



Universiteit  
Leiden  
The Netherlands

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

Achterberg, M.

### **Citation**

Achterberg, M. (2020, March 12). *Like me, or else: Nature, nurture and neural mechanisms of social emotion regulation in childhood*. Retrieved from <https://hdl.handle.net/1887/86283>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/86283>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/86283> holds various files of this Leiden University dissertation.

**Author:** Achterberg, M.

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

**Issue Date:** 2020-03-12

## CHAPTER THREE

# Control your anger! The neural basis of aggression regulation in response to negative social feedback

---

This chapter is published as: Achterberg M., Van Duijvenvoorde A.C.K., Bakermans-Kranenburg M.J. & Crone E.A. (2016), Control your anger! The neural basis of aggression regulation in response to negative social feedback., *Social cognitive and affective neuroscience* 11(5): 712-720.

## Abstract

Negative social feedback often generates aggressive feelings and behavior. Prior studies have investigated the neural basis of negative social feedback, but the underlying neural mechanisms of aggression regulation following negative social feedback remain largely undiscovered. In the current study participants viewed pictures of peers with feedback (positive, neutral, or negative) to the participant's personal profile. Next, participants responded to the peer feedback by pressing a button, thereby producing a loud noise towards the peer, as an index of aggression. Behavioral analyses showed that negative feedback led to more aggression (longer noise blasts). Conjunction neuroimaging analyses revealed that both positive and negative feedback were associated with increased activity in the medial prefrontal cortex (mPFC) and bilateral insula. In addition, more activation in the right dorsal lateral PFC (dlPFC) during negative feedback versus neutral feedback was associated with shorter noise blasts in response to negative social feedback, suggesting a potential role of dlPFC in aggression regulation, or top-down control over affective impulsive actions. This study demonstrates a role of the dlPFC in the regulation of aggressive social behavior.

*Keywords: social evaluation; social rejection; social acceptance; emotion regulation; functional magnetic resonance imaging (fMRI)*

## Introduction

People are strongly motivated to be accepted by others and to establish a sense of belonging. Receiving negative social feedback, therefore, is a distressing experience, related to serious negative consequences such as feelings of depression and anxiety (Nolan *et al.*, 2003). For some individuals, receiving negative social feedback can result in aggression towards people who have negatively evaluated or rejected them (Twenge *et al.*, 2001; Leary *et al.*, 2006; DeWall and Bushman, 2011; Chester *et al.*, 2014; Riva *et al.*, 2015; Chester and DeWall, 2016). However, the relation between negative social feedback and subsequent aggression is not well understood. In the current study we investigated the relation between receiving negative social feedback and subsequent aggression using neuroimaging, which allowed us to 1) examine the neural correlates of negative social feedback relative to neutral or positive feedback, 2) examine aggressive responses towards the person signaling negative social feedback, and 3) examine the association between the neural correlates of negative social feedback and behavioral aggression.

Social rejection and negative social feedback have previously been studied using a variety of experimental paradigms that manipulate social contexts. For example, the negative feelings associated with social rejection have been extensively studied using Cyberball, an online ball tossing game in which three players toss balls to each other, until at some point in the game, one of the players is excluded. It is consistently found that this type of social exclusion leads to feelings of distress, negative mood, and a decreased satisfaction of the need for a meaningful existence (Williams *et al.*, 2000; Williams, 2007). Neuroimaging studies point to a role of the midline areas of the brain, specifically the dorsal and subgenual anterior cingulate cortex (ACC), as well as the anterior insula, as important brain regions responding to social exclusion (Cacioppo *et al.*, 2013; Rotge *et al.*, 2015). Other studies have used a peer feedback social evaluation paradigm to study responses to both positive and negative social feedback. In such paradigms, participants believe they are socially evaluated by same-aged peers, based on first impressions of their profile picture (Somerville *et al.*, 2006; Gunther Moor *et al.*, 2010b; Hughes and Beer, 2013). These studies showed that dorsal ACC (dACC) activation was particularly activated in response to unexpected social feedback, irrespective of whether this was positive or negative (Somerville *et al.*, 2006), whereas ventral mPFC and ventral striatum activation was larger for positive feedback compared to negative feedback (Guyer *et al.*, 2009; Davey *et al.*, 2010; Gunther Moor *et al.*, 2010b).

More insight into the neural and behavioral correlates of social evaluation and rejection has been derived from studies testing the relation between social rejection and subsequent aggression. One study combined the Cyberball task in the scanner with a subsequent aggression index using a noise blast task outside of the scanner (Chester *et al.*, 2014). Individuals responded more aggressively

following the experience of social rejection, but intriguingly, these effects were dependent on whether the participant showed low or high executive control. Participants who scored high on executive control displayed lower aggression after social rejection, suggesting that executive control abilities may down-regulate aggression tendencies. It has been suggested that self-control relies strongly on the lateral prefrontal cortex (PFC), which is thought to exert top-down control over subcortical, affective, brain regions (such as the striatum) to suppress outputs that otherwise lead to impulsive response and actions (Casey, 2015). Transcranial magnetic stimulant (TMS) studies have indeed implicated a causal role for the lateral PFC in executing self-control when choosing long-term rewards (Figner *et al.*, 2010). Similarly, lateral PFC may have an important role in down-regulating aggression following rejection or negative social feedback. This hypothesis finds support in a study where participants had the opportunity to aggress to peers who had excluded them during Cyberball while undergoing transcranial direct current stimulation (tDCS) (Riva *et al.*, 2015). TDCS of the right ventrolateral (vl) PFC reduced participants' behavioral aggression to the excluders.

Taken together, prior studies suggested an important role of dorsal and ventral mPFC regions in processing negative and positive social feedback, but the exact contributions of these regions are not consistent across studies and may depend on the experimental paradigm. The first goal of this study was to disentangle effects of positive and negative feedback in a social evaluation paradigm (Somerville *et al.*, 2006). A novel component of this study relative to prior studies is that we included a neutral baseline condition, in which participants received neutral feedback on a subset of the trials. Based on prior research, we expected that positive social feedback would result in increased activation in the subgenual ACC (Somerville *et al.*, 2006) and the ventral striatum (Guyer *et al.*, 2009; Davey *et al.*, 2010; Gunther Moor *et al.*, 2010b). In contrast, we expected that negative social feedback would be associated with increased activity in the dACC/ dorsal medial PFC (dmPFC) and the insula. Prior studies remained elusive about whether dACC/mPFC and insula activity were associated with salient events per se (Somerville *et al.*, 2006) or social rejection specifically (Eisenberger *et al.*, 2003; Kross *et al.*, 2011). Therefore, we conducted conjunction analyses for both positive and negative feedback versus neutral baseline, as well as direct contrasts testing for differences between positive and negative social feedback.

Importantly, there may be individual differences in how participants respond to negative social feedback, which may be associated with increased neural activity in lateral PFC, as has been found in social rejection studies (Chester and DeWall, 2016). The second goal of this study was therefore to examine how individuals respond to negative social feedback, and if lateral PFC activity is related to aggression regulation following negative social feedback. Therefore, the paradigm included a second event where participants could

directly retaliate to the peer who judged them, by sending a loud noise blast (Twenge *et al.*, 2001; Chester *et al.*, 2014). Noise blast duration was measured after each trial within the fMRI task and therefore we could examine how neural activity related to individual differences in noise blast duration. On a behavioral level, we hypothesized that negative social feedback would trigger reactive aggression, i.e. longer noise blasts (Twenge *et al.*, 2001; Reijntjes *et al.*, 2011; Riva *et al.*, 2015). In addition, we hypothesized that less aggression (i.e., more aggression regulation, shorter noise blasts) would be related to increased activation in lateral PFC (Casey, 2015; Riva *et al.*, 2015) particularly during negative feedback.

## Methods

### Participants

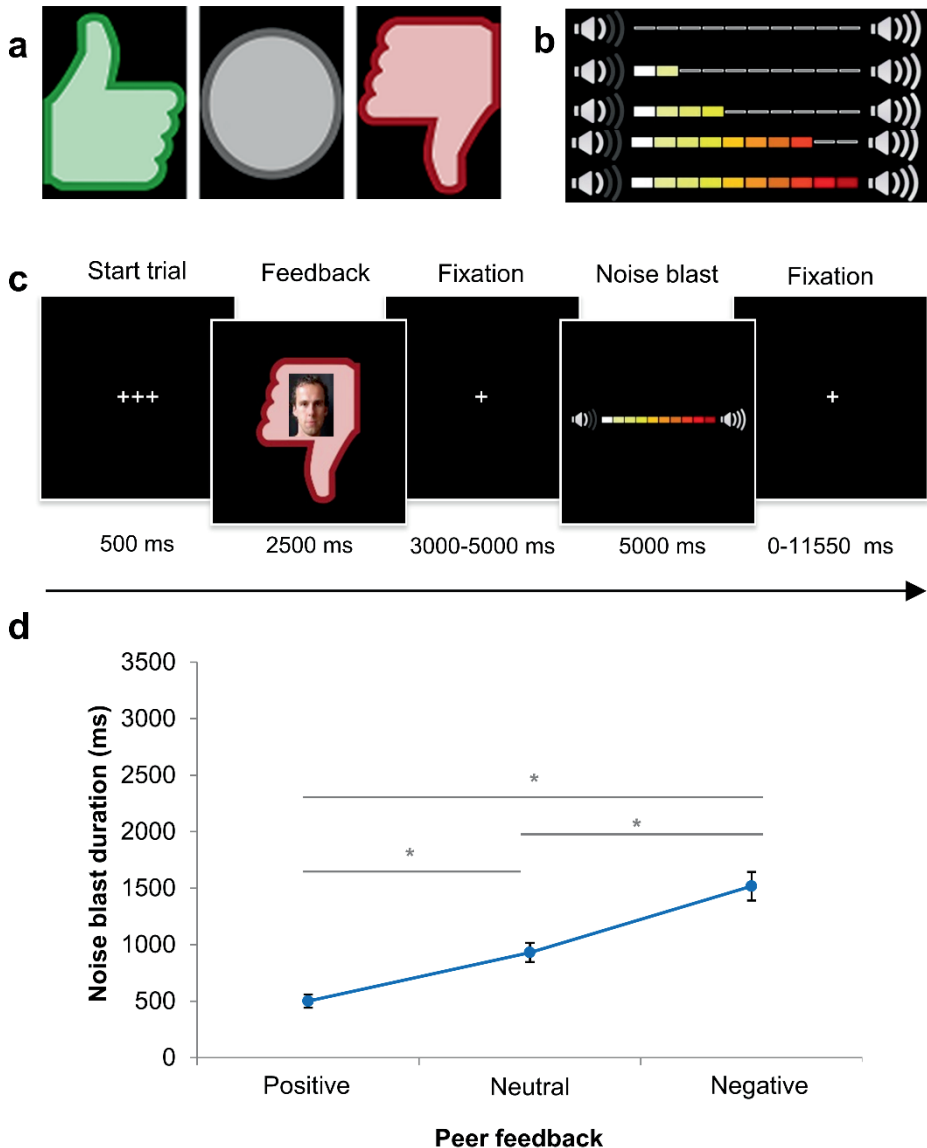
Thirty participants between the ages of 18 and 27 participated in this study (15 females,  $M=22.63$  years,  $SD=2.62$ ). They were either contacted from a participant database or they responded to an advert placed online. The institutional review board of the Leiden University Medical Center (LUMC) approved the study and its procedures. Written informed consent was obtained from all participants. All participants were fluent in Dutch, right-handed, and had normal or corrected-to-normal vision. Participants were screened for MRI contra indications and had no history of neurological or psychiatric disorders. All anatomical MRI scans were reviewed and cleared by a radiologist from the radiology department of the LUMC. No anomalous findings were reported.

Participants' intelligence quotient (IQ) was estimated with the subsets 'similarities' and 'block design' of the Wechsler Intelligence Scale for Adults, third edition (WAIS-III; Wechsler (1997)). All estimated IQs were in the normal to high range (95 to 135;  $M=113.92$ ,  $SD=9.23$ ). IQ scores were not correlated to behavioral outcomes of the Social Network Aggression Task (noise blast duration after positive, neutral, negative feedback and noise blast difference scores, all  $p$ 's > .244)

### Social Network Aggression Task

The Social Network Aggression Task (SNAT) was based on the social evaluation paradigm of Somerville *et al.* (2006) and Gunther Moor *et al.* (2010b). Prior to the fMRI session, participants filled in a profile page at home, which was handed in at least one week before the actual fMRI session. The profile page consisted of personal statements such as: "My favorite sport is...", "This makes me happy...", "My biggest wish is...". Participants were informed that their profiles were viewed by other individuals. During the SNAT participants were presented with pictures and feedback from same-aged peers in response to the participants' personal profile. This feedback could either be positive ('I like your profile', visualized by

a green thumb up); negative ('I do not like your profile'; red thumb down) or neutral ('I don't know what to think of your profile', grey circle), see Figure 1a.



**Figure 1.** Social Network Aggression Task. (a) The different feedback types: positive, neutral and negative. (b) Visual representation of intensity buildup of the volume bar. (c) Display of one trial and timing of the SNAT. (d) Noise blast duration across the different social feedback conditions. Asterisks indicate significant differences with  $p < .05$ .



Following each peer feedback (positive, neutral, negative), participants were instructed to send a loud noise blast to this peer. The longer they would press a button the more intense the noise would be, which was visually represented by a volume bar (Figure 1b). Participants were specifically instructed that the noise was not really sent to the peer, but that they had to imagine that they could send a noise blast to the peer, with the volume intensity of the participants' choice. This was done to reduce deception, and prior studies showed that imagined play also leads to aggression (Konijn *et al.*, 2007). Unbeknownst to the participants, the profile was not judged by others, and the photos were taken from an existing data base with pictures matching participants' age range (Gunther Moor *et al.*, 2010b). Peer pictures were randomly coupled to feedback, ensuring equal gender proportions for each condition. None of the participants expressed doubts about the cover story.

Prior to the scan session, the noise blast was presented to the participants twice during a practice session: once with stepwise buildup of intensity and once at maximum intensity. Two evaluation questions were asked after hearing the maximum intensity: '*How much do you like the sound?*' and '*How much do you dislike the sound?*'. Participants rated the sound on a 7-point scale, with 1 representing *very little* and 7 representing *very much*. In order to prevent that pressing the button during the experimental task would punish the participants themselves, they only heard the intensity of the noise blast during the practice session and not during the fMRI session. To familiarize participants with the task, participants performed six practice trials.

The SNAT consists of two blocks of 30 trials (60 trials in total), with 20 trials for each social feedback condition (positive, neutral, negative), that are presented semi randomized to ensure that no condition is presented more than three times in a row. Figure 1c displays an overview of one SNAT trial. Each trial starts with a fixation screen (500 ms), followed by the social feedback (2500 ms). After another fixation screen (jittered between 3000 and 5000 ms), the noise screen with the volume bar appears, which is presented for a total of 5000 ms. As soon as the participants starts the button press, the volume bar starts to fill up with a newly colored block appearing every 350 ms. After releasing the button, or at maximum intensity (after 3500 ms), the volume bar stops increasing and stays on the screen for the remaining of the 5000 ms. Before the start of the next trial, a fixation cross was presented (jittered between 0 and 11550 ms). The optimal jitter timing and order of events were calculated with Optseq 2 (Dale, 1999).

## Exit questions

Following the MRI session, three exit questions were asked: '*How much did you like reactions with a thumb up?*', '*How much did you like reactions with a circle?*', and '*How much did you like reactions with a thumb down?*'. Participants rated the

reactions on a 7-point scale, with 1 representing *very little* and 7 representing *very much*.

## **MRI data acquisition**

MRI scans were acquired with a standard whole-head coil on a Philips 3.0 Tesla scanner (Philips Achieva TX). The SNAT was projected on a screen that was viewed through a mirror on the head coil. Functional scans were collected during two runs T2\*-weighted echo planar images (EPI). The first two volumes were discarded to allow for equilibration of T1 saturation effect. Volumes covered the whole brain with a field of view (FOV)= 220 (ap) x 220 (rl) x 114.68 (fh) mm; repetition time (TR) of 2.2 seconds; echo time (TE) = 30 ms; sequential acquisition, 38 slices; and voxel size= 2.75 x 2.75 x 2.75 mm. Subsequently, a high-resolution 3D T1scan was obtained as anatomical reference (FOV= 224 (ap) x 177 (rl) x 168 (fh); TR=9.76 ms; TE=4.95 ms; 140 slices; voxel size 0.875 x 0.875 x 0.875 mm).

## **MRI data analyses**

### **Preprocessing**

MRI data were analyzed with SPM8 (Wellcome Trust Centre for Neuroimaging, London). Images were corrected for slice timing acquisition and rigid body motion. Functional scans were spatially normalized to T1 templates. Due to T1 misregistration, one participant was normalized to an EPI template. Volumes of all participants were resampled to 3x3x3 mm voxels. Data were spatially smoothed with a 6 mm full width at half maximum (FWHM) isotropic Gaussian kernel. Translational movement parameters never exceeded 1 voxel (<3 mm) in any direction for any participant or scan (movement range: 0.001-1.22 mm,  $M=0.055$ ,  $SD=0.036$ ).

### **First-level analyses**

Statistical analyses were performed on individual subjects' data using a general linear model. The fMRI time series were modeled as a series of two events convolved with the hemodynamic response function (HRF). The onset of social feedback was modeled as the first event with a zero duration and with separate regressors for the positive, negative, and neutral peer feedback. The start of the noise blast was modeled for the length of the noise blast duration (i.e., length of button press) and with separate regressors for noise blast after positive, negative, and neutral feedback. Trials on which the participants failed to respond in time were marked as invalid. Note that this happened rarely, on average 3.78% of the trials were invalid. The least squares parameter estimates of height of the best-fitting canonical HRF for each condition were used in pairwise contrasts. The pairwise comparisons resulted in subject-specific contrast images.

### Higher-level group analyses

Subject-specific contrast images were used for the group analyses. A full factorial ANOVA with three levels (positive, negative, and neutral feedback) was used to investigate the neural response to the social feedback event. We calculated the contrasts 'Positive versus Negative feedback', 'Positive versus Neutral feedback' and 'Negative versus Neutral feedback'. To investigate regions that were activated both after negative social feedback and after positive social feedback, we conducted a conjunction analysis to explore the main effect of social evaluation. Based on Nichols *et al.* (2005), we used the 'logical AND' strategy. The 'logical AND' strategy requires that all the comparisons in the conjunction are individually significant (Nichols *et al.*, 2005).

All results were False Discovery Rate (FDR) cluster corrected ( $p_{FDR} < .05$ ), with a primary voxel-wise threshold of  $p < .005$  (uncorrected) (Woo *et al.*, 2014). Coordinates for local maxima are reported in MNI space. To further visualize patterns of activation in the clusters identified in the whole brain regression analysis, we used the MarsBaR toolbox (Brett *et al.*, 2002) (<http://marsbar.sourceforge.net>).

In all behavioral repeated measures analyses, Greenhouse-Geisser (GG) corrections were applied when the assumption of sphericity was violated. When outliers were detected ( $Z$ -value  $< -3.29$  or  $> 3.29$ ), scores were winsorized (Tabachnick and Fidell, 2013).

## Results

### Behavioral analyses

#### Noise blast manipulation check

The ratings of how much participants liked the maximum intensity noise blast indicated that overall the noise blast was not liked ( $M=1.47$ ,  $SD=0.78$ ; range 1-4) and much disliked ( $M=5.67$ ,  $SD=1.30$ ; range 1-7). These results show that the noise blast was indeed perceived as a negative event by the participants.

#### Social feedback manipulation check

To verify whether participants differentially liked the social feedback conditions (positive, negative, neutral), we analyzed the exit questions with a repeated measures ANOVA. Analyses showed a significant main effect of type of feedback on feedback liking,  $F(2, 58)=53.63$ ,  $p < .001$  (GG corrected), with a large effect size ( $\omega^2 = 0.53$ ). Pairwise comparisons (Bonferroni corrected) showed that participants liked negative feedback ( $M=3.13$ ,  $SD=0.14$ ) significantly less than neutral feedback ( $M=4.23$ ,  $SD=0.14$ ,  $p < .001$ ) and positive feedback ( $M=5.23$ ,  $SD=0.16$ ,  $p < .001$ ). Participants also liked neutral feedback significantly less than that positive feedback ( $p < .001$ ).

### Noise blast duration

A repeated measures ANOVA was performed on noise blast duration after positive, negative, and neutral feedback. Results showed a significant main effect of type of social feedback on noise blast duration,  $F(2, 58)=75.57$ ,  $p<.001$  (GG corrected), with a large effect size ( $\omega^2 = 0.41$ ), see Figure 1d. Pairwise comparisons (Bonferroni corrected) revealed that noise blast duration after negative feedback ( $M=1517.08$ ,  $SD=126.94$ ) was significantly longer than noise blast duration after neutral feedback ( $M=930.41$ ;  $SD=84.77$ ,  $p<.001$ ), and after positive feedback ( $M=483.62$ ;  $SD=47.19$ ,  $p<.001$ ). Noise blast duration after neutral feedback was significantly longer than after positive feedback ( $p<.001$ ).

To derive a measure indicative of individual differences in aggression we calculated the differences in noise blast duration between negative *versus* neutral feedback and positive *versus* neutral feedback. The noise blast difference for positive-neutral was significantly negatively correlated to the noise blast difference for negative-neutral ( $r= -.48$ ,  $p=.008$ ), indicating that shorter noise blasts after positive feedback (compared to neutral feedback) were related to longer noise blasts after negative feedback (compared to neutral feedback). Next, noise blast differences were correlated with the exit questions. The difference of negative-neutral was positively correlated to the feedback liking of positive feedback ( $r= .39$ ,  $p=.032$ ) and negatively correlated to the feedback liking of negative feedback ( $r= -.57$ ,  $p=.001$ ), indicating that longer noise blasts after negative feedback were related to a stronger preference for positive social feedback and a stronger disfavor of negative social feedback see Figures S1a and S1b. Similarly, the noise blast difference of positive-neutral was negatively correlated to the feedback liking of positive feedback ( $r= -.42$ ,  $p=.021$ ) and positively correlated to the feedback liking of negative feedback ( $r= .73$ ,  $p<.001$ ), indicating that a stronger preference for positive social feedback and a stronger disfavor of negative social feedback were related to shorter noise blasts after positive feedback (see Figures S1c and S1d).

## fMRI whole brain analyses

### Social evaluation

The first goal was to examine neural activity in the contrast positive versus negative feedback at the moment of peer feedback. The contrast Positive > Negative feedback resulted in activation with local maxima in the bilateral lateral occipital lobes, left postcentral, and activation in the right and left striatum, extending into subgenual ACC (see Figure 2a, Table S1). The contrast Negative > Positive feedback did not result in any significant clusters of activation. Next, we tested how neural activity to positive and negative social feedback related to a neutral baseline condition. The contrast Negative > Neutral feedback resulted in activity in the bilateral insula and mPFC, see Figure 2b (Table S2). The reversed contrast (Neutral > Negative feedback) did not result in any significant clusters

of activation. The contrast Positive > Neutral feedback also revealed widespread activation in the bilateral insula and mPFC. In addition, the contrast resulted in increased activity in the ventral striatum, the subgenual ACC, as well as regions such as the occipital lobe, as shown in Figure 2c (Table S2). The reversed contrast (Neutral > Positive feedback) resulted in activity in the right insula and right postcentral gyrus (Table S2).

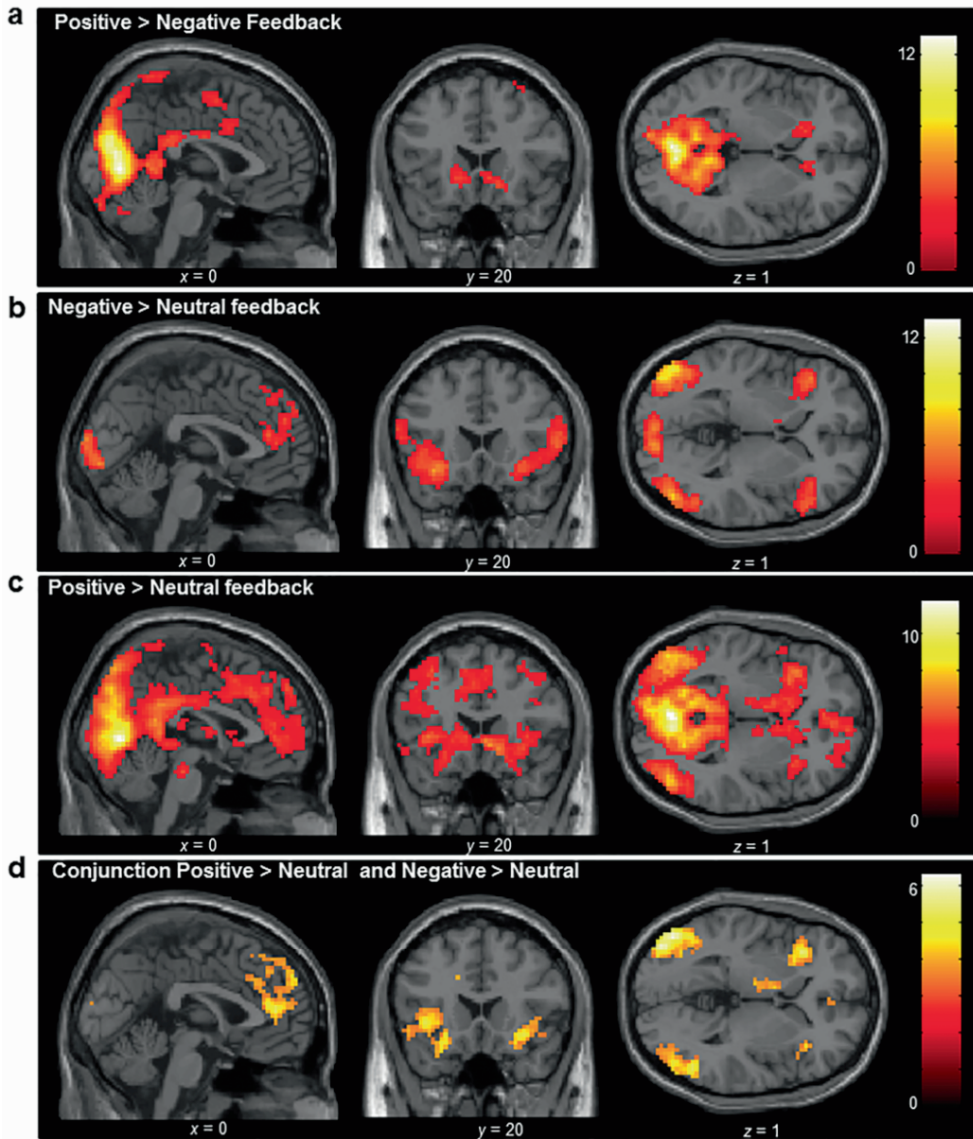
### **Social evaluation conjunction**

The analyses above suggested partially overlapping activation patterns for positive and negative social feedback, relative to a neutral baseline. To formally investigate the regions that were activated both after negative social feedback and after positive social feedback, we conducted a conjunction analyses to explore a main effect of social evaluation. Common activation across both positive and negative social feedback were observed in the insula and the mPFC, as well as the bilateral occipital lobes, including left Fusiform Face Area (FFA), see Figure 2d (Table S3).

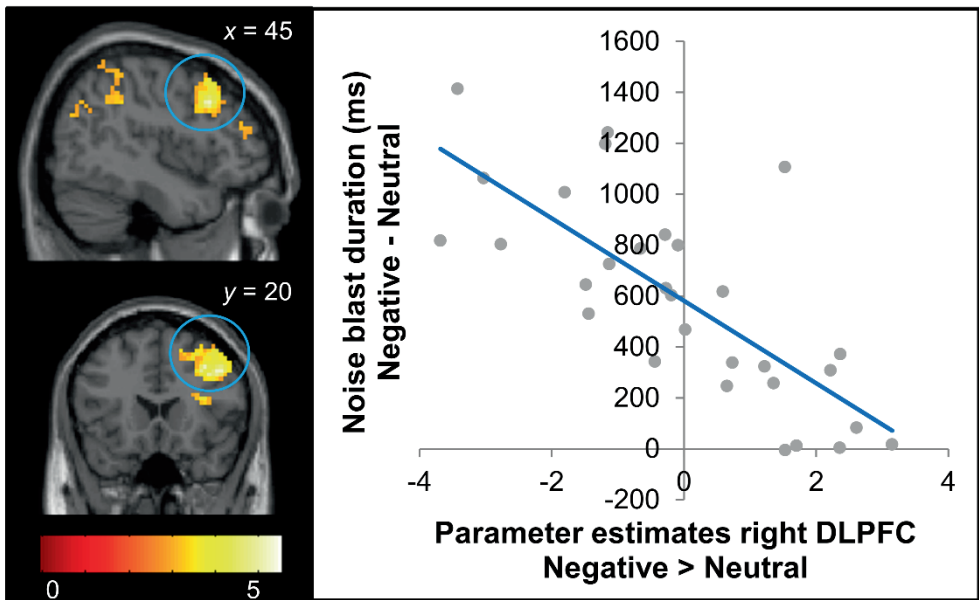
## **Brain-Behavior associations**

### **Noise blast duration**

To test the association between brain activity and behavior in response to negative social feedback, we conducted a whole brain regression analysis at the moment of receiving negative social feedback (relative to neutral feedback; Negative > Neutral), with the difference in noise blast duration after negative and neutral feedback as a regressor. This way, we tested how initial neural responses to feedback were related to subsequent aggression. The analyses revealed that increased activation in the right dorsal lateral PFC (dlPFC) was associated with smaller increases in noise blast duration after negative social feedback compared to neutral feedback, see Figure 3. A similar relation was observed for the left amygdala, left hippocampus, and bilateral superior parietal cortex (Table S4). The reversed contrast (positive relation between Negative > Neutral feedback and noise blast length difference) did not result in any significant activation.



**Figure 2.** Whole brain full factorial ANOVA conducted at group level for the contrasts (a) Positive>Negative feedback, (b) Negative>Neutral feedback, (c) Positive>Neutral feedback and (d) the conjunction of the Positive>Neutral and Negative>Neutral feedback contrasts. Results were FDR cluster corrected ( $PFDR < 0.05$ ), with a primary voxel-wise threshold of  $P < 0.005$  (uncorrected).



**Figure 3.** Brain regions in the contrast Negative>Neutral feedback that were significantly negatively correlated with the difference in noise blast duration after negative vs neutral feedback trials. Results were FDR cluster corrected ( $PFDR < 0.05$ ), with a primary voxel-wise threshold of  $P < 0.005$  (uncorrected). The right panel shows the negative relationship between difference in noise blast duration and right dLPFC (for visual illustration only, no statistical tests were carried out on the region of interest).

## Discussion

This study investigated the relation between negative social feedback and subsequent aggression, using neuroimaging. The goals of this study were threefold: 1) to disentangle neural signals of positive and negative social feedback, 2) to examine aggressive responses towards the person signaling negative social feedback, and 3) to test whether lateral PFC activity is related to aggression regulation after experiencing negative social feedback. To these ends, we developed a new social peer evaluation paradigm that included neutral feedback (to be able to compare positive and negative feedback to a neutral baseline) and the possibility to retaliate to the peer that gave the feedback (to be able to study aggression related to social feedback). In line with prior behavioral studies we found that negative social feedback was related to applying a longer noise blast towards the peer (Chester *et al.*, 2014). At the neural level, conjunction analyses showed that both negative and positive social feedback resulted in increased activity in the mPFC and the bilateral insula. Comparing the



conjunction analyses with the separate contrasts of negative and positive versus neutral feedback showed that positive feedback resulted in increased activity in the striatum and the ventral mPFC, whereas negative feedback activation merely overlapped with dorsal mPFC and insula activation observed following both positive and negative feedback. Finally, we found that increased lateral PFC activity after negative social feedback was associated with relative shorter noise blast durations after negative feedback, indicative of more aggression regulation.

Results of prior studies left undecided whether there is a unique neural coding for negative social feedback compared to positive social feedback. In this study we found that, consistent with prior studies (Guyer *et al.*, 2009; Davey *et al.*, 2010; Gunther Moor *et al.*, 2010b) there was increased activity in the ventral mPFC and the striatum after positive feedback. Numerous studies have shown that the striatum is involved in reward processing (for a review, see Sescousse *et al.* (2013)) and this fits well with theories suggesting that positive evaluations and social acceptance activates brain regions overlapping with those that are activated by the primary feelings of reward (Lieberman and Eisenberger, 2009). Notably, there was no neural activation that was specific for negative social feedback. In Cyberball paradigms, a number of studies observed specific heightened activity in insula and ACC in response to social rejection, which was interpreted as the feeling of social pain (Eisenberger and Lieberman, 2004; Lieberman and Eisenberger, 2009). There are several differences in the experimental paradigms, however, that may explain the divergent results. That is to say, in Cyberball paradigms social rejection is unexpected (for example, exclusion after a period of inclusion) and is therefore likely to violate social expectations. In contrast, in social evaluation paradigms such as used in the current study, equal proportions of negative, positive, and neutral feedback are presented, which may result in more equal saliency of negative and positive feedback. The current findings, which show enhanced insula and mPFC activity following both positive and negative feedback (relative to neutral feedback), suggest that the insula and mPFC in social evaluation paradigms might work as a salience network, and signal events that are socially relevant (Guroglu *et al.*, 2010; van den Bos *et al.*, 2011). Resting-state fMRI studies confirm that these regions are often active in concert, and have referred to this network as a salience network (Damoiseaux *et al.*, 2006; Jolles *et al.*, 2011; van Duijvenvoorde *et al.*, 2016a). Future research may disentangle the role of expectation violation in more detail by asking participants to make predictions about whether they expect to be liked (Somerville *et al.*, 2006; Gunther Moor *et al.* 2010), in combination with positive, negative, and neutral feedback.

An additional goal of this study was to examine the association between brain activation and behavioral responses to negative social feedback. A vast line of research has already shown that social rejection can result in retaliation (Twenge *et al.*, 2001; Leary *et al.*, 2006; DeWall and Bushman, 2011; Chester *et al.*, 2014; Riva *et al.*, 2015). Our study shows that receiving negative social



feedback is also followed by more aggressive behavior (i.e., by a longer noise blast towards the peer). In addition, we show that more activity in the right dlPFC is related to *less* aggression after negative social feedback (compared to neutral feedback), indicating that the lateral PFC is an important neural regulator of social aggression. Several studies on structural brain development have shown that the quality of brain connectivity between the PFC and the striatum is related to impulse control (Peper *et al.*, 2013; van den Bos *et al.*, 2014). That is to say, a large study on structural brain connectivity in typically developing individuals (258 participants, aged 8-25) revealed that less white matter integrity between subcortical and prefrontal brain regions was associated with more trait aggression (Peper *et al.*, 2015). Moreover, Chester and DeWall (2016) recently demonstrated that more functional connectivity between the nucleus accumbens and the lateral PFC during decisions about aggressive acts was related to less behavioral aggression. This study is the first study to investigate aggressive responses after positive, neutral, and negative feedback, and shows a role of the dlPFC in individual differences in the regulation of aggressive behavior.

Some limitations regarding this study need to be acknowledged. First, although the noise blast is often used as a measure of aggression (e.g., Bushman (2002); Chester *et al.* (2014); Riva *et al.* (2015)), our cover story stated that the peers would not hear the noise blast. That is to say, the aggression measure may reflect frustration and anger, and hypothetical aggression. Future research should further test the ecological validity of the noise blast as a measure of aggression by including additional measures of aggression or information on participants' histories of aggressive behavior. Secondly, our paradigm did not include an 'opt out' option, that is, we told participants to always push the noise blast button, even after positive feedback. This was done to keep task demands as similar as possible between the conditions. We explained that the noise would be very short and at very low intensity if the button was released as quickly as possible. However, participants may have wanted to refrain from any noise blast after positive feedback. Future research could take this into account by implementing options to respond either positive, neutral, or negative towards the peer, as can for example be implemented by using symbols (Jarcho *et al.*, 2013).

## Conclusion

In conclusion, we found evidence that the insula and mPFC generally respond to socially salient feedback, with no significant differentiation between negative and positive feedback. Positive social feedback received less attention in prior research and it has often been used as a baseline, but our findings show activation in the ventral mPFC and the striatum that is stronger for positive feedback. Additionally, the lateral PFC emerged as an important modulator for individual differences in aggression regulation. This may imply that individuals who show strong activation in the lateral PFC after negative social feedback may be better able to regulate behavioral impulses, and speculatively, impulsive responses in general (Casey *et al.*, 2011). This hypothesis that should be addressed in longitudinal research, including more general measures of impulsivity. An interesting direction for future research is to examine the neural mechanisms underlying social evaluation and aggression regulation processes in populations that are known for difficulties with response control and affect regulation, such as ADHD (Evans *et al.*, 2015), externalizing problems (Prinstein and La Greca, 2004), and depression (Nolan *et al.*, 2003; Silk *et al.*, 2014).

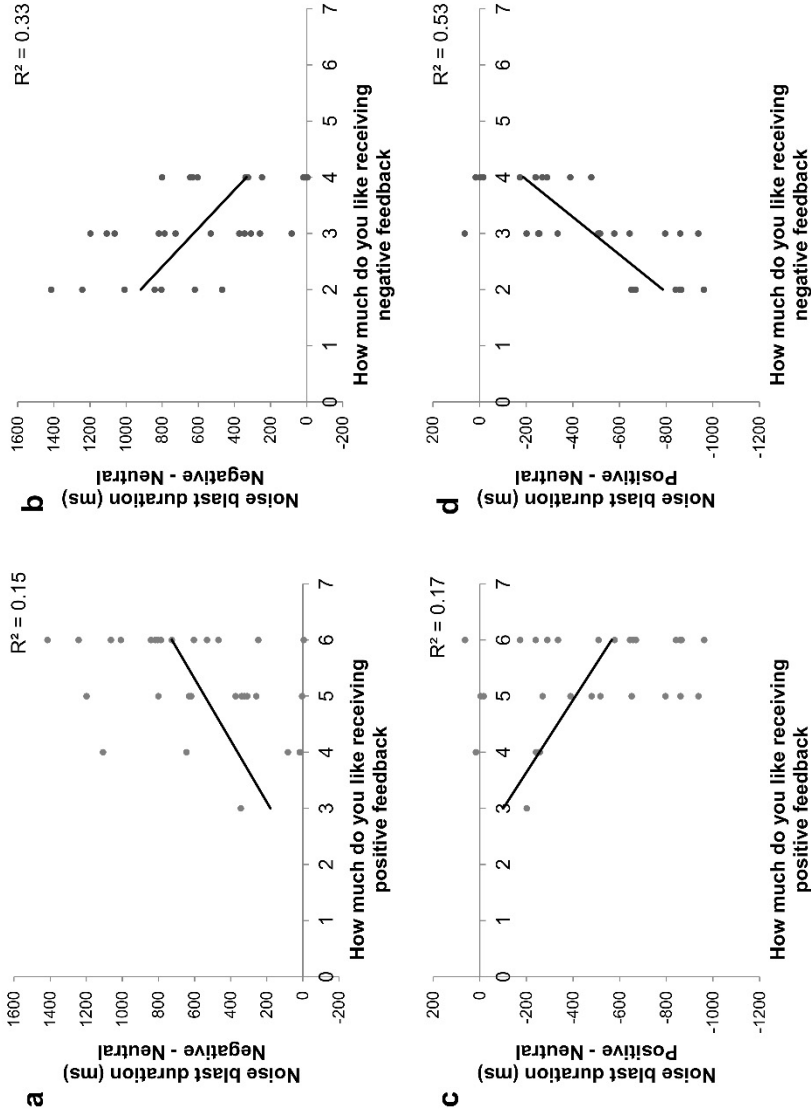
## Acknowledgments

The Consortium on Individual Development is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). MJBK was funded by the Netherlands Organization for Scientific Research (VICI) and the European Research Council (AdG 669249).

## Supplementary materials

**Table S1.** MNI coordinates for local maxima activated for the contrasts Positive > Negative feedback. The results were FDR cluster corrected ( $p_{FDR} < .05$ ), with a primary voxel-wise threshold of  $p < .005$  (uncorrected).

Area of Activation	<i>x</i>	<i>y</i>	<i>z</i>	Voxels	<i>T</i>
<i>Positive &gt; Negative feedback</i>					
Lingual Gyrus	0	-73	4	5491	12.99
R Intracalcarine Cortex	3	-76	13		12.91
Cuneal Cortex	0	-79	22		11.17
L Supramarginal Gyrus	-39	-37	40	951	6.61
L Supramarginal Gyrus	-48	-34	43		5.53
L Postcentral Gyrus	-54	-28	46		5.19
L Caudate	-12	23	-5	76	4.72
L Caudate	-21	29	-2		4.17
L Supplementary Motor Cortex	-12	2	52	206	4.43
L Supplementary Motor Cortex	-9	-10	55		3.95
R Supplementary Motor Cortex	3	-1	55		3.89
R Orbito Frontal Cortex	15	23	-8	90	4.34
R Orbito Frontal Cortex	18	17	-14		4.04
R Orbito Frontal Cortex	27	26	-8		2.84
R Superior Frontal Gyrus	18	11	46	102	4.30
R Middle Frontal Gyrus	30	8	52		4.27
R Middle Frontal Gyrus	33	14	64		3.48
L Precentral Gyrus	-60	5	28	82	4.08
L Precentral Gyrus	-54	-1	37		3.79



**Figure S1** Relation between noise blast duration difference scores and sensitivity to social evaluation. The y-axis displays the noise blast duration difference between different feedback types as measured by the exit questions. The x-axis displays the liking of the different feedback types. Participants sensitive to social feedback (i.e., liked receiving positive feedback and dislike receiving negative feedback) used a larger noise blast after negative feedback (a and b) and a shorter noise blast after positive feedback (c and d).

**Table S2.** MNI coordinates for local maxima activated for the contrasts Negative > Neutral feedback, Positive > Neutral feedback and Neutral > Positive feedback. Results were FDR cluster corrected ( $p_{\text{FDR}} < .05$ ), with a primary voxel-wise threshold of  $p < .005$  (uncorrected).

Area of Activation	x	y	z	Voxels	T
<i>Negative &gt; Neutral feedback</i>					
L Lateral Occipital Cortex	-48	-76	-2	3122	12.89
L Occipital Pole	-15	-97	22		8.82
L Occipital Pole	-9	-94	28		8.71
L Orbital Frontal Cortex	-33	20	-14	500	6.48
L Frontal Operculum Cortex	-36	29	4		5.51
L Insular Cortex	-33	23	-2		5.02
R Inferior Frontal Gyrus	57	26	7	346	6.12
R Orbital Frontal Cortex	45	29	-2		4.98
R Insular Cortex	36	17	-11		4.94
L Frontal Pole	-12	41	49	379	4.89
R Anterior Cingulate Cortex	6	35	16		4.65
L Frontal Pole	-12	50	46		4.56
<i>Positive &gt; Neutral feedback</i>					
L Lingual Gyrus	-3	-76	1	14183	14.39
R Lingual Gyrus	6	-73	1		12.87
R Lingual Gyrus	18	-73	-8		11.25
R Precentral Gyrus	45	-1	49	62	4.71
R Middle Frontal Gyrus	33	8	43		3.14

**Table S2.** (continued)

Area of Activation	x	y	z	Voxels	T
<i>Positive &gt; Neutral feedback</i>					
R Middle Frontal Gyrus	30	23	52	107	4.07
R Superior Frontal Gyrus	21	35	49		4.03
R Middle Frontal Gyrus	39	23	46		3.65
R Middle Frontal Gyrus	45	26	25	79	4.03
R Precentral Gyrus	36	5	31		3.47
R Middle Frontal Gyrus	39	32	31		3.35
L Superior Frontal Gyrus	-12	-4	67	110	3.98
L Superior Frontal Gyrus	-12	5	73		3.88
L Superior Frontal Gyrus	-18	-7	73		3.81
<i>Neutral &gt; Positive feedback</i>					
R Insular Cortex	36	-22	4	289	5.72
R Insular Cortex	30	-22	16		4.88
R Superior Temporal Gyrus	63	-16	4		4.75
R Postcentral Gyrus	33	-25	58	236	5.39
R Precentral Gyrus	33	-25	67		5.11
R Precentral Gyrus	33	-22	49		4.61

**Table S3.** MNI coordinates for local maxima activated for the conjunction of Positive > Neutral and Negative > Neutral. Results were FDR cluster corrected ( $p_{FDR} < .05$ ), with a primary voxel-wise threshold of  $p < .005$  (uncorrected).

Area of Activation	x	y	z	Voxels	T
<i>Conjunction of Positive &gt; Neutral &amp; Negative &gt; Neutral</i>					
L Lateral Occipital Cortex	-48	-76	-2	965	6.23
L Lateral Occipital Cortex	-48	-79	7		5.99
L Lateral Occipital Cortex	-51	-67	7		5.85
L Insular Cortex	-36	23	-2	320	5.06
L Insular Cortex	-30	14	-14		4.46
L Insular Cortex	-30	17	1		3.81
R Lateral Occipital Cortex	51	-61	-2	518	4.99
R Lateral Occipital Cortex	51	-79	7		4.62
R Lateral Occipital Cortex	42	-79	7		4.41
Cingulate Gyrus	0	44	10	367	4.79
Anterior Cingulate Gyrus	0	38	16		4.35
R Frontal Pole	3	56	13		3.95
R Insular Cortex	33	20	-11	94	4.58
R Orbito Frontal Cortex	36	26	1		3.81

**Table S4.** MNI coordinates for local maxima activated for the whole brain regression analyses. The contrast Negative > Neutral feedback with Negative - Neutral noise blast duration difference as negative regressor. Results were FDR cluster corrected ( $p_{FDR} < .05$ ), with a primary voxel-wise threshold of  $p < .005$  (uncorrected). dlPFC = dorsolateral prefrontal cortex.

Area of Activation	x	y	z	Voxels	T
<i>Negative &gt; Neutral feedback, Noise blast duration difference score as negative regressor</i>					
L Amygdala	-21	-7	-17	173	5.57
L Amygdala	-15	-10	-11		4.21
L Hippocampus	-36	-13	-17		4.01
R Middle Frontal Gyrus (dlPFC)	48	17	37	1144	5.17
R Middle Frontal Gyrus (dlPFC)	36	20	40		5.01
R Middle Frontal Gyrus (dlPFC)	39	14	34		4.57
L Superior Parietal Lobule	-24	-46	37	315	4.95
L Superior Parietal Lobule	-33	-43	52		4.87
L Supramarginal Gyrus	-39	-46	43		4.53
Thalamus	0	-7	16	105	4.62
L Thalamus	-15	-7	10		4.04
L Caudate	-12	-4	19		3.97
R Superior Parietal Lobule	30	-52	40	697	4.61
R Postcentral Gyrus	30	-31	37		4.29
R Lateral Occipital Cortex	15	-70	67		4.26