contrast of social rejection (*negative vs positive* social feedback) resulted in significant activation in -amongst others- the right putamen/thalamus, the bilateral IFG, and the MPFC (**Figure S4b,Table S6**). The contrast of social acceptance (*positive vs negative* social feedback) resulted in two large clusters of significant activation, one cluster in the prefrontal cortex (including the superior frontal gyrus and the left and right DLPFC) and one cluster

with local maxima in the occipital lobe (including the right and left lingual gyrus extending to more parietal regions and the precuneus) see **Figure S4c** and **Table S6**.

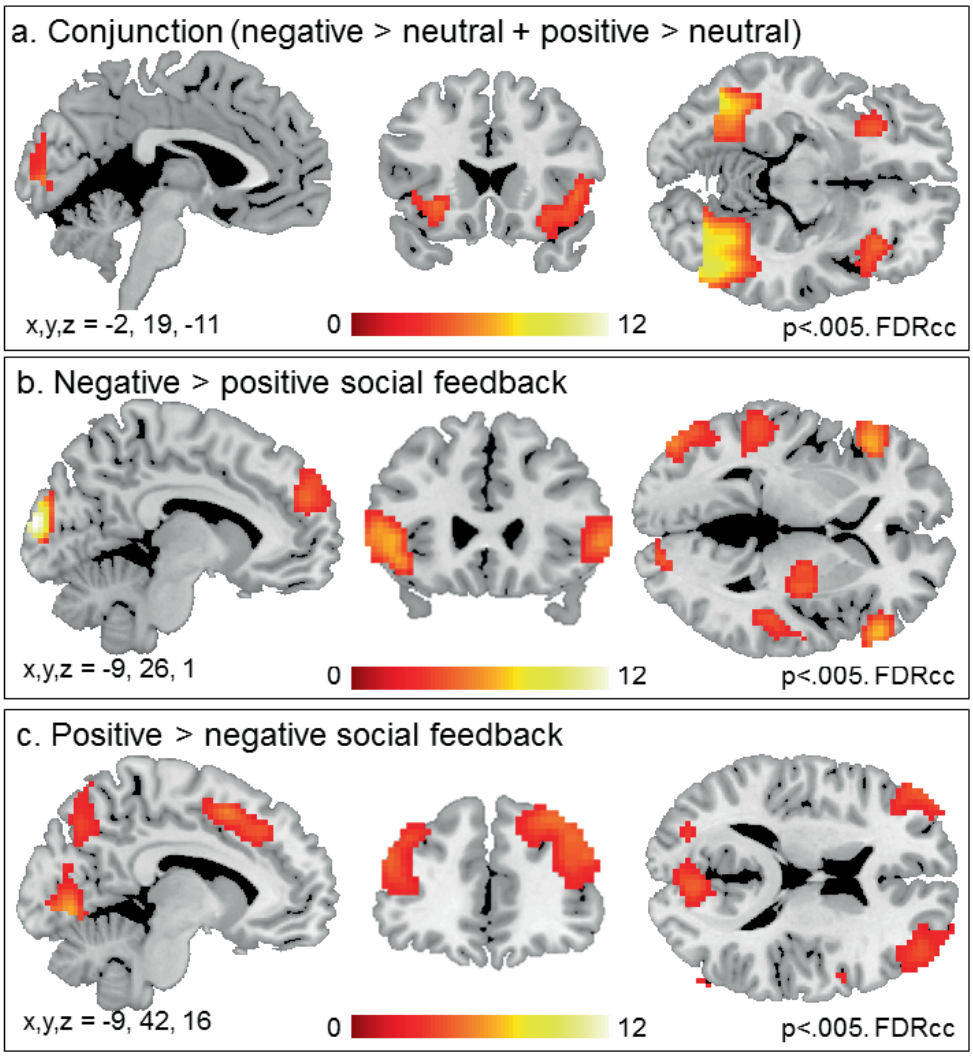


***Figure S4.*** *Whole brain analyses for all available MRI data at wave 2 (N=360). A) Neural activation for the general valence effects of social evaluation (Conjunction of negative>neutral and positive > neutral). B) Neural activation after social rejection. C) neural activation after social acceptance.*

**Table S1.** MNI coordinates for local maxima activated for the whole-brain contrasts in the reference group (N=41).

| Anatomical Region | Voxels | pFDRcc | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Social feedback > fixation*** | | | | | | |
| Right Fusiform Gyrus | 7710 | <.001 | 19.07 | 39 | -52 | -17 |
| | | | 18.87 | 39 | -79 | -11 |
| | | | 18.61 | 30 | -94 | 4 |
| Right Posterior Cingulate Cortex | 790 | <.001 | 6.53 | 3 | -55 | 31 |
| | | | 5.03 | 39 | -67 | 61 |
| | | | 4.87 | 36 | -61 | 43 |
| Right Inferior Frontal Gyrus | 542 | <.001 | 6.11 | 54 | 26 | 22 |
| | | | 6.08 | 60 | 29 | 28 |
| | | | 5.94 | 45 | 29 | 19 |
| left Rectal Gyrus | 158 | 0.009 | 5.95 | 0 | 65 | -17 |
| Right Superior Frontal Gyrus | 453 | <.001 | 4.87 | 15 | 50 | 49 |
| | | | 4.69 | -9 | 53 | 46 |
| | | | 4.28 | -12 | 38 | 55 |
| Left Angular Gyrus | 170 | 0.008 | 4.06 | -48 | -61 | 37 |
| | | | 3.57 | -57 | -55 | 52 |
| | | | 3.04 | -39 | -67 | 55 |
| ***Positive > negative social feedback*** | | | | | | |
| Right Lingual Gyrus | 908 | <.001 | 5.43 | 6 | -76 | -2 |
| | | | 5.25 | -18 | -85 | -8 |
| | | | 4.70 | 15 | -73 | -5 |
| Left Inferior/Middle Frontal Gyrus | 185 | 0.037 | 4.08 | -42 | 41 | 13 |
| | | | 4.06 | -36 | 47 | 13 |
| | | | 3.26 | -39 | 44 | 25 |
| Right Inferior Parietal Lobule | 170 | 0.037 | 3.89 | 57 | -34 | 55 |
| | | | 3.39 | 69 | -31 | 43 |
| | | | 3.30 | 63 | -16 | 28 |

**Table S1.** (continued)

| Anatomical Region | Voxels | pFDRcc | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Positive > neutral social feedback*** | | | | | | |
| Left Fusiform Gyrus | 3186 | <.001 | 6.43 | -27 | -79 | -11 |
| | | | 6.41 | 24 | -70 | -11 |
| | | | 5.98 | 12 | -76 | -8 |
| ***Negative > neutral social feedback*** | | | | | | |
| Left Middle Occipital Gyrus | 1958 | <.001 | 7.05 | -48 | -79 | 4 |
| | | | 6.10 | -12 | -97 | 16 |
| | | | 5.29 | 45 | -82 | 7 |

**Table S2.** Intra class coefficients between wave 1 and wave 2 brain activation in region of interest.

| ROI | contrast | ICC | 95% CI lower bound | 95% CI upper bound |
|---|---|---|---|---|
| Insula | negative > positive | -0.05 | -0.16 | 0.07 |
| | negative > neutral | 0.05 | -0.07 | 0.16 |
| | positive > neutral | -0.03 | -0.14 | 0.09 |
| IFG | negative > positive | -0.05 | -0.16 | 0.07 |
| | negative > neutral | 0.10 | -0.02 | 0.21 |
| | positive > neutral | 0.05 | -0.06 | 0.17 |
| mPFC | negative > positive | -0.08 | -0.20 | 0.03 |
| | negative > neutral | 0.06 | -0.05 | 0.17 |
| | positive > neutral | 0.03 | -0.09 | 0.14 |
| left DLPFC | negative > positive | 0.04 | -0.08 | 0.15 |
| | negative > neutral | 0.04 | -0.07 | 0.16 |
| | positive > neutral | 0.05 | -0.06 | 0.16 |

**Table S3.** Linear mixed effect model with noise blast duration as dependent variable. Output is based on type III ANOVA's using Satterthwaite's method. Significant effects are depicted in black fonts, insignificant effects in grey.

| **Linear Mixed Effect Models** | ***DF*** | | ***F*** | ***p*** |
|---|---|---|---|---|
| Condition | 2 | 2181.60 | 1033.61 | <0.001 |
| Wave | 1 | 2185.79 | 157.17 | <0.001 |
| Intervention Group | 1 | 217.81 | 0.07 | 0.795 |
| Conditon × Wave | 2 | 2181.60 | 16.06 | <0.001 |
| Conditon × Intervention | 2 | 2181.60 | 0.65 | 0.523 |
| Wave × Intervention | 1 | 2185.79 | 3.18 | 0.075 |
| Conditon × Wave × Intervention | 2 | 2181.60 | 0.14 | 0.874 |
| Estimated IQ | 1 | 406.16 | 0.01 | 0.928 |
| Gender | 1 | 217.86 | 1.21 | 0.273 |
| Conditon × Estimated IQ | 2 | 2181.60 | 13.55 | <0.001 |
| Conditon × Gender | 2 | 2181.60 | 2.26 | 0.104 |

**Table S4.** Linear mixed effect model with brain activation in regions of interest as dependent variable. Output is based on type III ANOVA's using Satterthwaite's method. Significant effects are depicted in black fonts, insignificant effects in grey.

| **Linear Mixed Effect Models** | | ***DF*** | ***F*** | ***p*** |
|---|---|---|---|---|
| ***Anterior Insula*** | | | | |
| Condition | 2 | 1526.24 | 27.79 | <0.001 |
| Wave | 1 | 1783.92 | 10.09 | <0.001 |
| Intervention Group | 1 | 181.80 | 0.00 | 0.953 |
| Conditon × Wave | 2 | 1526.24 | 2.06 | 0.127 |
| Conditon × Intervention | 2 | 1526.24 | 0.83 | 0.437 |
| Wave × Intervention | 1 | 1783.75 | 0.11 | 0.737 |
| Conditon × Wave × Intervention | 2 | 1526.24 | 0.93 | 0.394 |
| Estimated IQ | 1 | 313.02 | 1.88 | 0.171 |
| Gender | 1 | 182.36 | 0.19 | 0.663 |
| Conditon × Estimated IQ | 2 | 1526.24 | 0.61 | 0.544 |
| Conditon × Gender | 2 | 1526.24 | 0.83 | 0.435 |
| ***Inferior Frontal Gyrus*** | | | | |
| Condition | 2 | 1531.45 | 8.22 | <0.001 |
| Wave | 1 | 1804.24 | 0.54 | 0.461 |
| Intervention Group | 1 | 175.23 | 0.15 | 0.696 |
| Conditon × Wave | 2 | 1531.45 | 1.58 | 0.205 |
| Conditon × Intervention | 2 | 1531.45 | 0.60 | 0.549 |
| Wave × Intervention | 1 | 1804.11 | 3.17 | 0.075 |
| Conditon × Wave × Intervention | 2 | 1531.45 | 0.00 | 0.997 |
| Estimated IQ | 1 | 278.98 | 0.52 | 0.471 |
| Gender | 1 | 175.66 | 2.53 | 0.113 |
| Conditon × Estimated IQ | 2 | 1531.45 | 0.84 | 0.430 |
| Conditon × Gender | 2 | 1531.45 | 2.10 | 0.123 |

**Table S4. (**continued)

| **Linear Mixed Effect Models** | | ***DF*** | ***F*** | ***p*** |
|---|---|---|---|---|
| ***Medial Prefrontal Cortex*** | | | | |
| Condition | 2 | 1530.08 | 6.64 | 0.001 |
| Wave | 1 | 1790.59 | 5.61 | 0.018 |
| Intervention Group | 1 | 161.09 | 0.69 | 0.408 |
| Conditon × Wave | 2 | 1530.08 | 0.61 | 0.543 |
| Conditon × Intervention | 1 | 1790.41 | 2.43 | 0.119 |
| Wave × Intervention | 2 | 1530.08 | 0.26 | 0.769 |
| Conditon × Wave × Intervention | 1 | 314.32 | 0.03 | 0.853 |
| Estimated IQ | 1 | 161.60 | 0.93 | 0.337 |
| Gender | 2 | 1530.08 | 0.64 | 0.527 |
| Conditon × Estimated IQ | 2 | 1530.08 | 1.32 | 0.267 |
| Conditon × Gender | 2 | 1530.08 | 0.44 | 0.646 |
| ***Dorsolateral Prefrontal Cortex*** | | | | |
| Condition | 2 | 1532.09 | 8.21 | 0.000 |
| Wave | 1 | 1788.11 | 34.44 | 0.000 |
| Intervention Group | 1 | 187.98 | 0.00 | 0.993 |
| Conditon × Wave | 2 | 1532.09 | 2.53 | 0.080 |
| Conditon × Intervention | 2 | 1532.09 | 0.95 | 0.386 |
| Wave × Intervention | 1 | 1787.97 | 0.10 | 0.747 |
| Conditon × Wave × Intervention | 2 | 1532.09 | 0.04 | 0.958 |
| Estimated IQ | 1 | 300.71 | 5.67 | 0.018 |
| Gender | 1 | 188.49 | 0.05 | 0.827 |
| Conditon × Estimated IQ | 2 | 1532.09 | 4.21 | 0.015 |
| Conditon × Gender | 2 | 1532.09 | 1.98 | 0.138 |

**Table S5.** Linear mixed effect models with brain activation in regions of interest as dependent variable and noise blast duration added as factor. Output is based on type III ANOVA's using Satterthwaite's method. Significant effects are depicted in black fonts, insignificant effects in grey.

| Linear Mixed Effect Models | *DF* | | *F* | *p* |
|---|---|---|---|---|
| *Anterior Insula* | | | | |
| Condition | 2 | 1693.46 | 14.59 | <0.001 |
| Noise blast | 1 | 1908.4 | 5.47 | 0.019 |
| Wave | 1 | 1808.94 | 9.26 | 0.002 |
| Intervention Group | 1 | 181.67 | 0.01 | 0.907 |
| Conditon × Wave | 2 | 1659.61 | 1.18 | 0.306 |
| Conditon × Intervention | 2 | 1525.39 | 0.81 | 0.447 |
| Conditon × Noise blast | 2 | 1728.48 | 1.09 | 0.337 |
| Wave × Intervention | 1 | 1785.34 | 0.13 | 0.718 |
| Wave × Noise blast | 1 | 1913.4 | 1.74 | 0.188 |
| Conditon × Wave × Intervention | 2 | 1525.17 | 1.03 | 0.356 |
| Conditon × Wave × Noise blast | 2 | 1676.93 | 0.45 | 0.637 |
| Estimated IQ | 1 | 317.03 | 1.91 | 0.168 |
| Gender | 1 | 182.85 | 0.29 | 0.592 |
| Conditon × Estimated IQ | 2 | 1531.67 | 0.56 | 0.569 |
| Conditon × Gender | 2 | 1526.86 | 0.90 | 0.406 |
| *Inferior Frontal Gyrus* | | | | |
| Condition | 2 | 1709.10 | 5.45 | 0.004 |
| Noise blast | 1 | 1872.52 | 2.57 | 0.109 |
| Wave | 1 | 1830.86 | 0.83 | 0.363 |
| Intervention Group | 1 | 172.89 | 0.20 | 0.655 |
| Conditon × Wave | 2 | 1673.03 | 1.60 | 0.202 |
| Conditon × Intervention | 2 | 1527.06 | 0.52 | 0.592 |
| Conditon × Noise blast | 2 | 1752.47 | 0.34 | 0.711 |
| Wave × Intervention | 1 | 1804.98 | 3.14 | 0.077 |
| Wave × Noise blast | 1 | 1929.34 | 1.07 | 0.302 |
| Conditon × Wave × Intervention | 2 | 1526.81 | 0.01 | 0.994 |
| Conditon × Wave × Noise blast | 2 | 1695.95 | 0.06 | 0.940 |
| Estimated IQ | 1 | 280.10 | 0.80 | 0.373 |
| Gender | 1 | 174.04 | 2.89 | 0.091 |

**Table S5.** (continued)

| Linear Mixed Effect Models | *DF* | | *F* | *p* |
|---|---|---|---|---|
| Conditon × Estimated IQ | 2 | 1533.92 | 0.70 | 0.498 |
| Conditon × Gender | 2 | 1528.65 | 2.22 | 0.109 |
| *Medial Prefrontal Cortex* | | | | |
| Condition | 2 | 1699.33 | 5.73 | 0.003 |
| Noise blast | 1 | 1895.96 | 2.71 | 0.100 |
| Wave | 1 | 1816.12 | 8.20 | 0.004 |
| Intervention Group | 1 | 159.34 | 0.81 | 0.370 |
| Conditon × Wave | 2 | 1665.49 | 1.00 | 0.369 |
| Conditon × Intervention | 2 | 1529.67 | 0.52 | 0.594 |
| Conditon × Noise blast | 2 | 1736.36 | 0.48 | 0.621 |
| Wave × Intervention | 1 | 1791.29 | 2.22 | 0.136 |
| Wave × Noise blast | 1 | 1910.81 | 4.59 | 0.032 |
| Conditon × Wave × Intervention | 2 | 1529.44 | 0.20 | 0.815 |
| Conditon × Wave × Noise blast | 2 | 1684.20 | 1.31 | 0.271 |
| Estimated IQ | 1 | 316.05 | 0.00 | 0.959 |
| Gender | 1 | 160.39 | 1.17 | 0.282 |
| Conditon × Estimated IQ | 2 | 1536.06 | 0.39 | 0.675 |
| Conditon × Gender | 2 | 1531.15 | 1.39 | 0.248 |
| *Dorsolateral Prefrontal Cortex* | | | | |
| Condition | 2 | 1697.56 | 0.93 | 0.396 |
| Noise blast | 1 | 1911.70 | 4.32 | 0.038 |
| Wave | 1 | 1810.80 | 10.22 | 0.001 |
| Intervention Group | 1 | 187.45 | 0.00 | 0.958 |
| Conditon × Wave | 2 | 1664.25 | 3.34 | 0.036 |
| Conditon × Intervention | 2 | 1532.15 | 1.08 | 0.339 |
| Conditon × Noise blast | 2 | 1731.91 | 0.70 | 0.499 |
| Wave × Intervention | 1 | 1788.91 | 0.12 | 0.726 |
| Wave × Noise blast | 1 | 1918.14 | 0.07 | 0.797 |
| Conditon × Wave × Intervention | 2 | 1531.93 | 0.03 | 0.968 |
| Conditon × Wave × Noise blast | 2 | 1680.74 | 1.48 | 0.228 |
| Estimated IQ | 1 | 305.09 | 5.19 | 0.023 |
| Gender | 1 | 188.65 | 0.01 | 0.912 |
| Conditon × Estimated IQ | 2 | 1538.32 | 4.88 | 0.008 |
| Conditon × Gender | 2 | 1533.61 | 1.96 | 0.141 |

**Table S6.** MNI coordinates for local maxima activated for the whole brain contrast in the whole sample at wave 2 (N=358). Results were FDR cluster corrected ($p_{FDR}$<0.05), with a primary voxel-wise threshold of p<0.005.

| Anatomical Region | Voxels | pFDRcc | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Conjunction of negative>neutral and positive> neutral social feedback*** | | | | | | |
| Left Middle Occipital Gyrus | 3527 | <.001 | 12.45 | -48 | -79 | 1 |
| Right Fusiform Gyrus | | | 11.48 | 27 | -76 | -8 |
| Right Middle Occipital Gyrus | | | 10.43 | 48 | -73 | -2 |
| Left Insula | 206 | 0.024 | 5.47 | -30 | 26 | -8 |
| Left Insula | | | 3.42 | -42 | 17 | -2 |
| Left Insula | | | 3.06 | -39 | 23 | -17 |
| Right Inferior Frontal Gyrus | 266 | 0.013 | 4.96 | 48 | 20 | -2 |
| Right Inferior Frontal Gyrus | | | 4.80 | 33 | 26 | -14 |
| Right Insula | | | 3.75 | 39 | 32 | 4 |
| ***Negative > positive social feedback*** | | | | | | |
| Left Calcarine Gyrus | 554 | <.001 | 12.21 | -6 | -97 | 10 |
| Left Superior Occipital Gyrus | | | 11.78 | -12 | -94 | 19 |
| Right Superior Occipital Gyrus | | | 7.56 | 24 | -91 | 16 |
| Right Inferior Frontal Gyrus | 200 | 0.015 | 6.62 | 57 | 32 | 1 |
| Left Middle Occipital Gyrus | 608 | <.001 | 6.56 | -48 | 26 | 1 |
| Left Inferior Frontal Gyrus | | | 6.08 | -45 | 26 | -8 |
| Left Middle Temporal Gyrus | | | 5.57 | -54 | 8 | -23 |
| Left Middle Occipital Gyrus | 236 | 0.009 | 5.49 | -48 | -82 | 1 |
| Left Middle Occipital Gyrus | | | 4.42 | -54 | -73 | 1 |
| Left Middle Temporal Gyrus | | | 3.89 | -54 | -61 | 4 |
| Left Superior Medial Gyrus | 366 | 0.002 | 5.20 | -6 | 53 | 31 |
| Left Superior Frontal Gyrus | | | 4.80 | -18 | 47 | 31 |
| Left Superior Frontal Gyrus | | | 3.71 | -15 | 62 | 25 |
| Right Superior Temporal Gyrus | 567 | <.001 | 5.00 | 48 | -34 | 7 |
| Right Putamen | | | 4.90 | 33 | -13 | 4 |

**Table S6.** (continued)

| Anatomical Region | Voxels | pFDRcc | T | x | y | z |
|---|---|---|---|---|---|---|
| ***Negative > positive social feedback*** | | | | | | |
| Right Thalamus | | | 4.40 | 21 | -10 | -2 |
| Left Postcentral Gyrus | 337 | 0.003 | 4.69 | -42 | -16 | 40 |
| Left Postcentral Gyrus | | | 4.42 | -42 | -19 | 28 |
| Left SupraMarginal Gyrus | | | 3.83 | -42 | -37 | 28 |
| Left Middle Temporal Gyrus | 178 | 0.020 | 4.51 | -54 | -25 | -5 |
| Left Middle Temporal Gyrus | | | 4.34 | -57 | -37 | 1 |
| Left Middle Temporal Gyrus | | | 3.84 | -48 | -40 | 4 |
| Right Precentral Gyrus | 133 | 0.044 | 4.14 | 42 | -16 | 43 |
| Right Precentral Gyrus | | | 4.06 | 51 | -10 | 43 |
| Right Postcentral Gyrus | 313 | 0.003 | 3.89 | 24 | -34 | 61 |
| Right Postcentral Gyrus | | | 3.88 | 21 | -34 | 76 |
| Right Superior Parietal Cortex | | | 3.19 | 24 | -46 | 70 |
| ***Positive > negative social feedback*** | | | | | | |
| Right Lingual Gyrus | 3999 | <.001 | 13.80 | 6 | -73 | -2 |
| Right Lingual Gyrus | | | 9.61 | 21 | -70 | -5 |
| Left Lingual Gyrus | | | 8.94 | -18 | -85 | -2 |
| Right Middle Frontal Gyrus (DLPFC) | 4230 | <.001 | 7.72 | 39 | 35 | 43 |
| Right Superior Frontal Gyrus | | | 7.10 | 27 | 5 | 61 |
| Right Middle Frontal Gyrus (DLPFC) | | | 6.69 | 48 | 23 | 40 |