OXFORD

Social Cognitive and Affective Neuroscience, 2015, 1–10

doi: 10.1093/scan/nsv082
Original article

The pleasure of revenge: retaliatory aggression arises from a neural imbalance toward reward

David S. Chester and C. Nathan DeWall

Department of Psychology, University of Kentucky, Lexington, KY 40506, USA

Correspondence should be addressed to David S. Chester, Department of Psychology, University of Kentucky, 0003 Kastle Hall, Lexington, KY 40506, USA. E-mail: davidchester@uky.edu.

Abstract

Most of daily life hums along peacefully but provocations tip the balance toward aggression. Negative feelings are often invoked to explain why people lash out after an insult. Yet people might retaliate because provocation makes aggression hedonically rewarding. To test this alternative hypothesis, 69 participants underwent functional neuroimaging while they completed a behavioral aggression task that repeatedly manipulated whether aggression was preceded by an instance of provocation or not. After provocation, greater activity in the nucleus accumbens (NAcc) (a brain region reliably associated with reward) during aggressive decisions predicted louder noise blasts administered in retaliation. Greater NAcc activation was also associated with participants' history of real-world violence. Functional connectivity between the NAcc and a regulatory region in the lateral prefrontal cortex related to lower retaliatory aggression. These findings suggest that provocation tips the neural balance towards hedonic reward, which fosters retaliatory aggression. Although such pleasure of inflicting pain may promote retaliatory aggression, self-regulatory processes can keep such aggressive urges at bay. Implications for theory and violence reduction are discussed.

Key words: aggression; nucleus accumbens; reward; lateral prefrontal cortex; self-regulation

Introduction

Revenge is sweet and not fattening. –*Alfred Hitchcock*

Imagine your clenched fist cracking another person's jaw. How does it feel? The answer likely depends on several factors. Important among these is whether or not the target of the haymaker had recently provoked you. Stolen parking spots, rude emails and pugnacious parents often trigger such aggressive outbursts in daily life. Conventionally, scientific explanations of violence have focused on how negative feelings like anger precede aggressive responses to provocation (Berkowitz, 1989; Anderson and Bushman, 2002). Yet everyday experience and a budding literature suggest that revenge is sweet and retaliatory aggression may be driven by hedonic reward (Bushman *et al.*, 2001; Krämer *et al.*, 2007). The current research provides an empirical examination of this possibility.

We propose that provoked people respond aggressively because doing so is hedonically rewarding. Further, we predict that self-regulatory processes can intervene on this link between reward and aggression, reducing such behavior. We tested these hypotheses using behavioral and functional magnetic resonance imaging (fMRI) methods in a relatively large sample ($N = 69$).

The catharsis mystery

The idea that aggression can be experienced as pleasant dates back to psychology's infancy. Sigmund Freud began this trend by popularizing the notion of catharsis, which often took the form of 'releasing' anger through aggressive acts. The concept of cathartic aggression has thrived into modernity, as violent outbursts are commonly perceived as a viable means to replace negative affect with positive affect, which fuels aggressive behavior (see Bushman, 2002). Reflecting this trend, retaliatory aggression is greatest among individuals who expect that aggression will coincide with an improved mood (Bushman *et al.*, 1999, 2001). When provoked people believe that aggression

Received: 4 February 2015; **Revised:** 16 June 2015; **Accepted:** 24 June 2015

will not improve their mood, they no longer behave as aggressively (Bushman *et al.*, 2001). Yet does aggression actually feel good? Or is the notion of cathartic aggression based on a false premise?

The pleasure of inflicting pain

Preliminary behavioral evidence offers hints that provoked aggression is hedonically rewarding. In one study, participants rated aggressive responses to provocation as more pleasurable than unprovoked aggression (Ramírez *et al.*, 2005). Extrinsic rewards for retaliation reduced the self-reported enjoyment of aggressive tasks against provocateurs, which supports an intrinsically rewarding model of this behavior (Carré *et al.*, 2010). Participants' motivation to harm disliked outgroup members was positively correlated with the extent to which they recruited the muscles used for smiling (Cikara and Fiske, 2011). Taken together, these findings suggest that aggression can be rewarding, but that this experience is dependent upon a prior instance of provocation.

Neuroimaging evidence has corroborated the behavioral findings that retaliation is hedonically rewarding. After an insult, participants showed both greater aggression and greater left-hemispheric frontal asymmetry, an indicator of the activation of the behavioral approach system (BAS; Harmon-Jones and Sigelman, 2001). BAS activation is reliably linked to the hedonic experience of positive affect (Gray, 1994). However, approach motivation and hedonic reward differ in that the former represents a behavioral propensity and the latter refers to a subjective, valenced experience. Electroencephalography techniques have been crucial in understanding the role of the BAS in aggressive tendencies, but other neuroimaging techniques with the spatial resolution to explore subcortical functioning have yielded other insights into the neural correlates of aggressive behavior.

Research on the neural basis of punishment demonstrated that monetary penalties delivered to unfair individuals were associated with activity in two brain regions previously associated with reward processing: the caudate nucleus and ventromedial prefrontal cortex (VMPFC; de Quervain *et al.*, 2004; Lotze *et al.*, 2007). These findings have been interpreted as indicating that individuals experience pleasure in response to the punishment of individuals who are perceived to deserve such retribution. The dorsal MPFC (DMPFC) is also active when individuals selected the level of harm to inflict on a provocateur, suggesting that social cognitive processes (e.g. mentalizing) are critical components of aggression (Lotze *et al.*, 2007). The DMPFC, along with the dorsal anterior cingulate cortex, hippocampus and anterior insula, has robust associations with the experience of anger, angry rumination and displaced aggression in response to provocation (Denson *et al.*, 2008). These findings from the neuroscientific literature of punishment inform our basis for understanding the neural correlates of retaliatory aggression as such violent acts are themselves a form of punishment.

The seminal neuroimaging study to explore the neural correlates of retaliatory aggression showed greater activity in the caudate nucleus when participants exhibited greater retaliatory aggression (Krämer *et al.*, 2007). This caudate activity was interpreted as an indicator of reward. However, the caudate's role in the hedonic experience of reward is not clearly supported. This region, along with the rest of the dorsal striatum, functions more in the domains of behavioral response selection in service of motivational and goal states (Grahn *et al.*, 2008) and the habituation of rewarding behaviors (Graybiel, 2008). Large-scale meta-analyses of appetitive cue reactivity paradigms fail to reliably show caudate activity, challenging this region's putative role in the experience of hedonic reward (Chase *et al.*, 2011). If the caudate is not a reliable marker of hedonic reward, what brain region is?

The NAcc and VLPFC: their opposing relations to retaliatory aggression

NAcc and reward

The ventral striatum, specifically the nucleus accumbens (NAcc), is a brain region that is best described as 'a servant to many masters' (Floresco, 2015, p. 27). The NAcc is a crucial node in learning, motivation, and reward circuits that serves to promote goal attainment (Shohamy, 2011; Floresco, 2015). Of these various psychological processes, the NAcc is most reliably associated with the subjective experience of hedonic reward and pleasure (Diekhof *et al.*, 2012; Kühn and Gallinat, 2012; Bartra *et al.*, 2013; Berridge and Kringelbach, 2013). Using Neurosynth (neurosynth.org; Yarkoni *et al.*, 2011), a publically accessible, meta-analytic database of thousands of neuroimaging studies, we observed that the keyword 'reward' produces two relatively large reverse inference clusters of brain activity in the bilateral NAcc (left peak $Z = 25.42$, $x = 10$, $y = 8$, $z = -7$; right peak $Z = 25.00$, $x = 12$, $y = 9$, $z = -8$; 560 studies). Whereas the keyword 'learning' yielded relatively smaller and weaker clusters of NAcc activity (left peak $Z = 6.83$, $x = -12$, $y = 8$, $z = -10$; right peak $Z = 4.66$, $x = 12$, $y = 8$, $z = -10$; 807 studies). Together, this wealth of evidence demonstrates that the NAcc isn't merely a 'pleasure center', but still it is a crucial and robust neural correlate of the experience of reward.

NAcc and aggression

If retaliatory aggression is truly a rewarding behavior, it should be correlated with activity in the NAcc. In behavioral economics tasks such as the Ultimatum Game, punishment of defectors who had previously acted in an unfair manner was predicted by NAcc activity (Strobel *et al.*, 2011). However, this punishment took the form of removing a monetary reward. This type of behavior falls short of the typical definition of aggression, which is the act of intentionally harming others who are motivated to avoid the harm (Anderson and Bushman, 2002). NAcc activity has also been associated with passively viewing the suffering of disliked others (e.g. defectors, envied classmates, outgroup members; Singer *et al.*, 2006; Takahashi *et al.*, 2009; Cikara *et al.*, 2011). Suggesting a link to behavior, NAcc activity in response to the misfortunes of others also correlates with the motivation to harm them (Singer *et al.*, 2006; Cikara *et al.*, 2011). But to date, no research has clearly implicated the NAcc as the substrate of actual retaliatory aggression.

VLPFC, NAcc and aggression

Despite sharing no direct, anatomical connection, the NAcc exists in a regulatory equilibrium with lateral portions of the prefrontal cortex (Heatherton and Wagner, 2011). The ventral aspect of the lateral prefrontal cortex (VLPFC) shows functional connectivity with the NAcc and serves to inhibit and regulate the impulses it generates (Wagner *et al.*, 2013; Chester and DeWall, 2014). The right hemisphere of the VLPFC shows a particular regulatory function (Aron *et al.*, 2004; Chester and DeWall, 2014), though the VLPFC is also associated with other processes such as language generation (Nagel *et al.*, 2008). When

this balance between the VLPFC and NAcc is tipped in favor of the latter, self-regulatory failure occurs (Wagner *et al.*, 2013). The VLPFC has been specifically implicated in the inhibition of aggression (Mehta and Beer, 2010). Aggression most often expresses itself as a self-regulatory failure, when aggressive impulses over-ride self-regulation (Denson *et al.*, 2012). Thus, an imbalance that favors the reward-based processing of the NAcc and disadvantages the regulatory functions of the VLPFC might foster aggression.

Current research

We predicted that after provocation, NAcc activity during aggressive decision-making would be positively correlated with aggression. We did not expect to observe this correlation after no provocation. We also predicted that greater VLPFC-NAcc functional connectivity, representing inhibition of the NAcc, would be negatively correlated with retaliatory aggression. To test these hypotheses, 68 participants underwent fMRI while completing a modified, competitive task against an opponent that allowed them to administer extremely loud noise blasts (as used by Krämer *et al.*, 2007). We then assessed the relation between neural activity when participants were determining how loud of a noise blast to administer and their level of aggression (i.e. the volume of noise blasts) arising from that decision. These analyses were performed separately for aggressive decisions that followed provocation or not.

To assist in the reverse inference that NAcc activity would represent the experience of hedonic reward, 21 participants also completely a personality questionnaire (i.e. the Angry Mood Improvement Inventory; Bushman *et al.*, 2001) which measures the extent to which an individual's aggressive behavior is motivated by the desire to experience positive affect. We hypothesized that NAcc activity would positively correlate with this measure, reflecting that the NAcc activity does indeed reflect a positively valenced, rewarding experience during aggressive acts. See the Supplementary Materials available online for a description of two, large behavioral experiments (combined $N = 908$) that tested this supporting hypothesis.

Methods

Participants

Sixty-nine undergraduates participated in the experiment (68% female; age: $M = 18.70$, $SD = 0.93$). Previous neuroimaging research using this aggression paradigm tested approximately 20 participants (Krämer *et al.*, 2007). However, we sought to recruit far more participants in response to recent criticism of functional neuroimaging research's relatively small sample sizes (Button *et al.*, 2013). We attempted to enroll 80 participants, following the stop rule that participants had to complete the study by the end of the academic year in which they began the study. Participants received course credit and money as compensation. Potential participants were recruited from an introductory psychology subject pool based on their ability to be comfortable and safe in the MRI environment, as well as possessing no pathologies that might impact neural activity.

Materials

Angry mood improvement inventory (AMII). The AMII contains an eight-item subscale of particular relevance to our reverse inference issues with the NAcc, the Anger Expression—Out subscale (Bushman *et al.*, 2001). This subscale assesses the tendency to express angry mood outwardly as aggressive behavior (e.g. I strike out at whatever angers me). Three other eight item subscales measure the tendency to express anger internally, control anger's external expression and control anger's internal expression. Each item refers to a behavior that participants rate along a five-point scale which indicates the degree to which they would like to perform the given behavior *to try and feel better* when they are angry or furious.

Procedure

Participants arrived at our laboratory where they provided informed consent according to guidelines set by the University of Kentucky's Office of Research Integrity and were again screened for safety in the MRI environment. Then, participants completed a battery of questionnaires that included the Angry Mood Improvement Inventory (Bushman *et al.*, 2001) and the item 'have you ever been in a physical fight?' Only the final 21 participants of the study completed the AMII due to a clerical error. Although the AMII has shown excellent internal and test–retest reliability (Bushman *et al.*, 2001), its construct validity remains to be established. More specifically, our subscale of interest, the Anger Expression—Out subscale, has yet to be shown to predict increases in mood (i.e. positive affect) after an aggressive act. Towards this end, we conducted two behavioral experiments to establish the construct validity of this measure (for Methods and Results see Supplemental Materials available online).

Several days after the initial laboratory visit, participants arrived at the University of Kentucky's Magnetic Resonance Imaging and Spectroscopy Center. Participants were told that they would complete a task in the MRI scanner against a same-sex University of Kentucky student who was also in an MRI scanner connected over the internet. Participants then completed the aggression paradigm while undergoing fMRI.

To assess the neural correlates of both retaliatory and non-retaliatory aggression, we employed a version of the classic Taylor Aggression Paradigm that was adapted for the fMRI scanner (Taylor, 1967; Krämer *et al.*, 2007; Dambacher *et al.*, 2015). In this task, participants repeatedly compete against a fictitious opponent to see who can press a button faster, the loser of the competition then receives an aversive noise blast, the volume of which is determined by the opponent. The volume of the noise blast setting is the dependent measure of aggression. The aggression paradigm was implemented as a blocked design that closely matched the parameters of previous fMRI studies that have used this fMRI paradigm (Figure 1). The task was introduced to participants as a competitive reaction-time task in which participants would compete against a same-sex undergraduate student who was connected to them over the internet. Each of the task's 12 blocks began with a fixation trial that modeled baseline neural activity (20 s), which was then followed by the Aggression trial (7.5 s) in which participants set the volume of a noise blast to be delivered to their partner if their partner lost the competition. Volume settings ranged from 1 (not audible) to 4 (aversively loud). Participants then viewed a blank screen with a jittered duration (0.5/1.0/1.5 s) that was replaced by a Competition trial in which participants quickly pressed a button when a red square appeared on the screen (3.5/4.0/4.5 s). Participants then viewed their opponents' pre-programmed volume settings that were categorized as low (i.e. 1, 2) or high (i.e. 3, 4) in provocation (7.5 s). Finally, participants saw whether they won or lost the competition and received a noise blast if they lost (7.5 s). Retaliatory aggression trials were those that

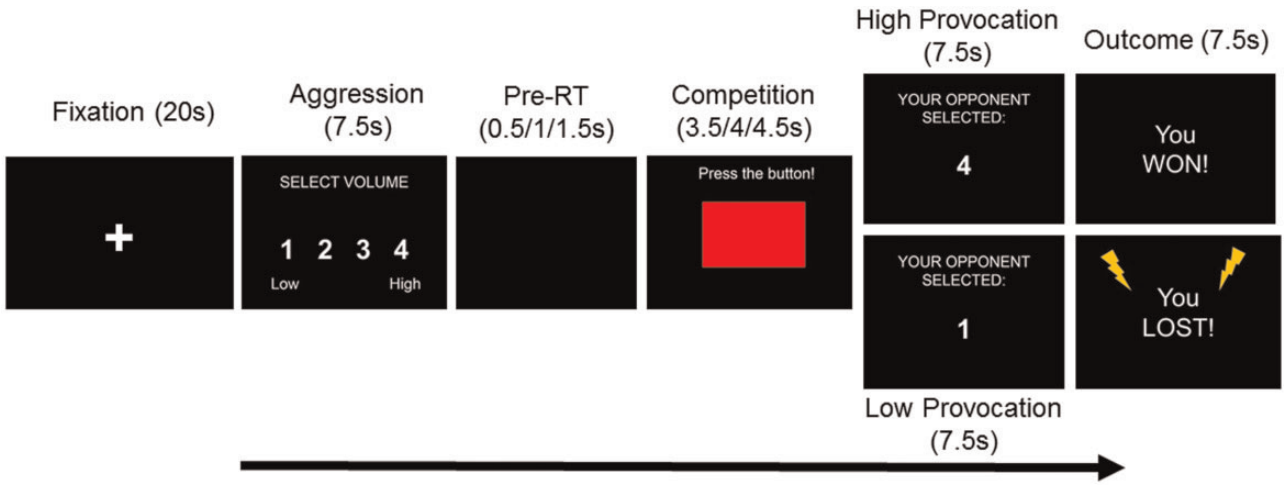

Fig. 1. Schematic of fMRI aggression task.

followed high levels of provocation and non-retaliatory aggression trials were those that followed low levels of provocation. The 12 blocks were characterized by a 5 Retaliatory and 7 non-retaliatory trials that were randomly ordered yet held constant across participants (see Table 1 for order). Wins and losses were also randomized yet held constant across participants.

fMRI data acquisition, preprocessing and analysis

All MRI data were obtained using a 3.0-tesla Siemens Magnetom Trio scanner using a 32-channel head coil. Echo planar BOLD images were acquired with a T2*-weighted gradient across the entire brain with a 3D shim (matrix size = 64 × 64, field of view = 224 mm, echo time = 28 ms, repetition time = 2.5 s, slice thickness = 3.5 mm, 40 interleaved axial slices, flip angle = 90°). To allow for registration to native space, a coplanar T1-weighted MP-RAGE was also acquired from each participant (1 mm^3 isotropic voxel size, echo time = 2.56 ms, repetition time = 1.69 s, flip angle = 12°).

The Oxford Center for Functional MRI of the Brain (FMRIB)'s Software Library (FSL version 5.0) was used to conduct all preprocessing and fMRI analyses (Smith *et al.*, 2004; Woolrich *et al.*, 2009). The first functional volume was removed to facilitate BOLD signal equilibration. Reconstructed functional volumes underwent head motion correction to the middle functional volume. Non-brain tissue was then removed from all functional and structural volumes. Functional volumes underwent slice-timing correction, pre-whitening, spatial smoothing with a 5-mm full width half maximum Gaussian kernel and high-pass temporal filtering (100 s cutoff).

Preprocessed fMRI data from the aggression task were then analyzed using a two-level general linear model approach. First, each participant's BOLD signal was modeled with a fixed effects analysis which modeled aggression trials as events using a canonical double-gamma hemodynamic response function with a temporal derivative. Aggression trials were separately modeled depending on whether they were preceded by a high provocation block (retaliatory aggression) or a low provocation block or no block (non-retaliatory aggression). Aggression trials were not modeled differently if they were preceded or followed by a win or loss because aggression reflects the *attempt* to harm another against their will (Anderson and Bushman, 2002), and thus occurs irrespective of whether the aggressive act (loud noise blast) is realized (win) or not (loss). Competition trials, pre-competition screens, opponent's volume setting trials and outcome trials, along with all six motion parameters were included as nuisance regressors into the model. Fixation trials were not modeled in this analysis. Linear contrasts then compared the each aggression condition to the implicit baseline (i.e. retaliatory aggression > baseline; non-retaliatory aggression > baseline). Resulting contrast images from this analysis were first linearly registered to native space structural volumes and then spatially normalized to a Montreal Neurological Institute stereotaxic space template image.

Table 1. Temporal order of aggression blocks

Block	Aggression type
1	Non-retaliatory
2	Retaliatory
3	Retaliatory
4	Non-retaliatory
5	Non-retaliatory
6	Retaliatory
7	Non-retaliatory
8	Non-retaliatory
9	Retaliatory
10	Retaliatory
11	Non-retaliatory
12	Non-retaliatory

Second, each participant's contrast volumes were fed into a group-level, mixed-effects analysis which created group average maps for both contrasts across the entire brain. Cluster-based thresholding (Worsley, 2001; Heller *et al.*, 2006) was applied to each image (cluster threshold: $Z > 2.3$, $P < 0.05$). Family-wise error correction based on Gaussian random field theory was then applied across the entire brain, Parameter estimates, in percent signal change units, were extracted separately for the left and right NAcc and right VLPFC region-of-interest (ROI) masks. The NAcc ROI masks were constructed from the Wake Forest University Pickatlas (Maldjian *et al.*, 2003) whereas the right VLPFC mask was constructed from the Automated Anatomical Labeling atlas, utilizing the orbital portion of the right inferior frontal gyrus (Tzourio-Mazoyer *et al.*, 2009). Outliers were determined as any datapoints ± 2.5 SDs from the sample mean.

Results

Noise blast settings from the aggression task spanned the four-point scale and showed a normal distribution around the midpoint of the scale (i.e. 2.5), $M = 2.27$, $SD = 0.82$. Noise blasts following high provocation, $M = 2.51$, $SD = 0.94$, were louder than those following low provocation, $M = 2.09$, $SD = 0.80$, $t(68) = 6.46$, $P < 0.001$, $d = 0.78$. No aggression scores extended beyond the ± 2.5 SD outlier cutoff. Of the 69 participants, 20 indicated they had been in a physical fight. Among the subset of the 21 participants who completed the AMII, all subscales of the AMII showed sufficient reliability, Cronbach's αs 0.76–0.90, except for the Anger Expression—In subscale, $\alpha = 0.60$, which was excluded from all subsequent analyses.

ROI analyses: NAcc activity and retaliatory aggression

Percent signal change values were extracted from the left and right NAcc separately across retaliatory and non-retaliatory trials. One female participant was deemed an outlier as her left NAcc activity was 4.10 SDs below the mean and was excluded from analysis. All analyses controlled for the effect of gender, which reliably influences aggression (Archer, 2004). After controlling for gender, left NAcc activity during retaliatory trials was associated with greater aggression, $\beta = 0.27$, $t(65) = 2.24$, $P = 0.029$ (Figure 2). This effect was not observed for the right NAcc, $\beta = 0.12$, $t(65) = 0.98$, $P = 0.329$, during non-retaliatory trials, $\beta = 0.07$, $t(65) = 0.54$, $P = 0.594$, or when the outlier was retained in the analysis, $\beta = 0.20$, $t(66) = 1.67$, $P = 0.099$. Retaliatory aggression's correlation with left NAcc activity during retaliatory trials was significantly stronger than with left NAcc activity during non-retaliatory trials, $Z = 2.23$, $P = 0.026$ (Lee and Preacher, 2013).

Left NAcc activity across all aggression trials was also associated with a substantially greater likelihood of having been in a physical fight, odds ratio (OR) = 38.86, Wald = 5.11, $P = 0.024$, even when the outlier was retained in the analysis, OR = 39.43, Wald = 5.22, $P = 0.022$. This effect was not observed for the right NAcc, OR = 6.20, Wald = 1.41, $P = 0.235$.

Suggesting that the NAcc activity represented a positively valenced experience of reward, individual differences in the tendency to use aggression to experience positive affect (as measured by the Anger Expression—Out subscale of the AMII) were associated with greater bilateral NAcc activity across both retaliatory and non-retaliatory trials, $\beta = 0.65$, $t(14) = 2.69$, $P = 0.018$, after controlling for the effects of gender and the other two AMII subscales. This association was not observed when the outlier was retained in the model, $\beta = 0.47$, $t(15) = 1.66$, $P = 0.118$. As detailed in our Supplemental Materials, the Anger Expression—Out subscale of the AMII is associated with the experience of positive affect in relation to aggression. Thus, it appears that reward-related activity in the NAcc is related to greater aggression both inside and outside the laboratory.

Functional connectivity analyses: NAcc-VLPFC coupling and retaliatory aggression

To assess correlations between aggression and functional connectivity between the right VLPFC and NAcc, we extracted the time series of the aggression task for each participant from the left NAcc and right VLPFC, using the ROI masks described previously. After isolating the timeseries of the Retaliatory and Non-Retaliatory Aggression trials, we correlated the right VLPFC and left NAcc time series yielding a Pearson's r coefficient for each participant and for each condition (as in Chester and DeWall, 2014; Denson *et al.*, 2014). Two female participants were deemed outliers as their functional connectivity estimates were over 2.5 SDs below the sample mean (i.e. −2.85, −2.54 SDs) and were excluded from analysis. After controlling for gender, functional connectivity between the right VLPFC and left NAcc during retaliatory trials was associated with lesser aggression, $\beta = -0.25$, $t(64) = -2.09$, $P = 0.041$ (Figure 3). This effect did not hold for non-retaliatory trials, $\beta = 0.13$, $t(64) = 1.08$, $P = 0.284$, or when the two outliers were retained in the model, $\beta = -0.16$, $t(66) = -1.34$, $P = 0.185$. Retaliatory aggression's correlation with left NAcc activity during retaliatory trials was significantly stronger than with left NAcc activity during non-retaliatory trials, $Z = -2.53$, $P = 0.011$ (Lee and Preacher, 2013). Because both retaliatory and non-retaliatory aggression trials were preceded by a 20 s fixation screen, there is no concern that a previous trial's BOLD response contaminated the connectivity estimates we obtained.

Whole brain analyses

To assess other potential neural correlates of aggression and to provide additional evidence for the association between the NAcc and retaliatory aggression, we conducted two whole brain analyses. In the first analysis, we contrasted neural activation associated with retaliatory > non-retaliatory aggression. However, because these trial types only represented the

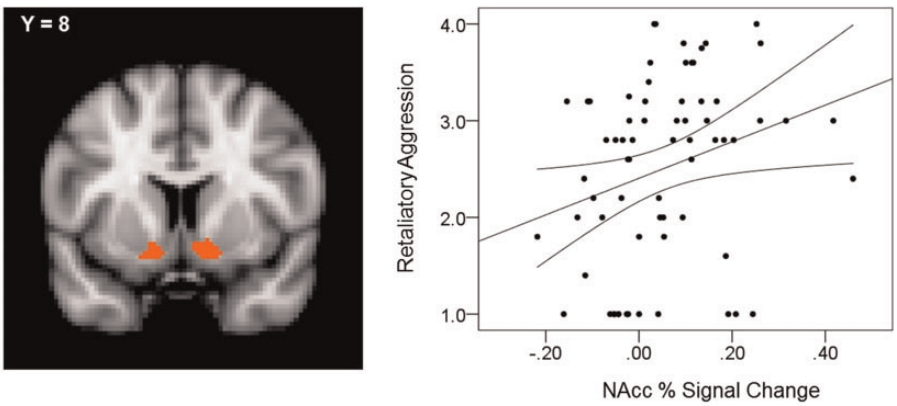

Fig. 2. Region-of-interest masks in red for the left and right NAcc, with Montreal Neurological Institute anatomical coordinates for the y axis. The scatterplot depicts positive association between the volume of noise blasts administered after provocation (i.e. retaliatory aggression) and percent signal change units in the left NAcc during retaliatory aggression trials (controlling for gender). Curved lines represent 95% CI. around the partial regression line.

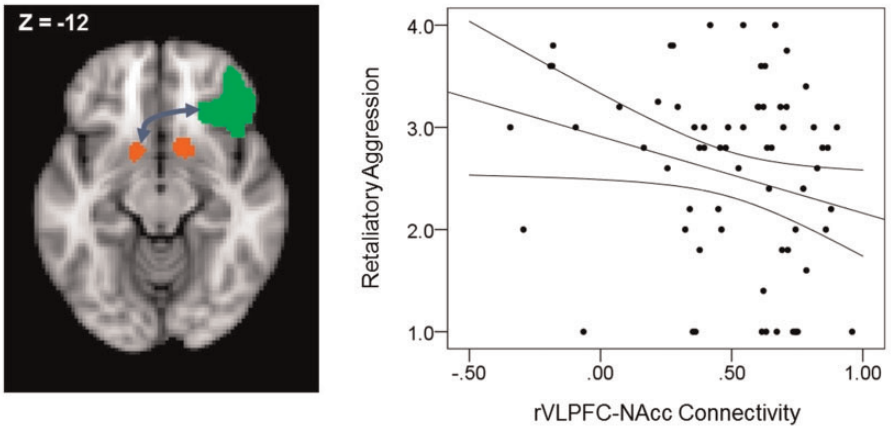

Fig. 3. Region-of-interest masks for the bilateral NAcc (red) and right VLPFC (green). The arrow represents functional connectivity between the left NAcc and right VLPFC. The scatterplot depicts a negative association between the volume of noise blasts administered after provocation (i.e. retaliatory aggression) and functional connectivity estimates between the left NAcc and right VLPFC (controlling for gender). Curved lines represent 95% C.I. around the partial regression line.

opportunity to aggress and did not capture associations with aggression itself, we conducted a whole-brain regression analysis in which retaliatory aggression scores were modeled as a regressor for neural activity during the retaliatory > non-retaliatory aggression contrast. This regression analysis allowed us to assess neural regions that were associated not only with retaliatory aggression trials, but with the actual levels of aggression participants displayed.

Main effect analysis. Retaliatory aggression trials, after being contrasted against non-retaliatory aggression trials, were associated with several large clusters of neural activity (Table 2). We replicated the findings of previous research in that retaliatory aggression trials were associated with activity in the inferior frontal gyrus, posterior cingulate, precentral gyrus, and superior temporal gyrus (Krämer *et al.*, 2007). However, we also observed greater, bilateral activity in the hippocampus that extended rostrally into the amygdala, as well as greater activity in the posterior insula and postcentral gyrus. No activated voxels were observed in the NAcc.

Regression analysis. When retaliatory aggression scores were correlated with brain activity, we saw a very similar pattern of results as to the main effect analyses (Figure 4; Table 3) with large clusters spanning across the amygdala, hippocampus, posterior insula, pre and postcentral gyri and superior temporal gyrus. However, these regression results, when underlaid beneath the main effect contrast activations, revealed additional, previously unobserved clusters in the left NAcc, dorsal striatum, dorsal anterior cingulate cortex and DMPFC (Figure 5).

Table 2. Whole-brain fMRI main effect results from the retaliatory > non-retaliatory aggression contrast (10 329 voxels)

Brain region	Peak Z	Peak MNI coordinates (x,y,z)
Hippocampus/parahippocampal Gyrus	5.17	−56, −14, 0
	4.38	−18, −22, −16
	4.06	36, −30, −14
	4.03	36, −20, −16
	3.94	30, −12, −18
	3.66	40, −42, −16
Inferior frontal gyrus	4.36	−30, 32, −18
Middle temporal gyrus	4.40	−62, −22, −18
Postcentral gyrus	4.83	−20, −38, 72
	3.88	−10, −42, 74
Posterior cingulate cortex	3.88	−10, −22, 44
Precentral gyrus	4.00	−16, −18, 74
	3.86	14, −28, 74
Superior parietal lobule	4.04	−12, −56, 70
Superior temporal gyrus	5.17	−56, −14, 0
	4.45	−58, −4, 6
	3.98	60, 4, −10
	3.60	56, −10, 2

Note: MNI = Montreal Neurological Institute.

Discussion

The damage done by human aggression is well known, yet what motivates this behavior is less understood. Perhaps the best-known causes of aggression are provocation and the negative feelings that result from it (Anderson and Bushman, 2002). However, provocation appears to render retaliatory aggression as a rewarding experience (Ramírez *et al.*, 2005), though a direct test of this hypothesis is lacking. In an effort to empirically substantiate this claim, we sought to implicate the function of the brain's reward circuitry as a substrate of retaliatory aggression. Supporting this prediction, activity in the left NAcc during aggressive decisions was associated with greater aggressive behavior after provocation. This association represents an initial step towards placing hedonic reward as a key contributor to retaliatory aggression.

In addition to this contribution, our findings largely replicated previous research on the neural correlates of retaliatory aggression that included regions of the cingulate cortex, dorsal striatum, insula, lateral and medial aspects of the prefrontal cortex and superior temporal gyri (Krämer *et al.*, 2007; Lotze *et al.*, 2007; Dambacher *et al.*, 2015). The absence of NAcc activity during retaliatory aggression from these previous studies may stem from their relatively smaller sample sizes and the associated ability to detect more subtle, subcortical neural activity. The large clusters of activation that we observed in the hippocampi and DMPFC during retaliatory aggression fit nicely with previous research that implicate these regions as critical neural substrates of anger and angry rumination in response to

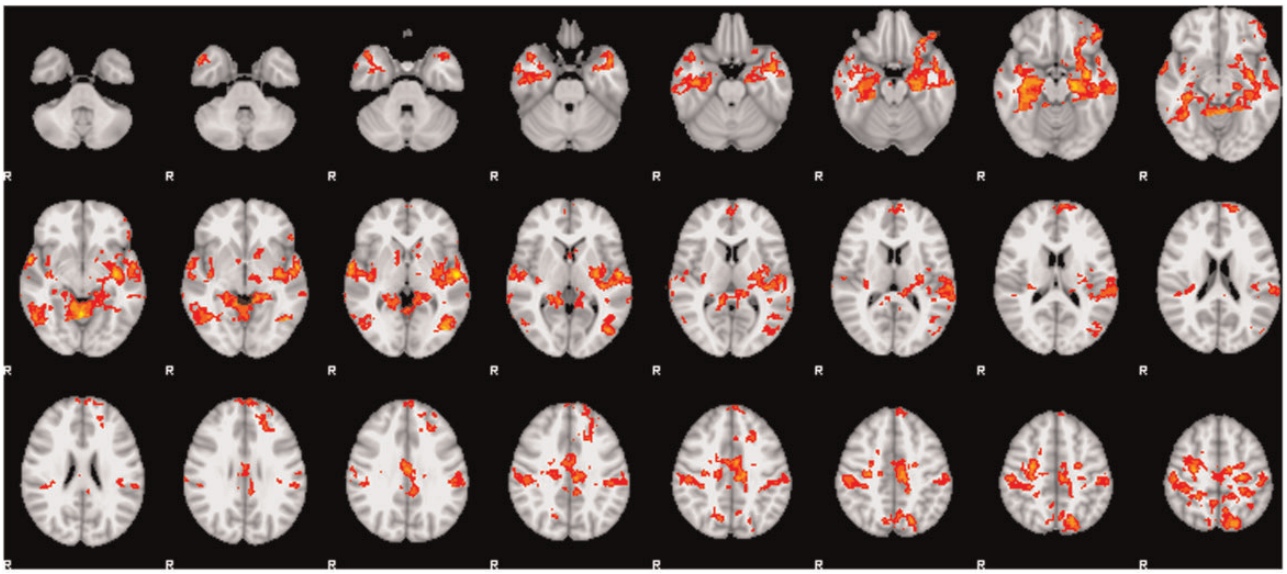

Fig. 4. Whole-brain fMRI regression analyses in which neural activity from the retaliatory > non-retaliatory aggression contrast was regressed onto participants' retaliatory aggression scores.

provocation (Denson *et al.*, 2008) and punishment (Lotze *et al.*, 2007). Future aggression research will benefit greatly from exploring the specific role that each of these regions play in motivating and regulating aggressive behavior. Modulation of these various regions through brain stimulation and psychoactive drugs presents itself as a promising avenue to disentangle their various contributions.

If greater NAcc activity promotes aggression, then the inhibition of this region should also predict inhibited aggression. We demonstrated that the more that participants' left NAcc showed functional connectivity with the right VLPFC during aggressive decisions, the less retaliatory aggression they perpetrated. This finding supports the central tenets of balance theory that construes self-regulatory failures (e.g. aggression) as arising from an imbalance in the brain that favors subcortical structures in relation to their regulatory counterparts in the prefrontal cortex (Heatherton and Wagner, 2011). Although we made a bilateral prediction regarding the NAcc, the effects we observed were specific to the left hemisphere. The specificity of our associations with the left and not the right NAcc are possibly due to the lateralization of positive valence to the dominant, left hemisphere even among such subcortical structures as the NAcc (Kühn and Gallinat, 2012; Berridge and Kringelbach, 2013).

This emphasis on hedonic reward as a potent contributor to aggression stands in stark contrast with conventional approaches to aggression research that emphasize the role of negative affect and aversive states like heat (Anderson, 1989; Berkowitz, 1989). However, our results are not meant to contend with such models, rather we hope that they add nuance to them. Consistent with previous research showing that aggression is often perceived and utilized as a means to alleviate negative affect (Bushman *et al.*, 1999, 2001), we expect that negative affect may motivate individuals to seek out sources of hedonic reward to regain affective homeostasis. As our results suggest, aggression may be such a perceived source of pleasure. These findings have clear implications for treatments and interventions that target reducing aggression. Indeed, if aggression is motivated by reward then such treatments should adopt practices from addiction treatment models that often seek to mitigate the role of cravings and anticipated reward.

Table 3. Whole-brain fMRI regression results, in which neural activity from the retaliatory > non-retaliatory aggression contrast was regressed onto retaliatory aggression scores (25 651 voxels)

Brain region	Peak Z	Peak MNI coordinates (x, y, z)
Cerebellum	4.93	2, −58, −8
Inferior occipital lobe	4.54	−42, 74, 0
Middle frontal gyrus	3.35	−24, 32, 30
Nucleus accumbens	2.85	−12, 4, −10
Postcentral gyrus	4.91	−22, −38, 72
	4.56	−36, −32, 64
Posterior insula	4.69	−36, −12, −8
Superior frontal gyrus/frontal pole	3.68	−22, 22, 38
	3.53	−4, 64, 30
	3.46	−22, 38, 28
	3.32	−6, 64, 16
	3.21	−18, 62, 20
Superior temporal gyrus	5.42	−56, −12, 2

Note: MNI = Montreal Neurological Institute.

A lingering question is whether the NAcc activity, we observed represents currently-felt reward, anticipated reward or some other psychological process altogether. Issues with reverse inference that are inherent in functional neuroimaging make this a difficult inquiry to resolve (Poldrack, 2006). It deserves to be noted that reverse inference is a problematic aspect of almost all research (e.g. does a self-report of anger truly represent the experience of that process?). Indeed, NAcc activity is not a pure neural signature of reward, given the role of the NAcc in broader learning and motivational circuitry (Floresco, 2015). Yet a wealth of meta-analytic evidence from hundreds of functional neuroimaging studies suggests that the NAcc is most reliably associated with reward (Diekhof *et al.*, 2012; Kühn and Gallinat, 2012; Bartra *et al.*, 2013; Berridge and Kringelbach, 2013). To support this reverse inference within our own study, we found that, in the context of our aggression task, NAcc activity represented positively valenced reward in the form of the observed correlation between

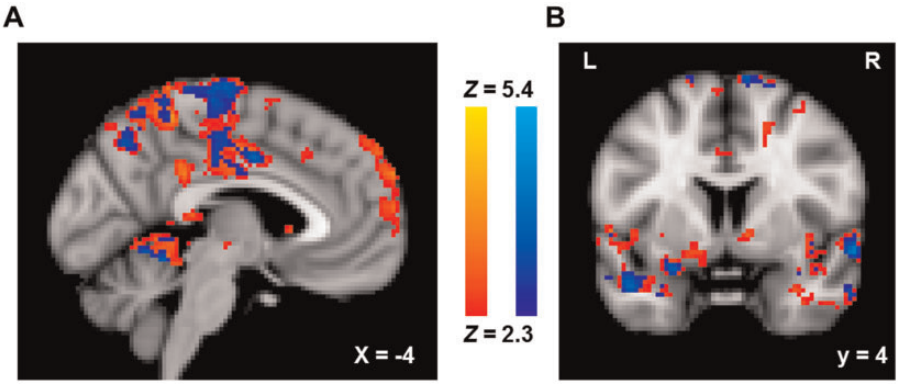

Fig. 5. Whole-brain fMRI analysis for the retaliatory > non-retaliatory aggression contrast in which red voxels represent neural activity regressed onto retaliatory aggression scores and blue voxels represent overlap with the main effect of trial type. Clusters that are specific to the regression analysis are depicted in both a (**A**) sagittal view demonstrating an MPFC cluster and a (**B**) coronal view depicting a cluster in the left NAcc.

NAcc activity and the tendency to use aggression to experience positive affect. This correlation suggests that, NAcc activity represented the anticipation of positive affect. Indeed, research on the NAcc appears to implicate it as being associated with the anticipation of a rewarding stimulus more than the current, hedonic sensation (Berridge and Kringelbach, 2013). As such, the NAcc activity we observed may represent the expected reward of aggression and less so the currently felt level of reward. This anticipatory function of the NAcc and its relation to aggression fits nicely with recent meta-theoretical evidence that anticipated emotion is a potent motivator of social behaviors (Baumeister *et al.*, 2007; DeWall *et al.*, in press).

We cannot disentangle the extent to which the NAcc activity we observed during retaliatory aggression was due to the reward of aggression itself, or the outcome of seeing the provocateur punished for their incendiary acts. Previous research linking the act of simply observing the punishment of provocateurs has also shown NAcc reactivity (Singer *et al.*, 2006; Krämer *et al.*, 2007). Future research should seek to disentangle the rewarding nature of aggression per se and the achievement of its intended outcome. Additionally, we cannot be sure that activity in the rVLPFC represented self-regulatory processes as this region has also been implicated in such psychological processes as language generation (Nagel *et al.*, 2008).

Yet why would these effects be observed only for retaliatory aggression? The behavioral literature shows very clearly that aggression is only rated as pleasant when it occurs after a provocation (Carré *et al.*, 2010). We speculate that this specificity of reward's association with retaliatory aggression is likely due to evolutionary forces that selected for a motivational system that spurred individuals towards inflicting reciprocal costs on those who reduced their reproductive fitness (McCullough *et al.*, 2013). Implicating the motivation to maintain justice, the positive affect associated with 'righting a wrong' appears to motivate aggressive behavior (Gollwitzer and Bushman, 2012). Thus, ancient retributive motivations and more modern desires for equity may underpin the specificity of reward in motivation retaliatory aggression. However, the ability of self-regulation to undermine the effect of reward on aggression offers hope for reducing human violence.

Acknowledgements

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are grateful for the assistance of Richard Milich and Donald Lynam for their assistance in developing and conducting this study.

Funding

This experiment was funded by a grant from the University of Kentucky's Center for Drug Abuse Research Translation (CDART; Sponsor: National Institute on Drug Abuse, Grant number: DA005312) to the last author.

Supplementary data

Supplementary data are available at *SCAN* online.

Conflict of interest. None declared.

References

Anderson, C.A. (1989). Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological Bulletin*, **106**, 74–96.

Anderson, C.A., Bushman, B.J. (1997). External validity of "true" experiments: The case of laboratory aggression. *Review of General Psychology*, **1**, 22.

Anderson, C.A., Bushman, B.J. (2002). Human aggression. *Annual Review of Psychology*, **53**, 27–51.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of General Psychology*, **8**, 291–322.

Aron, A.R., Robbins, T.W., Poldrack, R.A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, **8**, 170–7.

Bartra, O., McGuire, J.T., Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, **76**, 412–27.

Baumeister, R.F., Vohs, K.D., DeWall, C.N., Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, **11**, 167–203.

Berkowitz, L. (1989). Affective aggression: The role of stress, pain, and negative affect. In: Geen, R.G., Donnerstein, E., editors. *Human Aggression: Theories, Research, and Implications for Social Policy*. San Diego: Academic Press, pp. 49–72.

Berridge, K.C., Kringelbach, M.L. (2013). Neuroscience of affect: brain mechanisms of pleasure and displeasure. *Current Opinion in Neurobiology*, **23**, 294–303.

Bushman, B.J. (2002). Does venting anger feed or extinguish the flame? catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, **28**, 724–31.

Bushman, B.J., Baumeister, R.F., Phillips, C.M. (2001). Do people aggress to improve their mood? Catharsis beliefs, affect regulation opportunity, and aggressive responding. *Journal of Personality and Social Psychology*, **81**, 17–32.

Bushman, B.J., Baumeister, R.F., Stack, A.D. (1999). Catharsis, aggression, and persuasive influence: Self-fulfilling or self-defeating prophecies? *Journal of Personality and Social Psychology*, **76**, 367–76.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., *et al.* (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, **14**, 365–76.

Carré, J.M., Gilchrist, J.D., Morrissey, M.D., McCormick, C.M. (2010). Motivational and situational factors and the relationship between testosterone dynamics and human aggression during competition. *Biological Psychology*, **84**, 346–53.

Chase, H.W., Eickhoff, S.B., Laird, A.R., Hogarth, L. (2011). The neural basis of drug stimulus processing and craving: an activation likelihood estimation meta-analysis. *Biological Psychiatry*, **70**, 785–93.

Chester, D.S., DeWall, C.N. (2014). Prefrontal recruitment during social rejection predicts greater subsequent self-regulatory imbalance and impairment: neural and longitudinal evidence. *NeuroImage*, **101**, 485–93.

Cikara, M., Fiske, S.T. (2011). Bounded Empathy: Neural Responses to Outgroup Targets' (Mis)fortunes. *Journal of Cognitive Neuroscience*, **23**, 3791–803.

Dambacher, F., Sack, A.T., Lobbestael, J., Arntz, A., Brugman, S., Schuhmann, T. (2015). Out of control: evidence for anterior insula involvement in motor impulsivity and reactive aggression. *Social Cognitive and Affective Neuroscience*, **10**, 508–16.

Denson, T.F., DeWall, C.N., Finkel, E.J. (2012). Self-control and aggression. *Current Directions in Psychological Science*, **21**, 20–5.

Denson, T.F., Dobson-Stone, C., Ronay, R., von Hippel, W., Schira, M.M. (2014). A Functional Polymorphism of the MAOA Gene Is Associated with Neural Responses to Induced Anger Control. *Journal of Cognitive Neuroscience*, **26**, 1418–27.

Denson, T.F., Pedersen, W.C., Ronquillo, J., Nandy, A.S. (2008). The Angry Brain: Neural Correlates of Anger, Angry Rumination, and Aggressive Personality. *Journal of Cognitive Neuroscience*, **21**, 734–44.

De Quervain, D.J.-F., Fischbacher, U., Treyer, V., *et al.* (2004). The neural basis of altruistic punishment. *Science (New York, N.Y.)*, **305**, 1254–58.

DeWall, C.N., Baumeister, R.F., Chester, D.S., Bushman, B.J. (in press). How often does currently felt emotion predict social behavior and judgment? A meta-analytic test of two theories. *Emotion Review*.

DeWall, C.N., Finkel, E.J., Lambert, N.M., *et al.* (2013). The voodoo doll task: Introducing and validating a novel method for studying aggressive inclinations. *Aggressive Behavior*, **39**, 419–39.

Diekhof, E.K., Kaps, L., Falkai, P., Gruber, O. (2012). The role of the human ventral striatum and the medial orbitofrontal cortex in the representation of reward magnitude – An activation likelihood estimation meta-analysis of neuroimaging studies of passive reward expectancy and outcome processing. *Neuropsychologia*, **50**, 1252–66.

Giancola, P., Chermack, S. (1998). Construct validity of laboratory aggression paradigms: A response to Tedeschi and Quigley (1996). *Aggression and Violent Behavior*, **3**, 16.

Gollwitzer, M., Bushman, B.J. (2012). Do Victims of Injustice Punish to Improve Their Mood? *Social Psychological and Personality Science*, **3**, 572–80.

Grahn, J.A., Parkinson, J.A., Owen, A.M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, **86**, 141–55.

Gray, J.A. (1994). Three fundamental emotion systems. In: Ekman, P., Davidson, R.J., editors. The Nature of Emotion: *Fundamental Questions*. New York: Oxford University Press, pp. 243–7.

Graybiel, A.M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, **31**, 359–87.

Harmon-Jones, E., Sigelman, J. (2001). State anger and prefrontal brain activity: Evidence that insult-related relative left-prefrontal activation is associated with experienced anger and aggression. *Journal of Personality and Social Psychology*, **80**, 797–803.

Heatherton, T.F., Wagner, D.D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences*, **15**, 132–39.

Heller, R., Stanley, D., Yekutieli, D., Rubin, N., Benjamini, Y. (2006). Cluster-based analysis of fMRI data. *NeuroImage*, **33**, 599–608.

Krämer, U.M., Jansma, H., Tempelmann, C., Münte, T.F. (2007). Tit-for-tat: The neural basis of reactive aggression. *NeuroImage*, **38**, 203–11.

Kühn, S., Gallinat, J. (2012). The neural correlates of subjective pleasantness. *NeuroImage*, **61**, 289–94.

Lee, I.A., Preacher, K.J. (2013, September). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software].

Lotze, M., Veit, R., Anders, S., Birbaumer, N. (2007). Evidence for a different role of the ventral and dorsal medial prefrontal cortex for social reactive aggression: An interactive fMRI study. *NeuroImage*, **34**, 470–8.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, **19**, 1233–39.

McCullough, M.E., Kurzban, R., Tabak, B.A. (2013). Cognitive systems for revenge and forgiveness. *Behavioral and Brain Sciences*, **36**, 1–15.

Mehta, P.H., Beer, J. (2009). Neural mechanisms of the testosterone–aggression relation: the role of orbitofrontal cortex. *Journal of Cognitive Neuroscience*, **22**, 2357–68.

Nagel, I.E., Schumacher, E.H., Goebel, R., D'Esposito, M. (2008). Functional MRI investigation of verbal selection mechanisms in lateral prefrontal cortex. *NeuroImage*, **43**, 801–7.

Poldrack, R.A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, **10**, 59–63.

Ramírez, J.M., Bonniot-Cabanac, M.-C., Cabanac, M. (2005). Can Aggression Provide Pleasure? *European Psychologist*, **10**, 136–45.

Shohamy, D. (2011). Learning and motivation in the human striatum. *Current Opinion in Neurobiology*, **21**, 408–14.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature*, **439**, 466–9.

Smith, S.M., Jenkinson, M., Woolrich, M.W., *et al.* (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, **23**, S208–19.

Strobel, A., Zimmermann, J., Schmitz, A., *et al.* (2011). Beyond revenge: Neural and genetic bases of altruistic punishment. *NeuroImage*, **54**, 671–80.

Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., Okubo, Y. (2009). When your gain is my pain and your pain is my gain: neural correlates of Envy and Schadenfreude. *Science*, **323**, 937–39.

Taylor, S. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality*, **35**, 297–310.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., *et al.* (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**, 273–89.

Wagner, D.D., Altman, M., Boswell, R.G., Kelley, W.M., Heatherton, T.F. (2013). Self-regulatory depletion enhances neural responses to rewards and impairs top-down control. *Psychological Science*, **24**, 2262–71.

Woolrich, M.W., Jbabdi, S., Patenaude, B., *et al.* (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, **45**, S173–86.

Worsley, K.J. (2001). Statistical analysis of activation images. *Functional MRI: an Introduction to Methods*, **14**, 251–70.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, **8**, 665–70.