# Helping me, helping you: behavioral and neural development of social competence from childhood to adolescence

Dobbelaar, S.

# CHAPTER 2

# Development of social feedback processing and responses in childhood: an fMRI test-replication design in two age cohorts

*This chapter is published as:*
Dobbelaar, S., Achterberg, M., van Drunen, L., van Duijvenvoorde, A.C.K., van IJzendoorn, M.H., & Crone, E.A. (2022). Development of social feedback processing and responses in childhood: an fMRI test-replication design in two age cohorts. *Social Cognitive and Affective Neuroscience,* nsac039. doi:10.1093/scan/nsac039

## ABSTRACT

This study investigated behavioral and neural correlates underlying social feedback processing and subsequent aggressive behaviors in childhood in two age cohorts (test sample: n=509/n=385 and replication sample: n=354/n=195, 7-9 years old). Using a previously validated Social Network Aggression Task (Achterberg et al., 2020), we showed that negative social feedback resulted in most behavioral aggression, followed by less aggression after neutral and least aggression after positive feedback. Receiving positive and negative social feedback was associated with increased activity in the insula, medial prefrontal cortex and ventrolateral prefrontal cortex. Responding to feedback was associated with additional activation in the dorsolateral prefrontal cortex following positive feedback. This DLPFC activation correlated negatively with aggression. Furthermore, age analyses showed that older children showed larger reductions in aggression following positive feedback and more neural activation in dorsolateral prefrontal cortex to positive feedback compared to younger children. To assess the robustness of our results, we examined these processes in two independent behavioral/fMRI samples using equivalence testing, thereby contributing to replicable reports. Together, these findings demonstrate an important role of social saliency and regulatory processes where regulation of aggression rapidly develops between ages 7-9 years.


**Keywords:** social feedback processing, aggression, fMRI, childhood

## INTRODUCTION

Middle childhood, the developmental phase from approximately seven to ten years of age (DeFries et al., 1994), is an important period marked by rapid developmental changes in social competencies needed for developing adaptive social relations. For example, children more often experience and respond to social evaluations of peers. Prior research in adults and children showed that social rejection can lead to aggressive, retaliatory responses (Achterberg et al., 2017; Chester et al., 2014). An unanswered question, however, is how feedback processing and subsequent aggressive responses develop during childhood and which neural processes are involved. This study examined behavioral and neural responses to social feedback and subsequent aggression in middle childhood, using a test-replication design to optimize reliable and robust analyses.



### Neural correlates of social feedback processing

In prior studies, functional magnetic resonance imaging (fMRI) was used to specifically focus on neural activity while participants received social feedback (Achterberg et al., 2017, 2018; Gunther Moor et al., 2010; Guyer et al., 2012). The medial prefrontal cortex (MPFC), anterior insula (AI) and anterior cingulate cortex (ACC) showed enhanced activation during both positive and negative feedback compared to neutral feedback in prior studies in adults (Achterberg et al., 2016) and children (Achterberg et al., 2017, 2018, 2020). These regions have often been identified as regions involved in social cognition tasks, such as social evaluation and self-other referential processing (Apps et al., 2016; Blakemore & Mills, 2014; Crone et al., 2020; Yoon et al., 2018), but also in non-social tasks, such as cognitive control, attentional processes and memory (e.g., Euston et al., 2012; Menon & Uddin, 2010). In social feedback paradigms, the MPFC, AI and ACC might be specifically responsive to feedback that is salient (both positive and negative). Activation related to positive feedback has been reported in the ventral striatum, dorsolateral prefrontal cortex (DLPFC) and supplementary motor area (SMA), in adults (Achterberg et al., 2016), children and adolescents (Achterberg et al., 2017, 2018; Gunther Moor et al., 2010; Guyer et al., 2012). In contrast, negative feedback has been associated with activity in the superior medial prefrontal cortex (Achterberg et al., 2016, 2018), although it is not consistently observed across studies (Guyer et al., 2012). During negative feedback processing, enhanced DLPFC activation has been associated with subsequent aggression reduction (Achterberg et al., 2016, 2018; Riva et al., 2015). Given that replicability of fMRI findings in childhood is still relatively understudied and task-based fMRI shows low test-retest reliability (Elliott et al., 2020), the first aim of this study was to replicate previously observed neural

responses to social feedback and the relation to individual differences in aggression in a large sample of children aged 7-9-years (Open Science Collaboration, 2015; Schmidt, 2016; van IJzendoorn & Bakermans-Kranenburg, 2021).

## Neural correlates of responding to social feedback

The Social Network Aggression Task (Achterberg et al., 2016) has been developed to study the effects of receiving social feedback on subsequent aggressive responses. In this task, participants first received social peer feedback and were subsequently instructed to respond by sending a noise blast towards the peer giving the feedback. Negative feedback consistently resulted in the longest noise blasts compared to positive and neutral feedback, in adults and children (Achterberg et al., 2016, 2017, 2018; Chester et al., 2014). Table S1 shows an overview of previous findings in studies using the SNAT.

A still unexplored question of this study is which neural processes underly these responses to social feedback in childhood. In adults, studies on neural activity during aggressive following negative feedback show mixed results. For example, aggression following high (versus low) provocation, i.e., reactive aggression, has been associated with increased activation in the insula, ACC, MPFC, DLPFC, ventrolateral prefrontal cortex (VLPFC) (Dambacher et al., 2015; Krämer et al., 2007; Lotze et al., 2007; Repple et al., 2017), ventral striatum (Buades-Rotger et al., 2016; Chester & DeWall, 2016) and orbitofrontal cortex (OFC; Repple et al., 2017). On the other hand, increased aggression or punishment of unfair offers in an Ultimatum Game has also been related to *decreased* OFC and ventral MPFC activation (Beyer et al., 2015; Gilam et al., 2015; Mehta & Beer, 2010; S. F. White et al., 2014). Additionally, to our knowledge, only two prior studies investigated neural activation during forced aggressive responses following positive feedback. These studies reported increased activation in the DLPFC during responses to positive compared to negative feedback in the SNAT (van de Groep et al., 2021, 2022). Therefore, the second aim of our study was to investigate neural activity during responses to positive, neutral and negative social feedback.

In the SNAT, where participants are instructed to always send a noise blast, increased DLPFC activation following positive feedback (van de Groep et al., 2021) may possibly reflect inhibitory processes. DLPFC activation has previously been related to impulse control in non-social contexts (Blasi et al., 2006; Durston et al., 2002) and we hypothesize that it might play a similar role when responding to social feedback. Therefore, we additionally aimed to study whether individual differences in DLPFC activation following positive (versus negative) feedback was related to decreased aggressive responses following positive feedback.

## The current study

The current study tested the two processes of feedback processing and retaliation in the Leiden Consortium on Individual Development (L-CID), which consists of two independent age cohorts with overlapping time points in middle childhood (Crone et al., 2020), allowing for direct replication of our analyses. The experimental SNAT was used to measure neural activation on two events: the social feedback event, when participants received social feedback, and the noise blast event, when participants responded to social feedback by sending a noise blast.

Our first aim (*1*) was to replicate previously reported valence effects on neural activity during social feedback processing in AI, MPFC, VLPFC and DLPFC (Achterberg et al., 2020) and related brain-behavior correlations. We hypothesized that DLPFC activation following rejection is related to lower levels of aggression following negative feedback (Achterberg et al., 2018, 2020).

Our second novel aim (*2*) was (*a*) to investigate valence effects on neural activity during responses to social feedback and (*b*) to test whether individual differences in neural activity were meaningfully related to behavioral aggression. We expected that more activation following positive feedback relative to negative feedback would be related to decreased aggressive responses following positive feedback.

Finally, we performed two exploratory analyses (*3*). During childhood, the cognitive processes that are important for controlling behavior are still developing (Crone & Steinbeis, 2017). These developmental changes might already be noticeable during the course of middle childhood, since this period is marked by a rapid development in regulatory skills (Achterberg et al., 2020; Zelazo & Carlson, 2012). Therefore, we (*a*) explored age differences in aggression regulation between 7- and 9-year-olds to investigate how feedback processing and retaliation develop during childhood. Additionally, to further explore which children might be most prone to aggression following negative feedback, we (*b*) tested whether parental reported inhibitory control moderated the relation between activation in the affective salience network (AI and MPFC) during feedback and behavioral aggression, such that increased neural sensitivity to social feedback would be related to more reactive aggression only for children with low levels of inhibitory control (Chester et al., 2014).

# METHODS

## Participants

This study was part of the cohort-sequential longitudinal twin study of the Leiden Consortium on Individual Development (L-CID) (Crone et al., 2020; Euser et al., 2016). Children in the early childhood cohort (ECC) were followed from 3 to 9 years of age, whereas children in the middle childhood cohort (MCC) were followed from 7 to 13 years (Crone et al., 2020). The present study focused on the first visit of the MCC (test sample) and the fifth visit of the ECC (replication sample), which were lab visits including MRI scans. The overlap in age (7-9 years old) between cohorts allowed for a replication *within the same study* and for replication of effects that were previously reported by Achterberg et al. (2018, 2020) on the MCC data.

In the MCC, data was collected in 2015-2016. Data and characteristics of the test sample were previously reported in Achterberg et al. (2018). Our behavioral test sample consisted of 509 participants (mean age: 7.95 ± 0.67 years, 51% girls). In total, 124 participants were excluded from MRI analyses (Supplementary Materials, Figure S1), resulting in an MRI test sample of 385 participants (mean age: 7.99 ± 0.68 years, 53% girls).

In the ECC, data was collected in 2019. Our behavioral replication sample consisted of 354 participants (mean age: 8.00 ± 0.62 years, 53.1% girls). In total, 159 participants were excluded from MRI analyses (Supplementary Materials, Figure S1), resulting in an MRI replication sample of 195 participants (mean age: 8.08 ± 0.62 years, 57.4% girls). Sample characteristics are presented in Table 1. Prior to the first visit, informed consent was obtained from both parents. The study was approved by the Dutch Central Committee on Research Involving Human Subjects (CCMO).

**Table 1.** Demographic characteristics.

| | **Test sample (MCC)** | | **Replication sample (ECC)** | |
|---|---|---|---|---|
| | **Behavioral** | **MRI** | **Behavioral** | **MRI** |
| N | 509 | 385 | 354 | 195 |
| Girls (%) | 50.9% | 53.0% | 53.1% | 57.4% |
| Age (SD) | 7.95 (0.67) | 7.99 (0.68) | 8.00 (0.62) | 8.08 (0.62) |
| Age range | 7.02 – 9.68 | 7.02 – 9.68 | 6.93 – 9.62 | 7.02 – 9.49 |
| Left-handed (%) | 12.8% | 12.2% | 13.4% | 12.3% |
| IQ (SD) | 103.62 (11.77) | 104.03 (11.84) | 103.24 (10.69) | 103.21 (10.74) |
| IQ range | 72.50 – 137.50 | 72.50 – 137.50 | 75.0 – 130.0 | 77.50 – 130.0 |
| Caucasian (%) | 91% | 93% | 92% | 95% |
| SES low – middle - high (%) | 9 – 45 – 46% | 8 – 43 – 49% | 5 – 35 – 60% | 2 – 33 – 65% |

*Note.* SES = Socioeconomic status, based on parental education at T1.

## Measures

### *Social Network Aggression Task*

Feedback effects and subsequent aggressive responses were measured using the Social Network Aggression Task (SNAT), which has been validated as a reliable measure of aggression following social feedback (Achterberg et al., 2016, 2017). Participants filled out a personal profile at home and sent this back a week before the lab visit. During the lab visit, participants received feedback by peers on whether they liked the answers on their profile. This feedback could be positive (a green thumb up), neutral (a grey circle) or negative (a red thumb down). Subsequently, participants had to imagine sending a noise blast towards the peer who had given the feedback by pressing a button with their right index finger. They could decide the intensity of the noise blast by pressing the button for a longer duration (Figure 1A), which was used as a measure of behavioral aggression. Participants were specifically instructed to imagine sending the noise blast to reduce the amount of deception used in the task. Participants did not know that peers in the tasks were not real but were morphed photographs from an existing database. Each photograph was presented with either positive, neutral or negative feedback. The order of trials was pseudo-randomized.



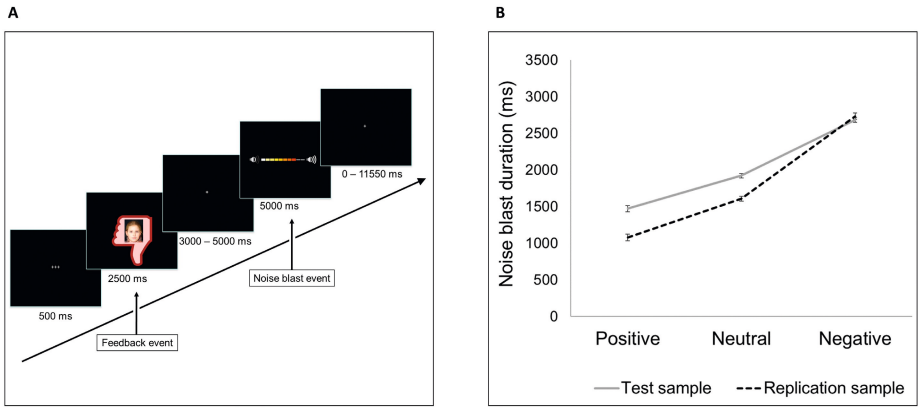


**Figure 1.** SNAT. (A) Schematic representation of a trial with negative social feedback. (B) Noise blast duration for each feedback condition in the original sample (solid line) and replication sample (dotted line). Error bars represent standard errors.

The SNAT consisted of sixty trials (three blocks of twenty trials) in the test sample. In the replication sample, we shortened the MRI session, based on findings that scan quality decreases with increasing length of the scanning procedure (Achterberg & van der Meulen, 2019). Since we found comparable main effects in the MCC data when

analyzing two versus three blocks of the SNAT, we chose to include forty trials (two blocks of twenty trials) in the replication sample.

### *Inhibitory control (TMCQ)*

To measure inhibitory control, the subscale Inhibitory Control of the parental reported version of the Temperament in Middle Childhood Questionnaire (TMCQ) was used (Simonds et al., 2007). The primary caregiver, who spent most time with the children at the start of the study, completed the questionnaire for both twin children separately. This subscale consisted of eight questions (e.g., "My child can stop him/herself from doing things too quickly") that were answered on a five-point scale (1 = 'extremely untrue', 5 = 'extremely true'). Cronbach's alpha was 0.67. A mean score over the seven items was computed, such that higher scores indicate more inhibitory control.

## MRI data acquisition

MRI scans were acquired on a Philips Ingenia 3.0 Tesla MR system, using a standard whole-head coil. Participants from both samples were scanned using the same MRI scanner. Foam inserts were added within the head coil to minimize head motion. Participants viewed the SNAT on a screen that was placed behind the scanner and could be viewed through a mirror on the head coil. Functional MRI scans were collected using T2*-weighted echo planar imaging (EPI). The first two volumes were discarded to allow for equilibration of T1 saturation effects (Field of View (FOV) = 220 (anterior-posterior, a-p) x 220 (right-left, r-l) x 111.65 (foot-head, f-h) mm; repetition time (TR) = 2.2 s, echo time (TE) = 30 ms; flip angle (FA) = 80°; sequential acquisition; 37 slices; voxel size = 2.75 x 2.75 x 2.75). In the test sample, the SNAT consisted of three blocks (block 1: 148 volumes, block 2: 142 volumes, block 3: 141 volumes, see Achterberg et al., 2018). In the replication sample, the SNAT consisted of two blocks (block 1: 148 volumes, block 2: 142 volumes). The modification between the two samples was introduced to decrease the total scan time (Achterberg & van der Meulen, 2019). In between blocks, scanning was paused to give the participant a small break. Additionally, a high-resolution 3D T1 scan was collected as anatomical reference (FOV = 224 (a-p) x 177 (r-l) x 168 (f-h) mm; 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

### *fMRI preprocessing*

fMRI data were analyzed in SPM8 (Wellcome Department of Cognitive Neurology, London) for consistency with the previously published study. First, images were corrected for slice timing acquisition and rigid body motion. Functional volumes were

spatially normalized to T1 templates using 12-parameter affine transform mapping and non-linear transformation involving cosine basis functions. Templates were based on MNI-305 stereotaxic space (Cocosco et al., 1997). Due to missing 3D T1 data, data of five participants in the test sample were normalized to an EPI template. Volumes of each participant were resampled to 3x3x3 mm voxels. Data were spatially smoothed using a 6 mm full-width-at-half-maximum isotropic Gaussian kernel. For all participants, translational movement parameters were calculated. Participants were included in the first level analysis when they had at least two runs of fMRI data with <3 mm maximum motion in every direction (x,y,z) (in line with the original study of Achterberg et al., 2018). We included detailed information on motion parameters and small spikes (i.e., motion between 0.9 and 3mm in any volume, see Siegel et al., 2014) in Table S8. Most participants showed a limited number of small spikes (participants with small spikes in more than 10% of volumes in test sample: n = 2 (10.8% and 13.4%); in replication sample: n = 4 (12.4% - 19.0%), see Figure S4). We allowed this minimal amount of motion in order to weigh up against the possible extra loss of trials (Siegel et al., 2014).



### *First level analyses*

We analyzed individual participant's data using a general linear model in SPM8. Two events were convolved with the hemodynamic response function (HRF) to model the fMRI time series. The onset of feedback delivery was modeled as the feedback event, with a zero duration and with valence of the feedback as separate regressors ('positive', 'neutral', 'negative'). The start of the noise blast was modeled as the noise blast event, with the HRF modeled for the length of the noise blast duration and with noise blast following positive, neutral and negative feedback as separate regressors (Achterberg et al., 2018; van de Groep et al., 2021, 2022). Trials on which participants did not respond in time were marked as invalid and excluded from analysis. Six motion regressors were added to the model as covariates of no interest. We used the least-squares parameter estimates of height of the best-fitting canonical HRF for each condition in pair-wise contrasts. The resulting subject-specific contrasts images were used in the group-level analyses.

### *Second level analyses*

To investigate feedback effects on neural responses during the social feedback event, we performed a full-factorial ANOVA with three levels (feedback: 'Positive', 'Neutral', 'Negative') on data of the replication sample. The whole-brain analyses of the test sample were previously reported in Achterberg et al. (2018). We calculated the 'Negative>Positive' and 'Positive>Negative' contrasts and investigated activation that was specific to both positive and negative feedback (compared to neutral

feedback) by performing a conjunction analysis (of the contrasts 'Negative>Neutral' and 'Positive>Neutral') with the 'logical AND' strategy. This strategy requires activation in both contrasts to be individually significant (Nichols et al., 2005).

In addition, we investigated feedback effects on neural responses during the noise blast event by performing a full-factorial ANOVA with three levels (noise after feedback: 'PositiveNoise', 'NeutralNoise', 'NegativeNoise'). These whole-brain analyses were conducted in both samples, since they were not previously reported. We explored the following contrasts: 'PositiveNoise>NegativeNoise' (and vice-versa), 'PositiveNoise>NeutralNoise' (and vice-versa), 'NegativeNoise>NeutralNoise' (and vice-versa).

All results were family wise error (FWE) cluster level corrected ($p_{FWEcc}$<.05) with an initial voxel-wise threshold of $p$<.005 (uncorrected) (Achterberg et al., 2020). We report coordinates for local maxima in MNI space. Results of the whole-brain analyses are reported in the Supplementary Materials. Untresholded statistical maps of the whole-brain contrasts are available on Neurovault (Gorgolewski et al., 2015) via https://neurovault.org/collections/GMFZUABO/.

### *Region of interest analyses*

For the social feedback event, we used the regions of interest (ROIs) that were previously used in Achterberg et al. (2020): bilateral AI, bilateral VLPFC, dorsal MPFC (DMPFC) and left DLPFC. Parameter estimates (PE, average Beta values) for each participant in the replication sample were extracted from the contrasts 'Positive>Fixation', 'Neutral>Fixation' and 'Negative>Fixation'.

For the noise blast event, we based our ROIs on whole-brain activation in the test sample and used the same ROIs in our replication sample. In the test sample, clusters of activation from the whole-brain contrasts were extracted using SPM8's MarsBar toolbox (Brett et al., 2002). Based on a priori hypotheses, we selected the following five regions: bilateral AI and bilateral VLPFC (from the 'AllNoise>Fixation' contrast), the bilateral MPFC and bilateral OFC (from the conjunction contrast of 'NegativeNoise>NeutralNoise' and 'PositiveNoise>NeutralNoise') and the bilateral DLPFC (from the conjunction contrast of 'PositiveNoise>NegativeNoise' and 'NeutralNoise>NegativeNoise'). These clusters of activation were masked with regions from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002) to construct our final ROIs (see https://osf.io/tc83e/ for 3D nifti files). For both samples, PEs for each participant were extracted from the contrasts 'PositiveNoise>Fixation', 'NeutralNoise>Fixation' and 'NegativeNoise>Fixation'.

## Statistical analyses

Outliers were defined as Z-scores below -3.29 or above 3.29 on each variable and these data points were excluded from subsequent analyses. No outliers were observed in the behavioral data. In the ROI data, maximally 2% of the data was defined as outliers.

### *Confirmatory analyses*

**Valence effects.** To test the effects of feedback condition on noise blast duration and ROI activation during feedback and noise blast, we used a linear mixed model approach in R (R Core Team, 2013) using the lme4 package (Bates et al., 2015). Data were fitted on the average noise blast duration (for behavioral analyses) and average PEs (for ROI analyses on the feedback and noise blast events) for each feedback condition (positive, neutral, negative). The use of linear mixed models allowed us to add two random factors to our model: ChildID to account for nesting of feedback conditions within children and FamilyID to account for nesting of children within families. Note that our participants were twins, who shared the same family environment within a twin pair. Feedback condition was added as fixed effect (three levels: positive, neutral, negative) and sex as covariate, including interaction effects with condition. Thus, we defined our linear mixed model in R as: Noise blast duration/ROI ~ Condition×Sex + (1|ChildID) + (1|FamilyID). We inspected the results with type III ANOVA's using Satterthwaite's method. Significant main effects were post-hoc inspected using least-square means with Kenward-Roger corrected degrees of freedom and Bonferroni-adjusted *p*-values (Achterberg et al., 2020).

Additionally, we checked in sensitivity analyses whether the addition of a random slope of condition on family level changed the results. We defined this model in R as: ROI ~ Condition×Sex + (1|ChildID) + (1+Condition|FamilyID) and checked whether the model fit increased compared to the original model, using log-likelihood tests with the anova function in R. For models on ROI activation during the feedback event, adding a random slope did not increase model fit (all $p > .05$). For models on ROI activation during the noise blast event, adding a random did increase the model fit (all $p < .05$). For this event, we report the results of the linear mixed models with random slope in Table S6 and S7.

Since the L-CID study included a randomized controlled trial, approximately 40% of the families received an intervention to promote positive parenting and sensitive discipline (VIPP-SD, see Euser et al., 2016, 2021) between T2 and T3 (Crone et al., 2020). As a sensitivity check, we repeated our replication analyses in the control group of the VIPP-SD. In an additional sensitivity analysis, we checked for possible IQ effects

by adding IQ as covariate to the linear mixed models. In both analyses, results did not meaningfully differ from results described in the paper (see Supplementary Materials).

**Brain-behavior analyses.** To test whether we could replicate the finding that increased DLPFC activation during feedback was related to shorter noise blasts following negative feedback, we used the DLPFC ROI from the whole brain regression on the test sample reported in (Achterberg et al., 2020). Using least square regressions, we specifically tested whether DLPFC activation (PE) in this ROI during the feedback event ('Negative-Neutral') negatively predicted noise blast duration (Δ negative-neutral) in the independent replication sample.

For the noise blast event, we performed a whole brain regression in the test sample ('PositiveNoise-NegativeNoise') with the difference in noise blast duration following positive and negative feedback as regressor (Δ positive-negative). Results were FWE voxel level corrected with $p_{FWE} < .05$. For our replication analysis, we tested whether DLPFC activation (PE) in this ROI during the noise blast event ('PositiveNoise–NegativeNoise') predicted noise blast duration (Δ positive–negative) in the replication sample. Activation in the left and right DLPFC was correlated, $r = .91$, $p < .001$, so we used the mean score in the replication analysis. Because twins are nested in families, the assumption of homoscedasticity is violated. To correct for this violation, we used heteroscedasticity-consistent standard errors (HCSE) estimators (Hayes & Cai, 2007) in our regression analyses.

**Equivalence testing.** To investigate replication effects, we used equivalence testing in R (TOSTER package 0.3.4; Lakens, 2017) in addition to null-hypothesis significance testing for effects that we could not replicate. Equivalence testing tests whether the hypothesis that replication effects are large enough to be considered meaningful, i.e., a smallest effect size of interest (SESOI; Lakens et al., 2018), is rejected. For our replication analyses, we defined the SESOI as the lower boundary of the confidence interval of the effect in the test sample (Lakens et al., 2018; Perugini et al., 2014).

### *Exploratory analyses*

Exploratory, we tested for age effects and for moderation effects by inhibitory control. For these analyses, we combined data of both samples to increase power.

**Age effects.** First, we tested for age effects on feedback processing and retaliation in middle childhood (behavioral: n = 863, MRI: n = 580). We added age (rounded to two decimal places, grand mean centered) to our linear mixed models in R, including interaction effects of age and feedback condition. Because we combined two samples, we controlled for cohort (MCC or ECC) in our analyses. Thus, our linear mixed model in R was defined as: Noise blast duration/ROI ~ Condition×Age + Condition×Sex + Condition×Cohort + (1|ChildID) + (1|FamilyID). In sensitivity analyses, we added a

random slope of condition on family level to the linear mixed model to check whether this increased model fit and changed the results. This linear mixed model in R was defined as: ROI ~ Condition×Age + Condition×Sex + Condition×Cohort + (1|ChildID) + (1+Condition|FamilyID). Again, model fit increased only for linear mixed models on ROI activation during the noise blast event.

**Moderation inhibitory control.** To test whether inhibitory control moderated the association between MPFC/AI activation during feedback and subsequent noise blast duration, we performed two moderation analyses (n = 549).

Moderation analyses were performed using the PROCESS macro (version 3.5) in R (Hayes, 2017). First, we performed a moderation analysis with AI activation during feedback ('Negative–Neutral') as independent variable, noise blast duration (Δ negative–neutral) as dependent variable and inhibitory control as moderator. Next, we performed a second moderation analysis with MPFC activation during feedback ('Negative–Neutral') as independent variable, noise blast duration (Δ negative-neutral) as dependent variable and inhibitory control as moderator. To control for sample effects, sample (test or replication) was added as covariate in both analyses.



# RESULTS

## Behavioral results

### *Valence effects*

**ECC Replication sample.** Replicating prior findings, we found a main effect of social feedback condition on noise blast duration, $F(2,708) = 490.76$, $p < .001$ (Figure 1B). Mean noise blast duration was longest following negative feedback, followed by shorter noise blasts following neutral and shortest following positive feedback (all pairwise comparisons $p < .001$). Additionally, the model revealed a sex effect ($F(1,179.86) = 9.55$, $p = .002$), indicating longer noise blast durations for boys than girls.

## Neural activation during feedback

### *Valence effects*

**ECC Replication sample.** We replicated previously reported valence effects on neural activation during feedback (Figure 2; Table S2 and S3). The differences that were found compared to the test sample were non-equivalent to zero (see Supplementary Materials).

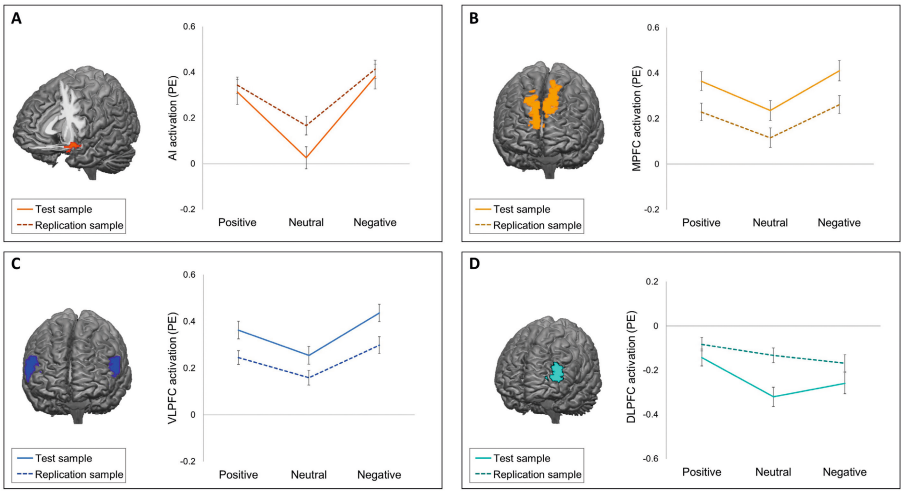


**Figure 2.** Neural activation (PEs) during the feedback event for each feedback condition in four ROIs in the test sample (solid lines) and replication sample (dotted lines). Error bars represent standard errors. (A) Activation in the AI. (B) Activation in the MPFC. (C) Activation in the VLPFC. (D) Activation in the DLPFC.

### *Brain-behavior analyses*

**ECC Replication sample.** We observed a significant negative correlation between noise blast duration following negative (vs neutral) feedback and DLPFC activation during negative feedback (vs neutral feedback), $r$ = -.16, $p$ = .023 (HCSE-corrected, Figure 4A). This effect was specific to noise blast duration following negative versus neutral feedback and negative versus positive feedback, as no correlations were found between DLPFC activation and noise blast duration following positive, negative or neutral, or positive versus neutral feedback (Table 2).

**Table 2.** Correlation analyses between DLPFC activation during feedback and subsequent noise blast duration in the test and replication sample.

| | | Noise blast duration negative - neutral | Noise blast duration negative - positive | Noise blast duration negative | Noise blast duration positive – neutral | Noise blast duration positive | Noise blast duration neutral |
|---|---|---|---|---|---|---|---|
| DLPFC* replication | *r* | -.16 | -.12 | -.02 | -.09 | -.11 | -0.08 |
| sample | *p* | .023 | .032 | .809 | .179 | .119 | .354 |
| DLPFC* test | *r* | -.06 | -.09 | .00 | -.02 | -.01 | .01 |
| sample | *p* | .249 | .124 | .955 | .695 | .899 | .852 |

*Note.* *P*-values are corrected with heteroscedasticity-consistent standard error estimates. *DLPFC ROI from whole-brain regression (with difference in noise blast duration negative – neutral as regressor) of Achterberg et al. (2020).

## Neural activation during noise blast

Noise blast neural activation has not been reported previously for any of the cohorts; therefore, we report first the results of the MCC test sample and then the results of the ECC replication sample.

### *Valence effects*

**MCC Test sample.** In all five ROIs of the noise blast event, we found a main effect of feedback condition on neural activation (Figure 3). However, patterns of activation differed between ROIs (see Table S5 for post-hoc test statistics). For the MPFC ($F(2,752.55) = 13.75$, $p < .001$) and OFC ($F(2,756.53) = 11.67$, $p < .001$), activation during the noise blast event was lower following neutral compared to negative feedback ($p \leq .01$) and lower following neutral compared to positive feedback ($p \leq .003$). For MPFC activity, we also observed an interaction effect of condition and sex, $F(2,752.55) = 3.89$, $p = .02$, indicating stronger condition effects for girls than for boys. For the DLPFC ($F(2,757.79) = 43.83$, $p < .001$), VLPFC ($F(2,745.60) = 29.73$, $p < .001$) and AI ($F(2,748) = 22.89$, $p < .001$), activation following negative feedback was significantly lower than following neutral feedback (all $p$'s$\leq$.01) and positive feedback (all $p$'s $< .001$). For DLPFC and VLPFC, but not AI, activation following neutral feedback was also lower compared to positive feedback (both $p$'s $< .001$).

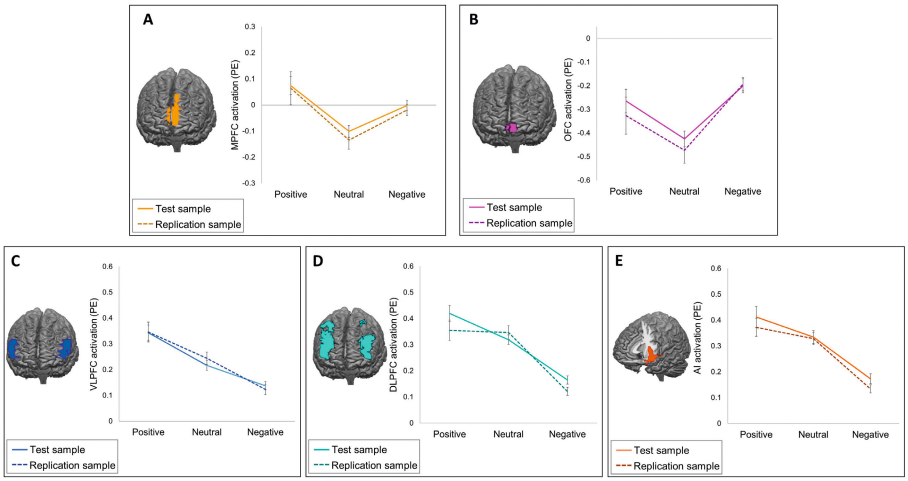


**Figure 3.** Neural activation (PEs) during the noise blast event for each feedback condition in five ROIs in the test sample (solid lines) and replication sample (dotted lines). Error bars represent standard errors. (A) Activation in the MPFC. (B) Activation in the OFC. (C) Activation in the VLPFC. (D) Activation in the DLPFC. (E) Activation in the AI.

**ECC Replication sample.** We repeated the ROI analyses in the replication sample and found similar patterns of activation (Figure 3) in all five ROIS (MPFC: $F(2,386.74) = 6.04$, $p = .003$; OFC: $F(2,383.42) = 6.69$, $p = .001$; DLPFC: $F(2,381.04) = 25.90$, $p < .001$; VLPFC: $F(2,378.81) = 20.79$, $p < .001$; AI: $F(2,381.45) = 29.14$, $p < .001$). However, some differences were found compared to the original sample. For the MPFC, the difference between negative feedback and neutral feedback was no longer significant, $p = .181$. Equivalence testing against raw equivalence bounds of -0.01 and 0.01 revealed that this effect was statistically non-equivalent to zero, $t(191) = 3.03$, $p = .999$. For OFC activation, the difference between positive feedback and neutral feedback was no longer significant, $p = .191$. Equivalence testing against equivalence bounds of -0.04 and 0.04 again revealed that this effect was non-equivalent to zero, $t(189) = 1.23$, $p = .890$. Finally, the difference between DLPFC activation in the positive and neutral feedback conditions was no longer significant, $p = 1$. Equivalence testing against equivalence bounds of -0.04 and 0.04 was non-significant, $t(189) = -0.76$, $p = .224$. Post-hoc statistics are presented in Table S5. Results were comparable using linear mixed models with random slope of condition, see Table S6 and S7.

### *Brain-behavior analyses*

**MCC Test sample.** A whole brain regression analysis with noise blast duration (negative – positive) as regressor resulted in increased activation in several areas including the left and right DLPFC (Figure 4B). Specifically, we found a negative relation between DLPFC activation during the noise blast and noise blast duration. A post-hoc correlation analysis showed that more DLPFC activation during the noise blast was associated with shorter noise blasts, $r = -.45$, $p < .001$ (HCSE-corrected, see also Table 3).

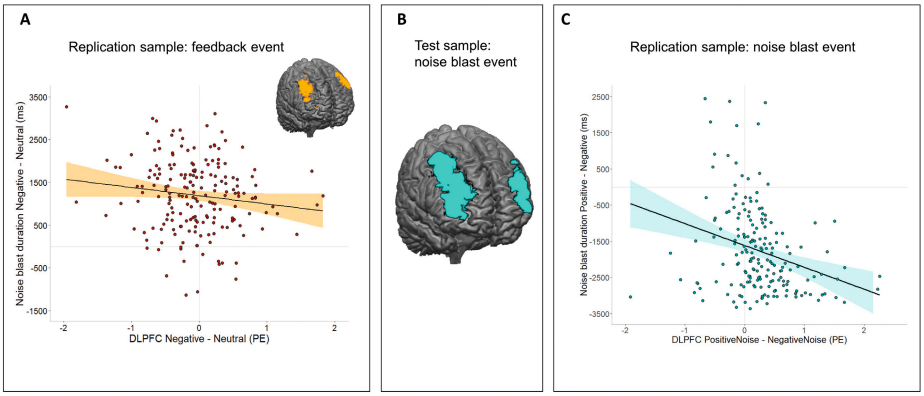


**Figure 4.** Brain-behavior relations during the feedback event. (A) Brain-behavior association during feedback in the replication sample. (B) Significant cluster of activation in the bilateral DLPFC in the test sample for the contrast PositiveNoise > NegativeNoise with noise blast (Δ positive – negative) as regressor. (C) Brain-behavior association during noise blast in the replication sample.

**ECC Replication sample.** Using the DLPFC from the whole brain regression in the test sample as ROI, we replicated the significant negative relation between noise blast duration (positive – negative) and activation in the DLPFC, $r$ = -.31, $p$ < .001 (HCSE-corrected, Figure 4C). Additional correlation analyses, however, revealed that this DLPFC activation was also negatively related to noise blast duration following positive versus neutral and negative versus neutral feedback (all $p \leq .001$, Table 3).

**Table 3.** Correlation analyses between DLPFC activation during noise blast and noise blast duration in the test and replication sample.

| | | Noise blast duration positive - negative | Noise blast duration negative - neutral | Noise blast duration negative | Noise blast duration positive – neutral | Noise blast duration positive | Noise blast duration neutral |
|---|---|---|---|---|---|---|---|
| DLPFC* replication sample | *r* | -.31 | -.46 | -.51 | -.30 | -.23 | -.44 |
| | *p* | < .001 | < .001 | < .001 | < .001 | .001 | < .001 |
| DLPFC* test sample | *r* | -.45 | -.42 | -.37 | -.26 | -.37 | -.34 |
| | *p* | < .001 | < .001 | < .001 | < .001 | < .001 | < .001 |

*Note.* *p*-values are corrected with heteroscedasticity-consistent standard error estimates. *DLPFC ROI from whole-brain regression (with difference in noise blast duration positive-negative as regressor) in original sample (MCC).

## Exploratory analyses

***Age effects*** We explored age-related effects, by testing the effects of age on the valence effects on noise blast duration and ROI activation. There was a significant main effect of age on noise blast duration, $\boldsymbol{F}(1,436.73) = 8.38$, $\boldsymbol{p} = .004$, $\eta^2_p = 0.02$, indicating higher mean noise blast duration for younger children. Results also revealed an interaction effect between age and feedback condition, $\boldsymbol{F}(2,1726) = 23.27$ $\boldsymbol{p} < .001$, $\eta^2_p = 0.03$, such that noise blast duration following negative feedback increased and noise blast duration following positive feedback decreased with age (Figure 5A).

We did not find any age effects on DLPFC, VLPFC, MPFC and AI activation during feedback (all $p > .251$).

During the noise blast, results revealed a main effect of age on activation in the DLPFC ($F(1,325.03) = 12.33$, $p < .001$, $\eta^2_p = 0.04$), VLPFC ($F(1,305.54) = 10.28$, $p = .001$, $\eta^2_p = 0.03$), MPFC ($F(1,297.98) = 4.38$, $p = .037$, $\eta^2_p = 0.01$) and AI ($F(1,332.56) = 6.30$, $p = .012$, $\eta^2_p = 0.02$), indicating increased activation for older children. In four ROIs, there was a significant interaction between condition and age (DLPFC: $F(2,1137.65) = 11.02$, $p < .001$, $\eta^2_p = 0.02$; VLPFC: $F(2,1123) = 9.16$, $p < .001$, $\eta^2_p = 0.02$; MPFC: $F(2,1140.72) = 4.58$, $p = .010$, $\eta^2_p = 0.01$; AI: $F(2,1120.37) = 4.73$, $p = .009$, $\eta^2_p = 0.01$). These interactions demonstrated that activation during aggression following positive feedback significantly increased with increasing age (all $p$'s $< .001$), whereas activation during aggression following negative and neutral feedback remained stable (all $p > .085$, see Figure 5C). The results were comparable when using linear mixed models with a random slope of condition, with the exception that the main and interaction effect of the MPFC were no longer significant (main effect of age: $F(1,328.57) = 3.75$, $p = .077$, $\eta^2_p = .01$; interaction effect: $F(2,423.29) = 2.59$, $p = .076$, $\eta^2_p = .01$). Thus, these results were less robust and should be interpreted with caution.

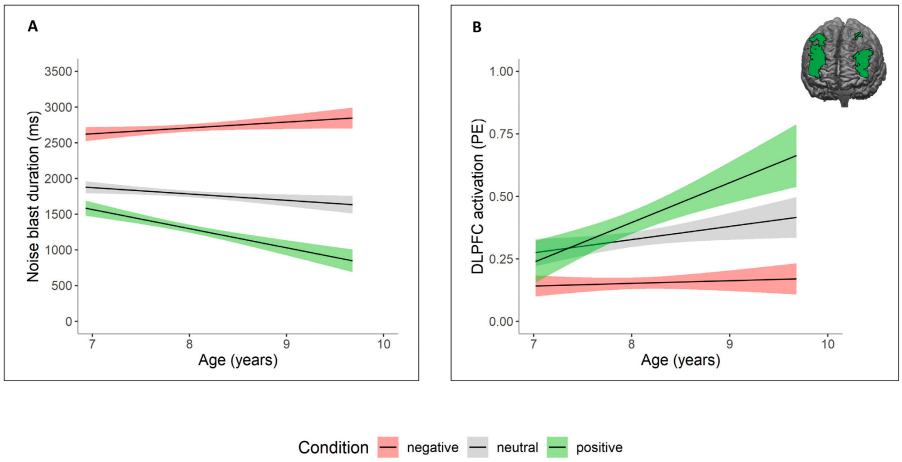


**Figure 5.** Age effects for each feedback condition. (A) Age effects on noise blast duration: older children show more behavioral differentiation between feedback conditions. (B) Age effects on DLPFC activation during noise blast: older children show increased activation following positive feedback. Similar relations were found for MPFC, VLPFC and insula activation.

### *Moderation inhibitory control*

To test whether inhibitory control moderated the relation between MPFC/AI activation and noise blast duration, we performed two moderation analyses. Results showed that there were no moderation effects of inhibitory control (see Supplementary Materials). In sensitivity analyses, we checked whether results were similar using data of the Stop Signal task as a behavioral proxy of inhibitory control (Williams et al., 1999). Again, we did not find moderation effects (see Supplementary Results).

## DISCUSSION

This study examined the behavioral and neural mechanisms of feedback processing and subsequent retaliatory responses in middle childhood, using a test-replication design. Specifically, our study answered three main questions. First, we replicated the behavioral and neural findings on feedback processing of Achterberg et al. (2020). Behaviorally, negative social feedback resulted in aggressive responses, which is a consistent finding in children (Achterberg et al., 2017, 2018) and adults (Chester et al., 2018). fMRI analyses revealed most activation in the AI and MPFC during positive and negative feedback and in the VLPFC during negative feedback, confirming prior work showing that these regions respond to socially salient events (Achterberg et al., 2018; Dalgleish et al., 2017) and rejection (Eisenberger et al., 2003). Second, in two

independent samples, we consistently found valence effects during responses to social feedback. The MPFC and OFC showed most activation following positive and negative feedback, and the DLPFC, VLPFC and AI showed least activation following negative feedback. Third, we observed brain-behavior relations during feedback and aggressive behavior: more DLPFC activation during negative feedback processing was related to less subsequent aggression, whereas more DLPFC activation during the noise blast event was generally related to less aggression (independent of the valence of the feedback). Age analyses revealed that older children showed less aggression in general and differentiated more in noise blasts following positive and negative feedback. Moreover, we found age effects in brain activation during the noise blast, but not during social feedback processing. That is, older children showed more neural activation during the noise blast following positive feedback. These findings indicate that responses to social feedback show developmental changes from middle to late childhood.

A novel question in this study was to examine the neural responses during responses to social feedback. Given that few studies incorporated a condition in which participants received positive feedback before being instructed to send a noise blast, it is interesting to note that in our study, sending noise blasts following positive feedback resulted in a wide network of activation, including the DLPFC, VLPFC and AI. These findings are in line with a recent study by van de Groep et al. (2021), who also reported increased activation in several regions including the lateral PFC when responding to positive compared to neutral and negative feedback. Possibly, responses to positive feedback might result in more cognitive control and intentional inhibition tendencies (Filevich et al., 2012) compared to negative and neutral feedback, especially in the SNAT paradigm where participants were instructed to always send a noise blast. Indeed, noise blast duration was shortest following positive feedback, which may be indicating intentional inhibitory processes. Activation in lateral prefrontal regions, ACC and supplementary motor area has previously been linked to inhibitory control and executive functioning (Crone & Steinbeis, 2017; Schel et al., 2014) and this network seems to be in place already in middle childhood (Engelhardt et al., 2019). Alternatively, however, activation in these regions following positive feedback could also be indicative of increased confusion when having to respond aggressively to positive feedback or a contradiction of feelings of fairness. Indeed, the insula and lateral prefrontal cortex were previously found to be involved in responding to conflicting information (Kim et al., 2010; Zaki et al., 2010) and in normative decision making (Buckholtz, 2015; Feng et al., 2015). Interestingly, in our study, the VLPFC and AI showed differential effects of feedback condition during the feedback and noise blast events, which may suggest a flexible role in both signaling for social saliency (Dalgleish

et al., 2017) and emotion regulation following threat (Zhao et al., 2021), as well as in inhibitory control processes (Nelson & Guyer, 2011; Swick et al., 2011). However, we did not use the exact same ROIs during the feedback and noise blast events, and thus these comparisons should be interpreted with caution.

Additionally, we examined brain-behavior relations between aggression and DLPFC activation. First, we replicated the finding that more DLPFC activation during negative (versus neutral) feedback was related to less aggression (Achterberg et al., 2020). This correlation was specific to DLPFC activation in the negative feedback condition, which could fit with studies on the role of the DLPFC in emotion regulation and reappraisal of negative events (Ochsner et al., 2012; Silvers & Guassi Moreira, 2019). During the noise blast event, specificity analyses revealed that more DLPFC activation was related to less aggression in general. The DLPFC has been found to play an important role in response inhibition in non-social contexts, such as in no-go paradigms (Blasi et al., 2006; Durston et al., 2002), and our findings suggest a similar role for the DLPFC in inhibitory control in a social context. Together, these brain-behavior relations reveal a robust regulatory mechanism in middle childhood.

In line with the notion that middle childhood is characterized by a rapid development in cognitive control functioning (Zelazo & Carlson, 2012), we observed age effects on behavioral aggression and neural activation during the noise blast event. Behaviorally, older children were less aggressive and showed more differentiation in aggression following negative and positive feedback. On a neural level, effects were most pronounced in the positive feedback condition, such that older children showed more activation in DLPFC, VLPFC, AI and MPFC when responding to positive feedback. Prior studies also revealed a decrease in aggression from middle to late childhood (Chen et al., 2011), with largest decreases in responses following positive feedback (Achterberg et al., 2020). Sensitivity to social evaluation and the importance of social belonging increases in adolescence (Somerville, 2013), which might cause older children to more often refrain from forced aggression following positive feedback. On the other hand, inhibitory control functions are still developing during childhood, which might possibly explain why we found increases in neural activation during aggression following positive feedback, but not in neural activation during feedback itself. In an exploratory moderation analysis, we tested the interaction of these two processes in predicting aggressive outcomes. However, we did not find any moderating effects of inhibitory control on the relation between neural social sensitivity and subsequent aggression in childhood, as was previously reported in adults (Chester et al., 2014). Prior studies showed that inhibition is a multi-dimensional construct (Ridderinkhof et al., 2014) and theoretical reviews distinguished between stimulus-driven and intentional inhibition, which were associated with separable

neural networks (Filevich et al., 2012; Schel et al., 2014). Possibly, the current paradigm relied most on intentional inhibition processes. Given that cognitive control functions continue to develop and social sensitivity peaks in adolescence (Somerville, 2013), an interesting direction for future research would be to explore whether the interaction of these processes are predictive of aggression in adolescence (Lickley & Sebastian, 2018).

This is the first study investigating neural mechanisms of both social feedback evaluation and subsequent responses in middle childhood, using a well-validated experimental task (Achterberg et al., 2016, 2017). By including two large independent samples, we investigated the robustness and replicability of these mechanisms in middle childhood, which is still a relatively understudied phase in terms of neural development. Previously, it has been indicated that neuroimaging results often include false positive results (Eklund et al., 2016), possibly because of power issues (Button et al., 2013; Turner et al., 2018). Also, Elliott et al. (2020) showed that task-based fMRI often shows low test-retest reliability and may often not be suitable to test for individual differences in small samples. However, the large sample sizes in our study compensated for these issues and may have increased the chance of replicating meaningful brain-behavior relations.

There were some limitations that should be considered as well when interpreting the results. First, there were differences between the two samples, such as the date of testing (2015 versus 2019), the number of trials or possible familiarity effects with the study (in the replication sample), that may have contributed to differences in results between the samples, such as smaller condition effects on ROI activation in the replication sample. However, since one might not expect to exactly replicate prior findings, we used equivalence testing to test whether a result would have been meaningful in the original study (Lakens et al., 2018). Indeed, all non-replicable effects in the replication sample were non-equivalent to zero and therefore still considered meaningful. Thus, by and large our study did not show non-replicability of the original results, and revealed some robust behavioral and neural effects on a group level and in individual differences analyses. An interesting approach for future studies would be to use correspondence testing, in which difference tests and equivalence testing are combined into one framework (Steiner & Wong, 2018). Second, since participants were instructed to always send a noise blast, it is difficult to discover whether some children would rather have refrained from aggression, or would have acted prosocially, when receiving positive feedback. Children who show both self-protective as well as prosocial behaviors were previously found to show decreased externalizing behaviors over time (Dobbelaar et al., 2021). To further disentangle these possible motives and relations to developmental outcomes, future studies might include both an aggressive and prosocial response option, such as a noise blast measure that can

both be increased and decreased. In keeping with prior studies, we modeled the HRF of the noise blast event as the duration of the noise blast length (van de Groep et al., 2021, 2022). Adding a prosocial response option could also help in overcoming the issue that noise blast events for the three feedback conditions were modeled with different noise blast durations and might help shed light on the specific function of the DLPFC when responding to social feedback. Another possibility to keep the different feedback conditions comparable is to let participants indicate the intensity of a noise blast by selecting a specific punishment (see e.g., Krämer et al., 2007; Chester et al., 2018). Third, in our exploratory analyses we tested for age effects cross-sectionally on a relatively narrow age range. To fully investigate developmental processes, it is necessary to investigate changes in within-subjects behavior longitudinally during a broader developmental period. The period from childhood to adolescence might be specifically important in the context of aggression regulation, given the changes in emotional reactivity and cognitive control (Achterberg et al., 2020; Crone & Steinbeis, 2017; Somerville, 2013; Yoon et al., 2018). However, our findings on a relatively small age range already revealed developmental effects on aggression regulation, confirming the importance of middle childhood in social cognitive development.

In conclusion, our results highlight the importance of the affective salience network and prefrontal regions in social feedback processing and subsequent responses in middle childhood. This phase is marked by an increase in social experiences, during which aggressive responses following peer rejection may lead to a negative spiral of even more peer rejection (Lansford et al., 2010). Thus, it is a crucial period to learn to regulate aggressive impulses. Our findings point towards the DLPFC as a flexible regulatory mechanism in both emotion regulation and inhibitory behavior. Although we note developmental effects in these processes, our results reveal a core neural basis for social evaluation and aggression already in middle childhood. Together, these findings aid to our understanding of why some children are more prone to aggression than others.

# SUPPLEMENTARY METHODS

## Participants

At the start of L-CID, address information was obtained from municipalities registers and invitations were sent to families with same-sex twins who lived in the Western municipalities of the Netherlands. All participants were fluent in Dutch, had normal or corrected to normal vision and no intellectual disability or visual or hearing impairments that could hinder their performance on behavioral tasks (Euser et al., 2016).

In the MCC, 512 children (256 families) were included at T1. Since three of these children did not have usable behavioral data due to technical issues, our behavioral test sample consisted of 509 participants (mean age: 7.95 ± 0.67 years, age range: 7.02-9.68 years, 51% girls, see Figure S1). Twenty-seven participants did not participate in the MRI procedure, due to anxiety, MRI contra-indications, no parental consent for participation in the MRI procedure and technical problems with the MRI system. Four participants were excluded from MRI analyses because of anomalous findings in brain anatomy, and 89 participants were excluded due to head movement beyond 3 mm in two or more blocks of the SNAT (3 blocks in total). Finally, an additional four participants were excluded because of preprocessing errors. This resulted in an MRI test sample of 385 participants (mean age: 7.99 ± 0.68 years, age range: 7.02-9.68 years, 53% girls, see Figure S1).

In the ECC, of the 476 participants that were included at T1, 360 were still participating at T5. Four participants did not have behavioral data, due to lack of parental consent (n = 2), technical errors (n = 1) and lack of child consent (n = 1). Since we only included participants that had at least five valid trials for each feedback type (positive, neutral, negative), we additionally excluded two participants. Therefore, our final behavioral replication sample consisted of 354 participants (mean age: 8.00 ± 0.62 years, age range: 6.93-9.62, 53.1% girls, see Figure S1). Sixty-four participants did not participate in the MRI procedure, due to MRI contra-indications, anxiety and lack of parental consent. An additional fifteen participants did not finish the SNAT in the MRI scanner and were therefore excluded from the MRI analyses. Eighty participants were excluded due to head movement beyond 3 mm in one or more blocks of the SNAT (2 blocks in total). This resulted in an MRI replication sample of 195 participants (mean age: 8.08 ± 0.62 years, age range: 7.02 – 9.49, 57.4% girls, see Figure S1).

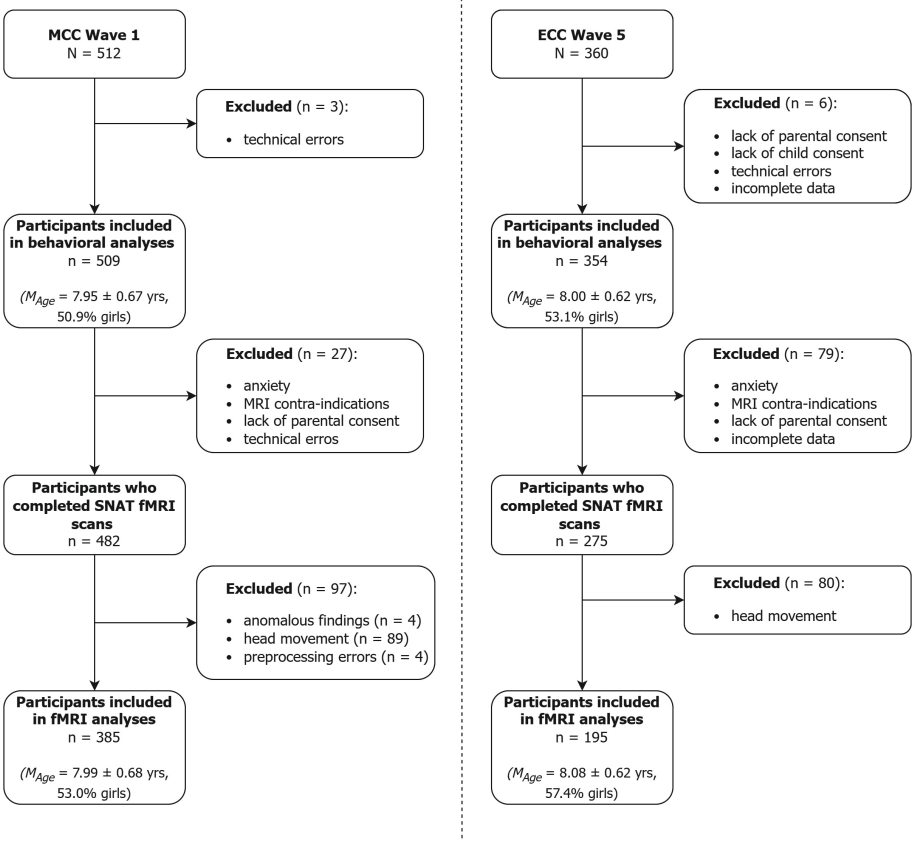


**Figure S1.** Flowchart of inclusion of participants in the middle childhood cohort (MCC, test sample) and early childhood cohort (ECC, replication sample).

# SUPPLEMENTARY RESULTS

## Neural activation during feedback

### *Whole-brain analyses*

**ECC Replication sample.** The "Positive > Negative" contrast resulted in neural activity in areas that were also reported in Achterberg et al. (2018), such as the lingual gyrus, left and right DLPFC and left OFC. In contrast to Achterberg et al. (2018), the reversed contrast, "Negative > Positive", did not show significant clusters of activation. Additionally, we investigated activation that was specific to negative and positive feedback by performing a conjunction analysis that combined the contrasts "Negative > Neutral" and "Positive > Neutral". Results showed increased activation in the left MPFC, ACC and occipital gyrus (Table S2).

### *Valence effects*

**ECC Replication sample.** Replicating prior findings, we observed a main effect of feedback condition in AI ($F(2,388.59) = 14.84$, $p < .001$), VLPFC ($F(2,388.46) = 6.93$, $p = .001$) and MPFC activation ($F(2,383.87) = 5.17$, $p = .006$; Figure 2). For the AI and MPFC, activation following neutral feedback was significantly decreased compared to negative feedback ($p \leq .01$) and positive feedback ($p \leq .04$). The VLPFC showed a similar pattern of activation, but the positive and neutral conditions did not differ in activation ($p = .07$), in contrast to prior findings in the test sample. Equivalence testing against raw equivalence bounds of -0.03 and 0.03 revealed that this difference was not statistically equivalent to zero, $t(192) = 1.48$, $p = .930$. Thus, we cannot reject the presence of a meaningful effect. Also in contrast to (Achterberg et al., 2020), no main effect was found for DLPFC activation, $F(2,382.42) = 1.78$, $p = .17$. Given that activation during positive feedback was significantly increased compared to neutral feedback in the test sample, but not in the replication sample ($p = .558$), we used equivalence testing against raw equivalence bounds of -0.06 and 0.06. Results revealed that the effect in the replication sample was not statistically equivalent to zero, $t(191) = -0.25$, $p = .400$. Post-hoc statistics are presented in Table S3.

## Neural activation during noise blast

### *Whole-brain analyses*

**MCC Test sample.** We explored feedback-specific effects on whole-brain activation during the noise blast event by investigating the six contrasts including the feedback conditions. The "PositiveNoise > NegativeNoise" contrast revealed a wide network of activation, including the medial and dorsal frontal regions and occipital gyrus.

The reversed contrast did not result in significant activation. For "PositiveNoise > NeutralNoise", we observed activation in the lingual gyrus, MPFC and inferior frontal gyrus. Again, the reversed contrast did not result in significant clusters of activation. For "NegativeNoise > NeutralNoise", we found increased activation in the MPFC, OFC and ACC. The reversed contrast, "NeutralNoise > NegativeNoise" showed a network of activation including the right superior frontal gyrus, supplementary motor area and occipital gyrus (see Table S4 and Figure S3).

**ECC Replication sample.** Similar to the test sample, the "PositiveNoise > NegativeNoise" contrast resulted in a wide network of activation, including the medial and dorsal frontal regions and occipital gyrus, whereas the reversed contrast did not result in significant clusters of activation. For "PositiveNoise > NeutralNoise", we observed activation in the lingual gyrus, MPFC and inferior frontal gyrus, whereas the reversed contrast did again not result in significant clusters of activation. For "NegativeNoise > NeutralNoise", we found activation in the MPFC and OFC. The reversed contrast, "NeutralNoise > NegativeNoise" resulted again in a wide network of activation, including the right superior frontal gyrus, supplementary motor area and parietal lobe (see Table S4).



### *Valence effects*

**Nucleus accumbens activation.** In an additional analysis, we tested for valence effects on nucleus accumbens (NAcc) activation during the noise blast, based on prior work that showed that aggression can also be rewarding (Chester & DeWall, 2016). We used the left and right nucleus accumbens from the structural FSL atlas as regions of interest. Left and right NAcc activity were significantly correlated (test sample: PositiveNoise: $r = .837$, $p < .001$; NeutralNoise: $r = .695$, $p < .001$, NegativeNoise: $r = .899$, $p < .001$; replication sample: PositiveNoise: $r = .887$, $p < .001$; NeutralNoise: $r = .877$, $p < .001$; NegativeNoise: $r = .688$, $p < .001$), so we took the left and right NAcc together into one measure of NAcc activity. We did not find feedback condition effects on NAcc activation during the noise blast in the test sample, $F(2,761.56) = 0.223$, $p = .800$, nor in the replication sample, $F(2,375.98) = 2.610$, $p = .075$ (see Figure S2). These findings show that NAcc activation does not differ between different types of social feedback in middle childhood.

## Moderation inhibitory control

To test whether inhibitory control moderated the relation between MPFC/AI activation and noise blast duration, we performed two moderation analyses. Noise blast duration was not predicted by AI activation during feedback, $b = 2.66$, $t = 0.07$, $p = .948$, 95% CI [-77.90, 83.23], nor by inhibitory control, $b = -90.46$, $t = -1.33$, $p = .185$, 95%

CI [-224.34, 43.43]. The interaction between AI activation and inhibitory control was also non-significant, $b$ = -47.92, $t$ = -0.62, $p$ = .535, 95% CI [-199.53, 103.70], indicating no significant moderation effect.

For MPFC activation, similar results were found: noise blast duration was not predicted by MPFC activation during feedback, $b$ = -56.53 $t$ = -1.15, $p$ = .251, 95% CI [-153.15, 40.10], nor by inhibitory control, $b$ = -72.14, $t$ = -1.05, $p$ = 0.292, 95% CI [-206.56, 62.29]. The interaction effect between MPFC activation and inhibitory control was non-significant as well, $b$ = -31.10, $t$ = -0.35, $p$ = .724, 95% CI [-203.90, 141.69]. Thus, we did not find any moderation effects of inhibitory control.

## Sensitivity analyses

### *Sensitivity analyses on participants with and without small motion spikes*

To test whether the inclusion of small spikes (i.e., motion between 0.9 and 3mm in any volume) might have affected our results, we checked whether condition effects differed for participants with small spikes (N = 112) and without small spikes (N = 83) in the replication sample. We performed repeated measures ANOVAs (Greenhouse-Geisser corrected in case of violation of the assumption of sphericity) with the occurrence of motion spikes added as between-subject variable (0 = no small spikes, 1 = small spikes). For neural activation in the DLPFC, VLPFC, MPFC and insula during the feedback event, there were no significant effects of motion group (all $p \geq .117$), nor any interaction effects of motion group * feedback condition (all $p \geq .124$). Similarly, for neural activation in the DLPFC, VLPFC, MPFC and OFC during the noise blast event, there were also no main effects of motion group (all $p \geq .394$), nor any interaction effects (all $p \geq .263$). We did find a significant main effect of motion group for insula activation during the noise blast, $F(1,186) = 5.701$, $p = .018$): participants with small spikes showed more activation in the insula compared to participants without small spikes. However, a follow-up analysis revealed that feedback conditions effects on insula activation were similar in both groups (group without spikes: $F(2,164) = 17.542$, $p < .001$; group with spikes: $F(2,208) = 14.729$, $p < .001$), and comparable to the effects found in the whole group. In both samples, insula activation following negative feedback was significantly lower than following neutral feedback (both $p$'s $< .001$) and following positive feedback (both $p$'s $< .001$). Thus, it seems that our results were minimally affected by motion.

### *Sensitivity analyses moderation effects by inhibitory control*

As a sensitivity analysis, we tested whether there were moderation effects of inhibitory control on the relation between neural sensitivity to social feedback and reactive aggression using an experimental behavioral measure instead of the questionnaire subscale Inhibitory Control. We used data of the Stop Signal task as behavioral measure of inhibitory control. The Stop Signal task (Williams et al., 1999) is a validated measure of inhibition of an already initiated response (Eagle et al., 2008). It consisted of four blocks of 50 trials. In this task, participants viewed colored arrows. For green arrows, participants had to respond as fast as possible by pressing the right or left button on a keyboard for right- and left pointed arrows respectively (Go trials, 75% of total trials). For red arrows, they were instructed to not push the button (NoGo trials, 25% of total trials). The time between the start of a trial (green arrow) and the stop signal (red arrow), also referred to as the Stop Signal Delay (SSD), was adjusted during the task based on the participant's performance. The SSD was increased (with 50ms) when participants successfully inhibited their response and decreased (with 50ms) when participants failed to inhibit their response. We used the Stop Signal Reaction Time (SSRT) as measure of inhibitory control. The SSRT was calculated as: median RT on Go trials – mean SSD. Higher SSRT scores represent less response inhibition. The Stop Signal task was administered in the same manner in the test and replication samples.

We performed the moderation analyses in R using the PROCESS macro (version 3.5) and controlled for sample effects by adding cohort as covariate. We performed two sensitivity moderation analyses. In the first analysis, we used AI activation during negative versus neutral feedback as independent variable, SSRT as moderator and noise blast duration (negative – neutral) as dependent variable. In the second analysis, we used MPFC activation during negative versus neutral feedback as independent variable.

For the moderation analysis with AI activation as independent variable (n = 561), noise blast duration was not predicted by AI activation during feedback, $b = 5.93$, $t = 0.15$, $p = .883$, 95%CI [-.73.10, 84.95], nor by SSRT, $b = -0.29$, $t = -0.62$, $p = .535$, 95% CI [-1.23, 0.64]. The interaction between AI activation and SSRT was also non-significant, $b = 0.85$, $t = 1.43$, $p = .154$, 95%CI [-0.32, 2.01], indicating no moderation effect by SSRT.

For the moderation analysis with MPFC activation as independent variable (n = 557), noise blast duration was not predicted by MPFC activation, $b = -61.40$, $t = -1.28$, $p = .201$, 95%CI [-155.55, 32.76], nor by SRRT, $b = -0.39$, $t = -0.82$, $p = .416$, 95%CI [-1.32, 0.54]. The interaction effect between MPFC activation and SSRT was also non-significant, $b = 0.44$, $t = 0.67$, $p = .503$, 95%CI [-0.85, 1.72], again indicating no moderation effect by SSRT.

### *Sensitivity analysis on the VIPP control group*

Since the L-CID study included a randomized controlled trial, approximately 40% of the families received the Video-feedback Intervention to Promote Positive Parenting and Sensitive Discipline in Twin Families (VIPP-Twins, Euser et al., 2016) between T2 and T3 (Crone et al., 2020). This intervention consisted of five biweekly home sessions during which parents received feedback on interactions with their children. The control condition consisted of six phone calls about the general development of the twins. To check whether this intervention could have influenced our results, we repeated our replication analyses (ECC replication sample) in the control group of the VIPP-Twins. For the behavioral analyses, this sample consisted of 213 participants (60% of behavioral sample, mean age: 8.03 ± 0.62 years, age range: 6.93 – 9.62, 52.6% girls). For the fMRI analyses, this sample consisted of 121 participants (62% of MRI sample, mean age: 8.06 ± 0.60 years, age range: 7.02 – 9.31, 53.7% girls). Note that the analyses in the MCC test sample could not have been affected by the intervention, since these data were collected at T1, prior to the administration of VIPP-Twins.

In the control group, all main effects of feedback condition were similar to the findings in the complete replication sample. We found main effects of feedback condition on noise blast duration, $F(2,426) = 328.63$, $p < .001$. Negative feedback resulted in the longest noise blast duration, followed by neutral feedback and positive feedback (all pairwise comparisons: $p < .001$). The main effect of sex, however, was no longer significant, $F(2, 107.1) = 2.91$, $p = .091$. In line with findings in the complete sample, we observed main effects of condition on ROI activation during feedback in the VLPFC, $F(2,241.23) = 8.04$, $p < .001$, insula, $F(2, 240.45) = 15.04$, $p < .001$ and MPFC, $F(2, 241.63) = 6.68$, $p = .001$, but not in the DLPFC, $F(2, 239.55) = 0.51$, $p = .600$. Pairwise comparisons revealed similar patterns of activation as in the complete sample, with the exception that MPFC activation in the positive and neutral conditions did no longer significantly differ ($p = .099$). For ROI activation during the retaliation event, we observed main effects of condition in all five ROIs (MPFC: $F(2, 238.86) = 5.05$, $p = .007$; OFC: $F(2,238.58) = 6.24$, $p = .002$; DLPFC: $F(2,236.05) = 18.94$, $p < .001$; Insula: $F(2,235.19) = 24.20$, $p < .001$, VLPFC: $F(2,234.94) = 15.05$, $p < .001$). All pairwise comparisons revealed similar patterns of activations as in the complete sample, with the exception that the difference in OFC activation in the positive and neutral condition was now significant ($p = .011$).

### *Sensitivity analysis with IQ as covariate*

In the analyses in Achterberg et al. (2020), IQ was added as covariate to the analyses. In the test sample, IQ was estimated at T1 with the subsets "similarities" and "block design" from the Wechsler Intelligence Scale for Children (WISC, 3rd edition), giving an estimation of both performal and verbal IQ. In the replication sample, however, IQ was estimated at T4 (at 5-7 years of age) with the subset "picture completion" from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI, 3rd edition), which only gave an estimation of performal IQ. Also, 23 participants in the replication sample did not have IQ data. Because of these differences between the samples, we decided not to add IQ in the main analyses but rather to check for possible effects in sensitivity analyses. Thus, the linear mixed models of the sensitivity analyses were defined in R as: Noise blast duration / ROI ~ Condition × Sex + Condition × IQ + (1|ChildID) + (1|FamilyID).



First, we checked for IQ effects on noise blast duration in the replication sample. We did not find any IQ effects on noise blast duration, $p = .969$. Controlled for IQ, we still observed a main effect of feedback condition, $p < .001$, and sex, $p = .005$.

Second, we checked for IQ effects on ROI activation during feedback in the replication sample. We found a main effect of IQ on MPFC activation, $F(1,179.09) = 4.077$, $p = .045$, indicating that children with a higher IQ showed less MPFC activation. We did not observe any other main effects of IQ on ROI activation (all $p \geq .439$), nor any interaction effects (all $p \geq .405$). All main effects of condition were similar with and without controlling for IQ.

Third, we checked for IQ effects on ROI activation during the noise blast event in the test sample. We found a main effect of IQ on MPFC and OFC activation (both $p \leq .040$), indicating more activation for children with a higher IQ. Furthermore, we found interaction effects of IQ and feedback condition for MPFC, DLPFC, insula and VLPFC activation (all $p \leq .039$), indicating more activation during retaliation following positive feedback for children with a higher IQ. Controlled for IQ, all main effects of feedback condition and sex remained similar to the results without IQ as covariate.

Finally, we checked for IQ effects on ROI activation during the noise blast event in the replication sample. We did not find any main effects of IQ on ROI activation (all $p \geq .438$), nor any interaction effects of IQ and feedback condition (all $p \geq .296$). All main effects of condition were similar with and without controlling for IQ.

**Table S1.** Overview of studies using the Social Network Aggression Task.

| Paper | Participants | SNAT event | Study type | Main findings |
|---|---|---|---|---|
| Achterberg et al. (2016) | 18-27y<br>N=30 | fMRI: feedback event | Cross-sectional | • Feedback: positive and negative (vs neutral) → MPFC and insula activity<br>• Brain-behavior: right DLPFC activation during negative (vs neutral) feedback → shorter noise blast duration following negative feedback |
| Achterberg et al. (2017) | 7-9y<br>N=19; N= 28; N=27 | fMRI: feedback event | Cross-sectional; Replication | • Feedback:<br>o Negative (vs positive) → amygdala, MPFC, left insula activity<br>o Negative (vs neutral) → AI, MPFC/ACCg activity |
| Achterberg et al. (2018) | 7-9y<br>N=385 (fMRI)<br>*Test sample* | fMRI: feedback event | Cross-sectional | • Feedback:<br>o Positive and negative (vs neutral) → AI and ACCg activity<br>o Negative (vs neutral) → MPFC, and IFC activity<br>o Positive (vs neutral) → caudate, SMA, DLPFC activity<br>• Brain-behavior: SMA/DLPFC activation during positive (vs negative) feedback → more aggression following negative feedback |
| Achterberg et al. (2020) | 7-9y & 9-11y<br>N=360<br>*Test sample* | fMRI: feedback event | Longitudinal | • Feedback: Insula, DMPFC and DLPFC activation increased over time<br>• Brain-behavior: increased DLPFC activation during negative feedback from middle to late childhood → decreased aggression over time |
| Dobbelaar et al. (2021) | 7-9y<br>N=496<br>*Test sample* | Behavioral | Cross-sectional | • Behavioral aggression combined with prosocial behavior → less externalizing behavior one year later |
| van de Groep et al. (2021) | 18-30y<br>N=40 | fMRI: feedback event + noise event | Cross-sectional | • Feedback: positive and negative (vs neutral)→ increased insula and ACC activity<br>• Noise blast: positive (vs neutral/negative) → increased DLPFC activity<br>• Brain-behavior: DLPFC activation during negative (vs positive) feedback → shorter noise blast duration following negative feedback |
| van de Groep et al. (2022) | 18-30y<br>N=54 (antisocial history); N=39 (control) | fMRI: feedback event + noise event | Cross-sectional | • Feedback:<br>o Positive (vs neutral) → increased insula and ACC activity<br>o Insula activation: higher for antisocial trajectories (vs control)<br>o DLPFC activation: higher for desisting (vs persisting/control) antisocial trajectory<br>• Noise blast: positive (vs neutral/negative) feedback → increased DLPFC and ACC activity<br>• Brain-behavior: ACC activation during noise blasts to positive feedback → shorter noise blast following positive feedback<br>o Strongest for desisting (vs persisting/control) antisocial trajectory |

**Table S2.** MNI coordinates for local maxima activated for the whole-brain contrasts of the feedback event in the replication sample.

| | | | | MNI coordinates | | |
|---|---|---|---|---|---|---|
| **Anatomical region** | **Voxels** | ***p*FWEcc** | ***T*** | ***x*** | ***y*** | **Z** |
| ***Positive* > *Negative*** | | | | | | |
| Right lingual gyrus | 1252 | < .001 | 8.19 | 6 | -73 | -2 |
| Left lingual gyrus | | | 5.17 | -15 | -88 | -11 |
| Left fusiform gyrus | | | 4.57 | -24 | -73 | -8 |
| Left middle frontal gyrus | 907 | < .001 | 4.46 | -30 | 38 | 49 |
| | | | 4.45 | -27 | 20 | 61 |
| Left middle orbital gyrus | | | 4.17 | -42 | 59 | -5 |
| Right superior frontal gyrus | 554 | < .001 | 4.15 | 24 | 14 | 61 |
| Right middle frontal gyrus | | | 4.13 | 33 | 38 | 46 |
| | | | 3.74 | 45 | 38 | 34 |
| Right inferior parietal lobule | 239 | .024 | 3.92 | 57 | -46 | 46 |
| | | | 3.55 | 54 | -55 | 52 |
| Right supramarginal gyrus | | | 3.14 | 57 | -31 | 49 |
| ***Negative* > *Positive*** | | | | | | |
| n.s. | | | | | | |
| ***Conjunction Negative* > *Neutral* and *Positive* > *Neutral*** | | | | | | |
| Right middle occipital gyrus | 2633 | < .001 | 6.43 | 48 | -79 | 4 |
| Left inferior occipital gyrus | | | 6.29 | -45 | -73 | -5 |
| Left middle occipital gyrus | | | 6.06 | -48 | -79 | 4 |
| Left anterior cingulate cortex | 285 | .011 | 3.46 | -3 | 50 | 4 |
| Right anterior cingulate cortex | | | 3.27 | 9 | 47 | 16 |
| Left superior medial cortex | | | 3.27 | -3 | 56 | 22 |

*Note.* Results were FWE cluster corrected ($p_{FWEcc} < .05$), with a primary voxel-wise threshold of $p < .005$.

**Table S3.** Statistics on the post-hoc test of the linear mixed effect model with brain activation during the feedback event as dependent variable in the replication sample.

| | Estimated difference score | *t*-value | df | *p*-value | 95% CI |
|---|---|---|---|---|---|
| **MPFC** | | | | | |
| *Positive - Neutral* | 0.13 | 2.53 | 392 | .035 | [0.01, 0.24] |
| *Negative – Neutral* | 0.15 | 2.96 | 392 | .010 | [0.03, 0.26] |
| *Negative – Positive* | 0.02 | 0.43 | 392 | 1.000 | [-0.10, 0.14] |
| **DLPFC** | | | | | |
| *Positive - Neutral* | 0.05 | 1.33 | 391 | .558 | [-0.05, 0.16] |
| *Negative – Neutral* | -0.02 | -0.49 | 390 | 1.000 | [-0.12, 0.08] |
| *Negative – Positive* | -0.08 | -1.81 | 392 | .212 | [-0.18, 0.02] |
| **VLPFC** | | | | | |
| *Positive - Neutral* | 0.09 | 2.30 | 391 | .066 | [0.00, 0.19] |
| *Negative – Neutral* | 0.15 | 3.66 | 392 | <.001 | [0.05, 0.24] |
| *Negative – Positive* | 0.05 | 1.37 | 391 | .513 | [-0.04, 0.15] |
| **Insula** | | | | | |
| *Positive - Neutral* | 0.18 | 3.82 | 392 | <.001 | [0.06, 0.29] |
| *Negative – Neutral* | 0.24 | 5.24 | 393 | <.001 | [0.13, 0.35] |
| *Negative – Positive* | 0.07 | 1.43 | 393 | .464 | [-0.05, 0.18] |

*Note.* Output is based on least-square means with Kenward-Roger corrected degrees of freedom and Bonferroni-adjusted *p*-values. Abbreviations: MPFC = medial prefrontal cortex, DLPFC = dorsolateral prefrontal cortex, VLPFC: ventrolateral prefrontal cortex.

**Table S4.** MNI coordinates for local maxima activated for the whole-brain contrasts of the noise blast event in the test and replication sample.

| | | | | MNI coordinates | | |
|---|---|---|---|---|---|---|
| **Anatomical region** | **Voxels** | ***p*FWEcc** | ***T*** | ***x*** | ***y*** | **Z** |
| ***Test sample: PositiveNoise > NegativeNoise*** | | | | | | |
| Right calcarine gyrus | 29769 | < .001 | 10.97 | 21 | -91 | -2 |
| Left fusiform gyrus | | | 10.32 | -30 | -67 | -14 |
| Right inferior occipital gyrus | | | 10.31 | 33 | -85 | -2 |
| ***Test sample: NegativeNoise > PositiveNoise*** | | | | | | |
| n.s. | | | | | | |
| ***Test sample: PositiveNoise > NeutralNoise*** | | | | | | |
| Right lingual gyrus | 6459 | < .001 | 8.42 | 21 | -91 | -5 |
| Left fusiform gyrus | | | 8.25 | -27 | -67 | -14 |
| | | | 8.00 | -27 | -79 | -11 |
| Left supramarginal gyrus | 8306 | < .001 | 8.13 | -66 | -55 | 28 |
| Right superior medial gyrus | | | 7.20 | 12 | 41 | 55 |
| Right inferior frontal gyrus | | | 7.03 | 54 | 20 | -2 |
| ***Test sample: NeutralNoise > PositiveNoise*** | | | | | | |
| n.s. | | | | | | |
| ***Test sample: NegativeNoise > NeutralNoise*** | | | | | | |
| Left mid orbital gyrus | 1005 | < .001 | 6.45 | 0 | 53 | -11 |
| | | | 6.42 | -3 | 65 | -11 |
| Left superior medial gyrus | | | 4.58 | -6 | 65 | 13 |
| Left angular gyrus | 311 | 0.042 | 5.33 | -60 | -67 | 22 |
| Inferior parietal cortex | | | 4.47 | -48 | -76 | 34 |
| Left middle temporal gyrus | | | 4.04 | -69 | -46 | 4 |
| ***Test sample: NeutralNoise > NegativeNoise*** | | | | | | |
| Right superior frontal gyrus | 27372 | < .001 | 10.66 | 27 | -1 | 55 |
| Right SMA | | | 10.58 | 3 | 14 | 52 |
| Right middle occipital gyrus | | | 9.86 | 30 | -70 | 37 |
| ***Replication sample: PositiveNoise > NegativeNoise*** | | | | | | |
| Right lingual gyrus | 20790 | < .001 | 9.37 | 24 | -88 | -5 |
| Right inferior occipital gyrus | | | 8.58 | 42 | -64 | -11 |

**Table S4.** (Continued)

| | | | | MNI coordinates | | |
|---|---|---|---|---|---|---|
| **Anatomical region** | **Voxels** | ***p*FWEcc** | ***T*** | ***x*** | ***y*** | **Z** |
| Right middle occipital gyrus | | | 8.49 | 36 | -91 | 4 |
| ***Replication sample: NegativeNoise > PositiveNoise*** | | | | | | |
| n.s. | | | | | | |
| ***Replication sample: PositiveNoise > NeutralNoise*** | | | | | | |
| Right middle occipital gyrus | 959 | < .001 | 6.05 | 39 | -91 | 7 |
| Right fusiform gyrus | | | 4.50 | 33 | -64 | -8 |
| Right lingual gyrus | | | 4.48 | 24 | -88 | -5 |
| Right angular gyrus | 293 | .013 | 4.74 | 57 | -52 | 37 |
| | | | 3.64 | 63 | -61 | 25 |
| Left calcarine gyrus | 1236 | < .001 | 4.48 | -12 | -82 | 4 |
| Left fusiform gyrus | | | 4.48 | -24 | -73 | -8 |
| Left middle occipital gyrus | | | 4.34 | -30 | -94 | 4 |
| Left supramarginal gyrus | 341 | .006 | 3.94 | -60 | -52 | 37 |
| Left middle temporal gyrus | | | 3.83 | -54 | -61 | 22 |
| Left angular gyrus | | | 3.37 | -42 | -55 | 22 |
| ***Replication sample: NeutralNoise > PositiveNoise*** | | | | | | |
| n.s. | | | | | | |
| ***Replication sample: NegativeNoise > NeutralNoise*** | | | | | | |
| Left mid orbital gyrus | 264 | .021 | 4.55 | 0 | 65 | -11 |
| | | | 2.78 | -9 | 44 | -5 |
| Left superior medial gyrus | | | 4.36 | -3 | 71 | 1 |
| ***Replication sample: NeutralNoise > NegativeNoise*** | | | | | | |
| Right inferior parietal lobe | 18180 | < .001 | 8.20 | 51 | -40 | 52 |
| Right SMA | | | 8.12 | 6 | 17 | 49 |
| Right superior frontal gyrus | | | 7.90 | 27 | 5 | 58 |

*Note.* Results were FWE cluster corrected ($p_{FWEcc} < .05$), with a primary voxel-wise threshold of $p < .005$. Abbreviations: SMA = Supplementary Motor Area.

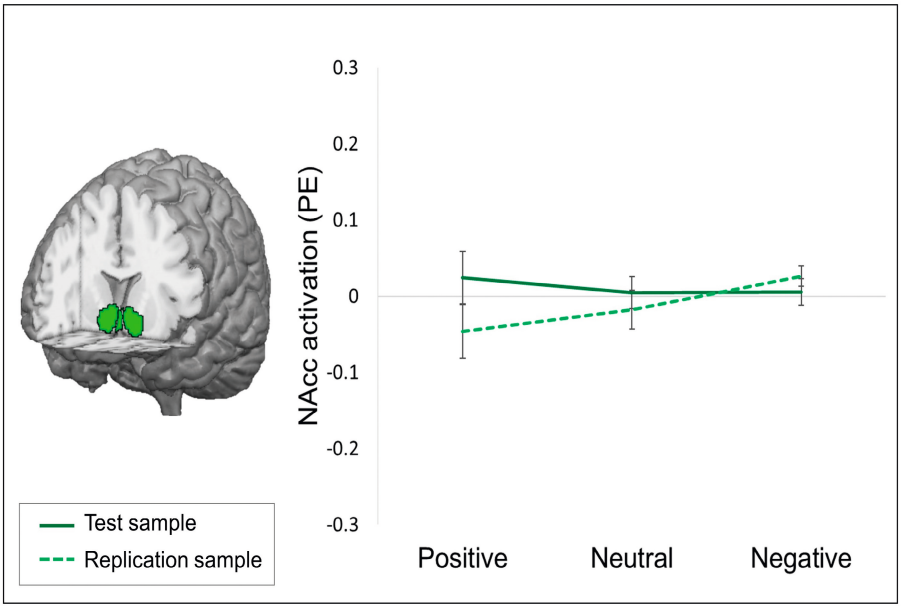


**Figure S2.** Neural activation (parameter estimates) during the noise blast event for each feedback condition in the nucleus accumbens (NAcc) in the test sample (solid lines) and replication sample (dotted lines). Error bars represent standard errors.

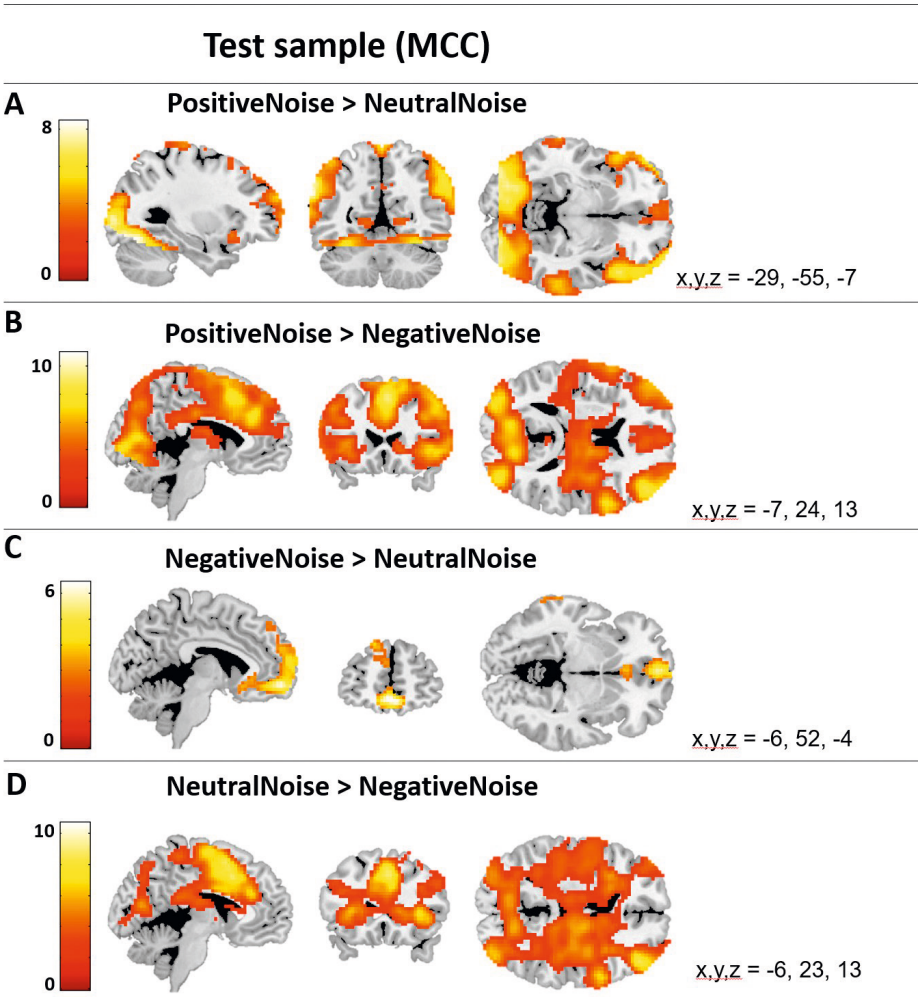


**Figure S3.** Whole-brain activation during the noise blast event in the test sample for (A) the contrast PositiveNoise > NeutralNoise; (B) the contrast PositiveNoise > NegativeNoise; (C) the contrast NegativeNoise > NeutralNoise; (D) the contrast NeutralNoise > NegativeNoise.

**Table S5.** Statistics on the post-hoc test of the linear mixed effect model with brain activation during the noise blast event as dependent variable.

| | Estimated difference score | *t*-value | df | *p*-value | 95% CI |
|---|---|---|---|---|---|
| **MPFC** | | | | | |
| *Positive - Neutral* | | | | | |
| Test sample | 0.17 | 5.22 | 764 | <.001 | [0.09, 0.25] |
| Replication sample | 0.20 | 3.45 | 392 | .002 | [0.06, 0.35] |
| *Negative – Neutral* | | | | | |
| Test sample | 0.10 | 2.90 | 763 | .012 | [0.01, 0.18] |
| Replication sample | 0.11 | 1.88 | 390 | .181 | [-0.03, 0.25] |
| *Negative – Positive* | | | | | |
| Test sample | -0.08 | -2.32 | 766 | .063 | [-0.16, 0.00] |
| Replication sample | -0.09 | -1.57 | 391 | .349 | [-0.23, 0.05] |
| **OFC** | | | | | |
| *Positive - Neutral* | | | | | |
| Test sample | 0.16 | 3.33 | 767 | .003 | [0.04, 0.28] |
| Replication sample | 0.14 | 1.86 | 389 | .191 | [-0.04, 0.32] |
| *Negative – Neutral* | | | | | |
| Test sample | 0.23 | 4.68 | 764 | <.001 | [0.11, 0.34] |
| Replication sample | 0.27 | 3.64 | 389 | <.001 | [0.09, 0.46] |
| *Negative – Positive* | | | | | |
| Test sample | 0.06 | 1.35 | 765 | .534 | [-0.05, 0.18] |
| Replication sample | 0.13 | 1.78 | 389 | .228 | [-0.05, 0.32] |
| **DLPFC** | | | | | |
| *Positive - Neutral* | | | | | |
| Test sample | 0.10 | 3.75 | 763 | <.001 | [0.04, 0.17] |
| Replication sample | 0.01 | 0.35 | 390 | 1 | [-0.08, 0.10] |
| *Negative – Neutral* | | | | | |
| Test sample | -0.15 | -5.55 | 766 | <.001 | [-0.22, -0.09] |
| Replication sample | -0.22 | -6.04 | 389 | <.001 | [-0.31, -0.13] |
| *Negative – Positive* | | | | | |
| Test sample | -0.25 | -9.28 | 766 | <.001 | [-0.32, -0.19] |
| Replication sample | -0.24 | -6.35 | 391 | <.001 | [-0.33, -0.15] |

**Table S5.** (Continued)

| | Estimated difference score | *t*-value | df | *p*-value | 95% CI |
|---|---|---|---|---|---|
| **VLPFC** | | | | | |
| *Positive - Neutral* | | | | | |
| Test sample | 0.12 | 4.65 | 761 | <.001 | [0.06, 0.19] |
| Replication sample | 0.10 | 2.78 | 389 | .017 | [0.01, 0.19] |
| *Negative – Neutral* | | | | | |
| Test sample | -0.08 | -2.99 | 762 | .009 | [-0.14, -0.02] |
| Replication sample | -0.13 | -3.61 | 388 | .001 | [-0.21, -0.04] |
| *Negative – Positive* | | | | | |
| Test sample | -0.20 | -7.63 | 763 | <.001 | [-0.27, -0.14] |
| Replication sample | -0.23 | -6.39 | 389 | <.001 | [-0.31, -0.14] |
| | | | | | |
| **Insula** | | | | | |
| *Positive - Neutral* | | | | | |
| Test sample | 0.07 | 2.05 | 763 | .123 | [-0.01, 0.16] |
| Replication sample | 0.04 | 1.17 | 390 | .730 | [-0.04, 0.12] |
| *Negative – Neutral* | | | | | |
| Test sample | -0.16 | -4.55 | 762 | <.001 | [-0.24, -0.08] |
| Replication sample | -0.20 | -5.91 | 388 | <.001 | [-0.28, -0.12] |
| *Negative – Positive* | | | | | |
| Test sample | -0.23 | -6.59 | 764 | <.001 | [-0.32, -0.15] |
| Replication sample | -0.24 | -7.08 | 389 | <.001 | [-0.32, -0.16] |

*Note.* Output is based on least-square means with Kenward-Roger corrected degrees of freedom and Bonferroni-adjusted *p*-values. Abbreviations: MPFC = medial prefrontal cortex, OFC = orbitofrontal cortex, DLPFC = dorsolateral prefrontal cortex, VLPFC: ventrolateral prefrontal cortex.

**Table S6.** Statistics on the main effect of condition in the linear mixed effect model with a random slope for condition and with brain activation during the noise blast event as dependent variable.

| | degrees of freedom | *F*-value | *p*-value | $\eta^2_p$ |
|---|---|---|---|---|
| **MPFC** | | | | |
| Test sample | 2, 277.19 | 12.83 | <.001 | 0.08 |
| Replication sample | 2, 157.47 | 5.24 | .006 | 0.06 |
| **OFC** | | | | |
| Test sample | 2, 314.69 | 14.48 | <.001 | 0.08 |
| Replication sample | 2, 147.19 | 8.71 | <.001 | 0.11 |
| **DLPFC** | | | | |
| Test sample | 2, 269.61 | 30.49 | <.001 | 0.18 |
| Replication sample | 2, 148.51 | 32.10 | <.001 | 0.30 |
| **VLPFC** | | | | |
| Test sample | 2, 262.50 | 18.61 | <.001 | 0.12 |
| Replication sample | 2, 153.81 | 14.79 | <.001 | 0.16 |
| **Insula** | | | | |
| Test sample | 2, 307.41 | 18.93 | <.001 | 0.11 |
| Replication sample | 2, 147.70 | 27.19 | <.001 | 0.27 |

*Note.* Output is based on type III ANOVA's using Satterthwaite's method. Abbreviations: MPFC = medial prefrontal cortex, OFC = orbitofrontal cortex, DLPFC = dorsolateral prefrontal cortex, VLPFC: ventrolateral prefrontal cortex.

**Table S7.** Statistics on the post-hoc test of the linear mixed effect model with a random slope for condition and with brain activation during the noise blast event as dependent variable.

| | Estimated difference score | *t*-value | df | *p*-value | 95% CI | Equivalence testing |
|---|---|---|---|---|---|---|
| **MPFC** | | | | | | |
| *Positive - Neutral* | | | | | | |
| Test sample | 0.17 | 4.66 | 220 | <.001 | [0.08, 0.27] | |
| Replication sample | 0.19 | 2.68 | 124 | .025 | [0.02, 0.37] | |
| *Negative – Neutral* | | | | | | |
| Test sample | 0.10 | 3.19 | 213 | .005 | [0.02, 0.17] | $p = .997$ |
| Replication sample | 0.11 | 2.32 | 116 | .066 | [-0.01, 0.23] | |
| *Negative – Positive* | | | | | | |
| Test sample | -0.08 | -1.91 | 221 | .173 | [-0.18, 0.02] | |
| Replication sample | -0.08 | -1.05 | 125 | .881 | [-0.27, 0.11] | |
| **OFC** | | | | | | |
| *Positive – Neutral* | | | | | | |
| Test sample | 0.17 | 3.12 | 218 | .006 | [0.04, 0.30] | $p = .890$ |
| Replication sample | 0.12 | 1.27 | 125 | .624 | [-0.11, 0.35] | |
| *Negative – Neutral* | | | | | | |
| Test sample | 0.22 | 5.09 | 210 | <.001 | [0.11, 0.33] | |
| Replication sample | 0.28 | 4.13 | 120 | <.001 | [0.12, 0.44] | |
| *Negative – Positive* | | | | | | |
| Test sample | 0.06 | 0.94 | 211 | 1 | [-0.09, 0.20] | |
| Replication sample | 0.16 | 1.64 | 125 | .131 | [-0.08, 0.39] | |
| **DLPFC** | | | | | | |
| *Positive - Neutral* | | | | | | |
| Test sample | 0.10 | 3.25 | 222 | .004 | [0.03, 0.18] | $p = .302$ |
| Replication sample | 0.02 | 0.42 | 128 | 1 | [-0.10, 0.14] | |
| *Negative – Neutral* | | | | | | |
| Test sample | -0.15 | -6.41 | 194 | <.001 | [-0.21, -0.09] | |

**Table S7.** (Continued)

| | Estimated difference score | *t*-value | df | *p*-value | 95% CI | Equivalence testing |
|---|---|---|---|---|---|---|
| Replication sample | -0.23 | -7.40 | 121 | <.001 | [-0.30, -0.15] | |
| *Negative – Positive* | | | | | | |
| Test sample | -0.25 | -7.48 | 181 | <.001 | [-0.34, -0.17] | |
| Replication sample | -0.25 | -4.93 | 128 | <.001 | [-0.37, -0.13] | |
| **VLPFC** | | | | | | |
| *Positive - Neutral* | | | | | | |
| Test sample | 0.13 | 4.15 | 222 | <.001 | [0.06, 0.21] | |
| Replication sample | 0.10 | 2.54 | 124 | .037 | [0.00, 0.20] | |
| *Negative – Neutral* | | | | | | |
| Test sample | -0.08 | -3.36 | 215 | .003 | [-0.14, -0.03] | |
| Replication sample | -0.13 | -4.15 | 116 | <.001 | [-0.20, -0.05] | |
| *Negative – Positive* | | | | | | |
| Test sample | -0.21 | -6.05 | 224 | <.001 | [-0.30, -0.13] | |
| Replication sample | -0.23 | -5.049 | 126 | <.001 | [-0.34, -0.12] | |
| **Insula** | | | | | | |
| *Positive - Neutral* | | | | | | |
| Test sample | 0.08 | 1.92 | 212 | .171 | [-0.02, 0.17] | |
| Replication sample | 0.03 | 0.82 | 127 | 1 | [-0.06, 0.13] | |
| *Negative – Neutral* | | | | | | |
| Test sample | -0.16 | -5.06 | 226 | <.001 | [-0.24, -0.08] | |
| Replication sample | -0.20 | -6.66 | 119 | <.001 | [-0.27, -0.13] | |
| *Negative – Positive* | | | | | | |
| Test sample | -0.24 | -5.07 | 278 | <.001 | [-0.35, -0.13] | |
| Replication sample | -0.23 | -5.56 | 126 | <.001 | [-0.33, -0.13] | |

*Note.* Output is based on least-square means with Kenward-Roger corrected degrees of freedom and Bonferroni-adjusted *p*-values. Abbreviations: MPFC = medial prefrontal cortex, OFC = orbitofrontal cortex, DLPFC = dorsolateral prefrontal cortex, VLPFC: ventrolateral prefrontal cortex.

**Table S8.** Motion information of the final test and replication samples (after exclusion of participants with >3mm movement).

| | Test sample (n=385) | Replication sample (n=195) |
|---|---|---|
| ***Motion parameters:*** | | |
| Mean (SD) | 0.15 (0.07) | 0.14 (0.08) |
| ***Number of small spikes (0.9 – 3mm):*** | | |
| Mean (SD) | 5.27 (7.00) | 3.87 (7.29) |
| Range (min – max) | 0 – 37 | 0 - 55 |
| Total number of images per participant | 283 – 431 | 290 |
| ***Number of participants:*** | | |
| With 0 spikes (0.9 – 3mm) | 117 | 83 |
| With small spikes (0.9 – 3mm) | 268 | 112 |

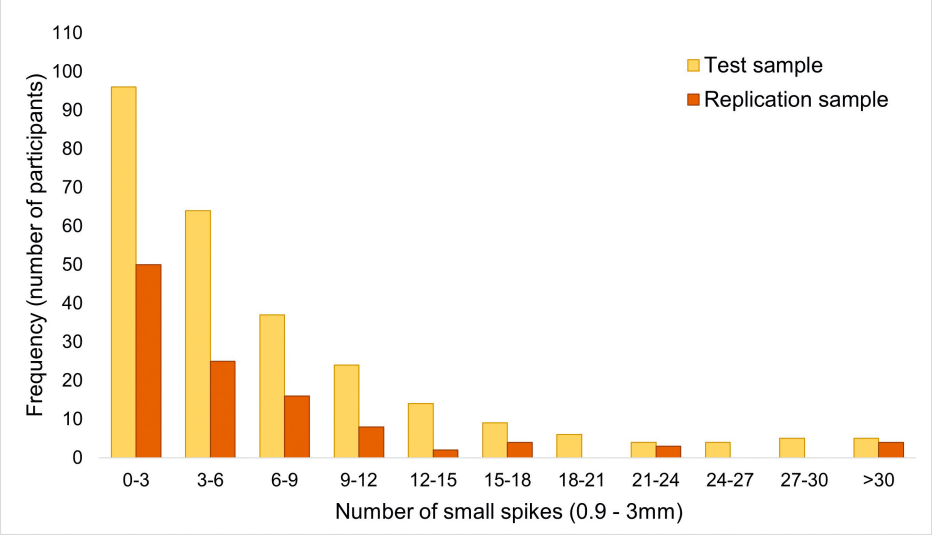


**Figure S4.** Number of participants with small spikes (0.9 – 3mm) in the test sample (in yellow) and replication sample (in red). Note that the total number of images was larger in the test sample (283-431 images) than in the replication sample (290 images).