A neurobiological association of revenge propensity during intergroup conflict

Xiaochun Han[a], Michele J. Gelfand[b], Bing Wu[c], Ting Zhang[a], Wenxin Li[d], Tianyu Gao[a], Chenyu Pang[a], Taoyu Wu[a], Yuqing Zhou[a], Shuai Zhou[c], Xinhuai Wu[c,1], Shihui Han[a,1]

[a]School of Psychological and Cognitive Sciences, PKU-IDG/McGovern Institute for Brain Research, Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

[b]Department of Psychology, University of Maryland, USA

[c]Department of Radiology, The 7th Medical Center of PLA General Hospital, Beijing, China

[d]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China

Address correspondence to:
Shihui Han Ph.D. Prof.
School of Psychological and Cognitive Sciences
Peking University
52 Haidian Road, Beijing 100080, China
Phone: (86) 10-6275-9138
Email: shan@pku.edu.cn
or
Xinhuai Wu. M.D.
Department of Radiology
PLA Army General Hospital
5 Nanmencang, Beijing, 100070, China
Phone: (86) 10-8400-8368
Email: bei925@sina.com

Abstract

Revenge during intergroup conflict is a human universal, but its neurobiological underpinnings remain unclear. We address this by integrating functional MRI and measurements of endogenous oxytocin in participants who view an ingroup and an outgroup member's suffering that is caused mutually (Revenge group) or respectively by a computer (Control group). We show that intergroup conflict encountered by the Revenge group is associated with an increased level of oxytocin in saliva compared to in the Control group. Furthermore, the medial prefrontal activity in response to ingroup pain in the Revenge but not Control group mediates the association between endogenous oxytocin and the propensity to give painful electric shocks to outgroup members regardless of whether they were directly involved in the conflict. Our findings highlight an important neurobiological correlate of revenge propensity which may be implicated in conflict contagion across individuals in the context of intergroup conflict.

Revenge, which refers to taking actions of harming someone in retaliation for an injury (*Elshout et al., 2015; Jackson et al., 2019*), is a global phenomenon and a causal factor in many homicides and transgenerational conflicts (*Kopsaj, 2016; Jackson et al., 2019*). While revenge is an aggressive act, not all aggressive acts represent vengeance. For example, unsolicited acts of aggression, like deviance, incivility, and bullying, would not count as revenge (*Raver and Barling, 2008; Jackson et al., 2019*). Revenge often occurs between families or clans when an outgroup member brings harm to an ingroup member which, in turn, induces retaliation upon outgroup members (*Ericksen and Horton, 1992*). According to early social psychological theories (*Allport et al., 1954; Brewer, 1999*), a desire to help the ingroup ("ingroup love") and/or an aggressive motivation to hurt the outgroup ("outgroup hate") may drive participation in intergroup conflict by taking revenge. Recent behavioral research using economic games indeed suggests that ingroup love plays a key role in driving economic punishment toward outgroup (*Halevy et al., 2008; de Dreu, 2010; De Dreu et al., 2010; Halevy et al., 2012*). Yet despite the severe social consequences of revenge, its neurobiological underpinnings remain unclear. Building upon previous findings (*Halevy et al., 2008; de Dreu, 2010; De Dreu et al., 2010; Halevy et al., 2012*), we suggest that there may be a neurobiological mechanism that links perceived ingroup pain caused by an outgroup and the propensity to seek revenge upon an outgroup during intergroup conflict. The present work specifically examined the hormonal (i.e., oxytocin) and neural responses to ingroup suffering caused by an outgroup that predict revenge propensity against outgroups.

Previous brain imaging research has revealed neural responses to ingroup/outgroup members' suffering, yet they have been done in contexts that lack the key character of real-life intergroup conflict, i.e., ingroup and outgroup members causing each others' pain. Functional magnetic resonance imaging (fMRI) studies have identified increased activity in both the empathy network (e.g., the anterior midcingulate (aMCC) and anterior insula (AI))

and theory-of-mind network (e.g., the medial prefrontal cortex (mPFC) and temporoparietal junction (TPJ)) in response to ingroup pain (*Hein et al., 2010; Cikara et al., 2011; Han, 2018*). Outgroup pain, on the other hand, is related to enhanced activity in the reward system (e.g., the ventral striatum and nucleus accumbens, *Hein et al., 2010; Cikara et al., 2011; Luo et al., 2015*). In addition, the mPFC activity in response to perceived pain is associated with decisions to help ingroup members (*Hein et al., 2010; Mathur et al., 2010*), and the activity in the nucleus accumbens predicts decisions not to help outgroup members (*Hein et al., 2010; Luo et al., 2015*). These findings highlight ingroup favoritism in brain responses to others' pain as neural underpinnings of ingroup love but leave open a critical question: Are brain responses to ingroup pain inflicted by outgroup members during intergroup conflict associated with subsequent revenge? Specifically, it is unclear whether activities in the empathy and/or theory-of-mind networks in response to perceived ingroup pain are associated with revenge motives during intergroup conflict. If revenge aims to bring suffering to an outgroup to get reward during intergroup conflict, one may expect the involvement of the reward system in decision making related to outgroup punishment (*Hein et al., 2010; Cikara et al., 2011; Han, 2018*). However, if the goal of revenge is to help ingroup members who suffer from physical harm caused by an outgroup (*Lickel et al., 2006*), the mPFC, which responds to ingroup pain and is associated with ingroup help (*Hein et al., 2010; Mathur et al., 2010*), may be associated with tendencies to punish the outgroup.

Previous fMRI research has also examined neurobiological correlates of punishment decision-making pertaining to those who have violated social norms in economic games (*Seymour et al., 2007; Krueger and Hoffman, 2016*). Punishment decisions to prevent social norm violations have been associated with increased activities in both the empathy and theory-of-mind networks including the aMCC, AI, and mPFC (*Krueger and Hoffman, 2016*). Yet these studies focused on brain activities related to punishment decisions rather than

neurobiological mechanisms that link perceived ingroup suffering to propensity to seek revenge upon outgroups. Punishment decisions during previously studied economic games were likely motivated by prevention of social norm violations rather than by perceived physical harm to ingroup members caused by outgroup members, which characterizes most of revenge behavior in real-life situations.

Finally, at the hormone level, recent research reported increased levels of urinary oxytocin (OT) — a nine amino acid peptide synthesized in hypothalamic cells — in Chimpanzees immediately before and during border patrols and intergroup encounters (*Samuni et al., 2017*). Likewise, intranasal administration of OT (vs. placebo) in humans enhanced empathic neural responses to ingroup pain (*Sheng et al., 2013*) and individuals' contributions to ingroup payoffs (*De Dreu et al., 2010; de Dreu, 2010*). OT administration also promotes motivation to sacrifice outgroup targets (*De Dreu et al., 2011*) and facilitates within-group coordination for successful outgroup attack during economic games (*Zhang et al., 2019*). These findings shed light on a functional role of the oxytocinergic system in decision making related to outgroup punishment. However, there has been little direct evidence for modulations of endogenous OT in humans during intergroup conflict (but see Levy et al., 2016). In addition, neural architectures that mediate endogenous OT and revenge propensity during intergroup conflict have been largely unexplored. Among the brain regions in which activities are sensitive to ingroup pain, the mPFC contains OT-sensitive neurons (*Ninan, 2011*). The mPFC, cingulate, and insula express OT receptors (*Gimpl and Fahrenholz, 2001; MacDonald and MacDonald, 2010; Boccia et al., 2013*) and mPFC/aMCC activities are modulated by administered OT (*Sabihi et al., 2014; Eckstein et al., 2015; Liu et al., 2017; Wang et al., 2017*). However, to date, whether the neural systems involved in empathy or theory-of-mind link endogenous OT to revenge propensity during intergroup conflict has yet to be examined.

A key challenge to empirically address these issues is the need for an experimental paradigm of intergroup conflict that can be used in a neuroimaging laboratory setting to measure neurobiological responses to perceived ingroup physical pain caused by an outgroup and revenge propensity to bring physical harm to the outgroup. Another challenge for empirical research on the neurobiological association of revenge propensity during intergroup conflict is to disentangle the effect of the key component of revenge (i.e., to punish outgroup members for their harming to one's ingroup) from other concomitant but nonessential factors, including perceived group identity (Kahn et al., 2017), negative evaluation of the outgroup (*Schiller et al., 2014*), and decreased empathy for outgroup pain (*Hein et al., 2010; Cikara et al., 2011; Han, 2018*). These factors themselves may lead to negative treatment of outgroup members based on ingroup biases in social behavior that occur even in the absence of intergroup conflict. It is therefore necessary to examine neurobiological responses in two conditions in which ingroup biases in emotions, attitudes, and behavior are matched but the motive to punish the outgroup is different. That is, in the revenge condition, individuals punish outgroup members because they bring physical harm to the ingroup, whereas in the control condition, all else being equal, individuals punish outgroup members to show their ingroup favoritism in the absence of perceived intergroup conflict.

Toward these ends, we developed a new neural-behavioral paradigm that simulates real-life revenge during intergroup conflict. In this paradigm participants viewed an ingroup and an outgroup member who gave each other painful electric shocks (Revenge group) or received electric shocks given by a computer (Control group) during a competitive game. The key difference between the two conditions is whether outgroup members brought physical harm to ingroup members during an intergroup conflict while all other aspects of the experimental manipulations were the same for the two groups. We measured participants' salivary levels of OT, brain responses to perceived ingroup pain, and revenge propensity to

bring physical harm to the outgroup. These measures allowed us to investigate the neurobiological correlates of how harm to an ingroup member caused by an outgroup member inspires an uninvolved ingroup member to punish outgroup members.

Based on the findings of increased endogenous OT immediately before and during border patrols and intergroup encounters in Chimpanzees (*Samuni et al., 2017*), we hypothesized that endogenous OT in humans is also sensitive to intergroup conflict and that salivary levels of OT would increase in the Revenge compared to Control groups. Since brain regions in both the empathy and theory-of-mind network express OT receptors (*Gimpl and Fahrenholz, 2001; MacDonald and MacDonald, 2010; Boccia et al., 2013*) and activities in both networks are modulated by administered OT (*Sabihi et al., 2014; Eckstein et al., 2015; Liu et al., 2017; Wang et al., 2017*), activities in both networks in response to ingroup members' pain may be associated with endogenous OT in the context of intergroup conflict that involves physical harm. We tested this hypothesis by conducting whole-brain analyses of neural responses to perceived ingroup pain. Specifically, we searched for brain regions in which salivary levels of OT predicted activity in response to ingroup pain caused by outgroup. Our whole-brain analyses revealed that salivary levels of OT were associated with mPFC activity in the Revenge group. Accordingly, we further tested whether the mPFC activity predicted propensity to punish outgroup members and mediated the association between endogenous OT and revenge propensity. The results of these analyses allowed us to test the association between endogenous OT and mPFC responses to ingroup pain as a neurobiological correlate of revenge propensity during intergroup conflict. To provide a broad test of the neurobiological underpinnings of revenge propensity, we also examined whether the tendency to retaliate against outgroup members who are not directly involved in the conflict, or what has been termed "vicarious retribution" (*Lickel et al., 2006; Gelfand et al., 2012; Lee et al., 2013*), has the same neurobiological association.

In what follows, after describing our experimental design, we first present behavioral results that show comparable ingroup biases in emotions and attitudes in the Revenge and Control groups We then examined whether the Revenge compared to the Control group showed higher endogenous OT levels after witnessing ingroup members' pain caused by outgroup members. Thereafter, we report results of whole-brain analyses that identified neural responses to outgroup-inflicted ingroup pain that were predicted by salivary levels of OT. Finally, we report evidence for the association between revenge propensity and mPFC activity in response to ingroup pain caused by the outgroup, as well as evidence that the mPFC activity mediates the association between endogenous OT and revenge propensity. These results together suggest an association between endogenous OT and mPFC activity in response to ingroup pain as a neurobiological correlate of revenge propensity during intergroup conflict.

Results

Behavioral paradigm

We developed a new behavioral paradigm to examine neurobiological associations of the key component of revenge behavior (i.e., to punish outgroup in response to physical harm to ingroup caused by outgroup) while the effects of ingroup biases in emotions and attitudes were controlled. We adopted a between-subjects design by recruiting two independent samples of healthy adults to test our hypotheses.

The paradigm for Revenge group (n=40, all males) had three phases and 6 players (4 participants and 2 confederates). In *Phase 1*, the 6 players played a game to form an ingroup and an outgroup. Each group consisted of one confederate and two participants (see Figure 1A). *Phase 2* introduced initial conflict by inviting the participants to watch an ingroup member (Involved_Ingroup target) and an outgroup member (Involved_Outgroup target), both played by the confederates, interact in a competitive game during which the winner gave

painful or non-painful electric shocks to the rival. In *Phase 3*, the participants underwent fMRI scanning. In four scans, they were informed that Involved_Ingroup and Involved_Outgroup targets continued the competitive game and applied shocks to each other. On each trial the participants first viewed a photo of the Involved_Ingroup or Involved_Outgroup target to indicate the loser of one trial and had to judge his group identity (i.e., ingroup or outgroup) by pressing a button. A lightning (or round) symbol was then displayed to inform the winner's decision of giving a painful (or non-painful) shock followed by the target's face with painful or neutral expressions to indicate that the target was experiencing a painful or non-painful shock (Figure 1B). Because group members during intergroup conflict are often regarded as an entity of interchangeable members (*Lickel et al*., *2006; Lee et al*., *2013*), we also examined generic neurobiological associations of tendencies to retaliate against outgroup members regardless of their direct involvement in conflict. To this end, in additional four fMRI scans, the participants were presented with photos of an ingroup and an outgroup member who were not directly involved in the conflict (Uninvolved_Ingroup and Uninvolved_Outgroup targets) and judged their group identity before viewing their painful or neutral expressions. After fMRI scanning the participants were asked to report how willing they were to punish a target by giving painful shocks (1 = not painful at all, 9 = extremely painful) to estimate their tendencies to punish outgroup members.

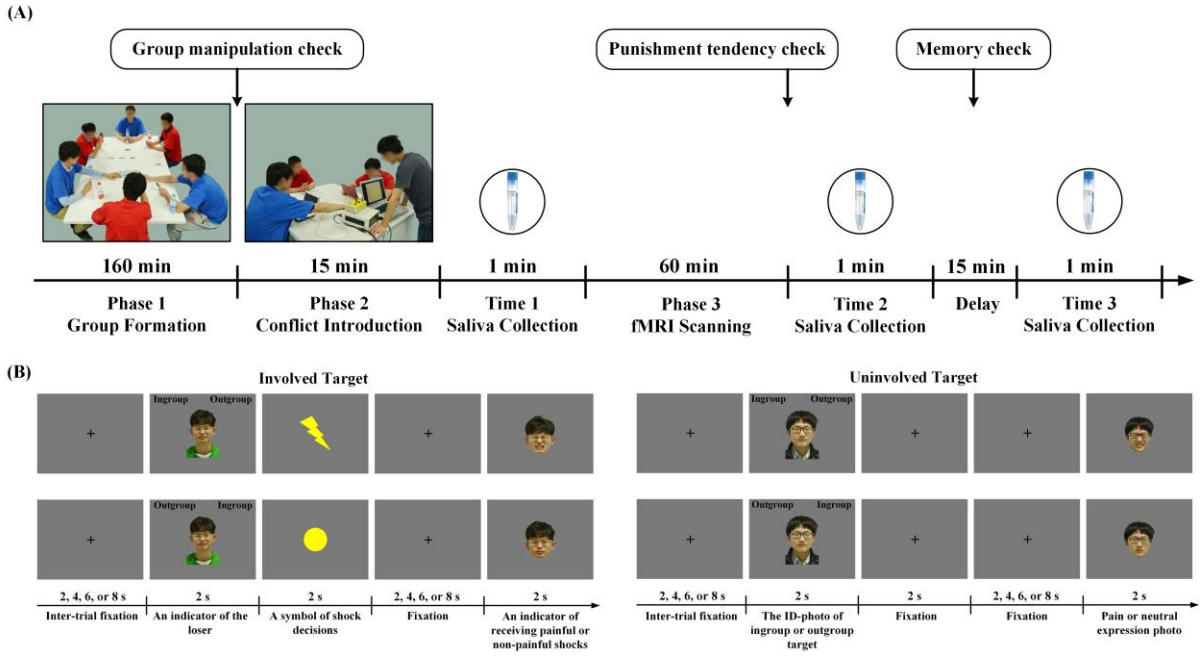

Figure 1. Experimental procedure and behavioral results. (A) Experimental procedure. Phase 1 assigned 4 participants and 2 confederates into two groups who played a game to create group affiliation. In Phase 2 the experimenter (in a grey T-shirt) introduced a participant (in a red T-shirt facing the experimenter) to witness a conflict between an ingroup member and an outgroup member (both played by the confederates) who played a competitive game. During fMRI scanning (Phase 3) the participant witnessed ingroup and outgroup members who were directly involved or uninvolved in conflict. Saliva was collected at three points in time. (B) Trial structure during fMRI scanning. An ID-photo of Involved_Ingroup or Involved_Outgroup target indicated the loser of the game and the participant had to judge his group identity. A yellow circle or lightning symbol then indicated a non-painful or painful shock. After a fixation, a photo of the loser's face with painful or neutral expression was displayed to indicate that he was experiencing a painful or non-painful shock. When ID-photos of uninvolved targets were presented, the participant also judged their group identities and passively viewed a following photo of the target with neutral or painful expression.

We recruited a Control group (n=40, all males) to control for the effects of perceived group identity, ingroup biases in emotions and attitudes, and ingroup favoritism in empathic brain activity on the potential neurobiological association of revenge propensity. The scenario for the Control group was the same as that for Revenge group except that, during Phases 2 and 3, the participants were informed that Involved_Ingroup and Involved_Outgroup targets played a competitive game with a computer, respectively, and received painful or non-painful electric shocks given by the computer. Thus, an outgroup bias in tendency to apply painful electric shocks was driven by outgroup derogation in the Control group but by revenge in return for ingroup members' suffering produced by outgroup in the Revenge group.

We collected saliva from both Revenge and Control groups at three points in time (i.e., Time 1: after introduction of initial intergroup conflict; Time 2: outside the scanner immediately after fMRI scanning; Time 3: 15 minutes later, Figure 1A) to estimate changes of endogenous OT. By comparing OT results of the Revenge and Control groups we sought to determine whether endogenous OT increases immediately after initially witnessing an intergroup conflict (Time 1), and whether such effects, if observed, would be enlarged after additional experiences of intergroup conflict (Time 2). We also assessed whether the level of endogenous OT predicted brain responses to perceived ingroup suffering during intergroup conflict. Finally, we examine whether brain responses to ingroup pain mediate the association between endogenous OT and individuals' inclinations to seek revenge by giving painful electric shocks to outgroup members.

Punishment tendencies in Revenge and Control groups

Revenge and Control groups were matched in age, education, and psychological traits (see Supplemental file 1 for the demographic information and psychological traits of the participants). In order to assess the effect of the key component of revenge (i.e., to punish outgroup members in return for their harm to ingroup), we controlled other concomitant but

nonessential factors by collecting self-reports of emotions, attitudes, and punishment tendencies. After the group formation in Phase 1, participants in Revenge and Control groups reported similar ingroup favoritism in feelings of closeness (Ingroup vs. Outgroup: 4.72±0.98 vs. 3.27±0.99; $F(1,86) = 136.23$, $p < 0.001$, $\eta^2_p = 0.613$). After fMRI scanning, participants were asked to report their emotions and attitudes related to ingroup and outgroup targets on a Likert Scale (1=not at all, 9=extremely strong). Participants from both the Revenge and Control groups reported similar ingroup favoritism in emotions and attitudes (see Figure 2 - Figure supplement 1, Supplementary file 2, Supplementary file 3 for statistical details). When viewing ingroup vs. outgroup targets' pain, the participants in both conditions reported greater empathy (6.58±1.80 vs. 6.15±1.91; $F(1,78) = 8.21$, $p = 0.006$, $\eta^2_p = .095$), unpleasantness (4.23±2.09 vs. 3.77±1.73; $F(1,78) = 5.37$, $p = 0.026$, $\eta^2_p = .064$), anger (2.66±1.88 vs. 2.11±1.35; $F(1,78) = 9.68$, $p = 0.004$, $\eta^2_p = .110$), and fear (2.64±1.97 vs. 2.36±1.72; $F(1,78) = 3.95$, $p = 0.050$, $\eta^2_p = .048$, all FDR corrected,). On the contrary, the participants in both conditions reported greater schadenfreude when viewing outgroup vs. ingroup targets' pain (2.22±1.60 vs. 1.61±1.00; $F(1,78) = 14.91$, $p < 0.002$, FDR corrected, $\eta^2_p = .160$). Moreover, the participants in both conditions reported greater trust (5.53±1.38 vs. 4.51±1.49; $F(1,78) = 28.62$, $p < 0.002$, FDR corrected, $\eta^2_p = .268$) and likability (ingroup: 5.32±1.53; outgroup: 4.51±1.47; $F(1,78) = 23.57$, $p < 0.002$, FDR corrected, $\eta^2_p = .232$) of ingroup than outgroup members. Finally, participants in both conditions reported greater tendencies to punish outgroup targets (3.23±1.91 vs. 2.26±1.53; $F(1,78) = 27.19$, $p < 0.002$, FDR corrected, $\eta^2_p = .259$, Figure 2A). Importantly, ingroup favoritism in attitudes, emotions, and punishment tendencies did not differ significantly between the Revenge and Control groups and between the involved and uninvolved targets (see Supplementary file 2 for statistical details). Accordingly, the results of self-report measures indicate that the ingroup/outgroup manipulation was successful in both the Revenge and Control groups. In

addition, the results of similar ingroup biases in attitudes, emotions, and punishment tendencies in the two groups suggest that any differences in the neurobiological measures across the Revenge and Control groups cannot be attributed to differences in group affiliation or ingroup favoritism in emotions and attitudes between the two groups.

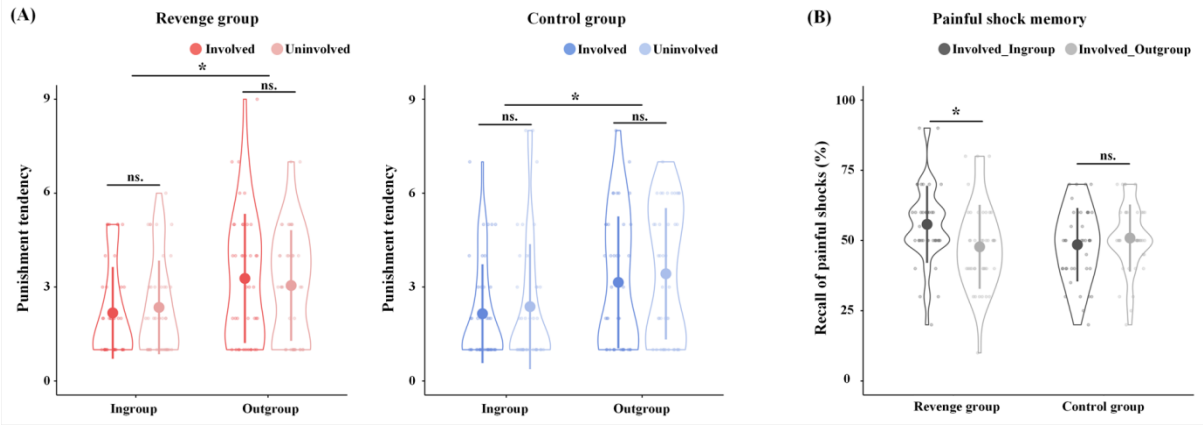

Figure 2. Behavioral results. (A) Self-report of punishment tendencies. (B) Memory of painful shocks applied to Involved_Ingroup and Involved_Outgroup targets during scanning. Shown are group means (big dots), standard deviation (bars), measures of each individual (small dots), and distribution (violin shape).

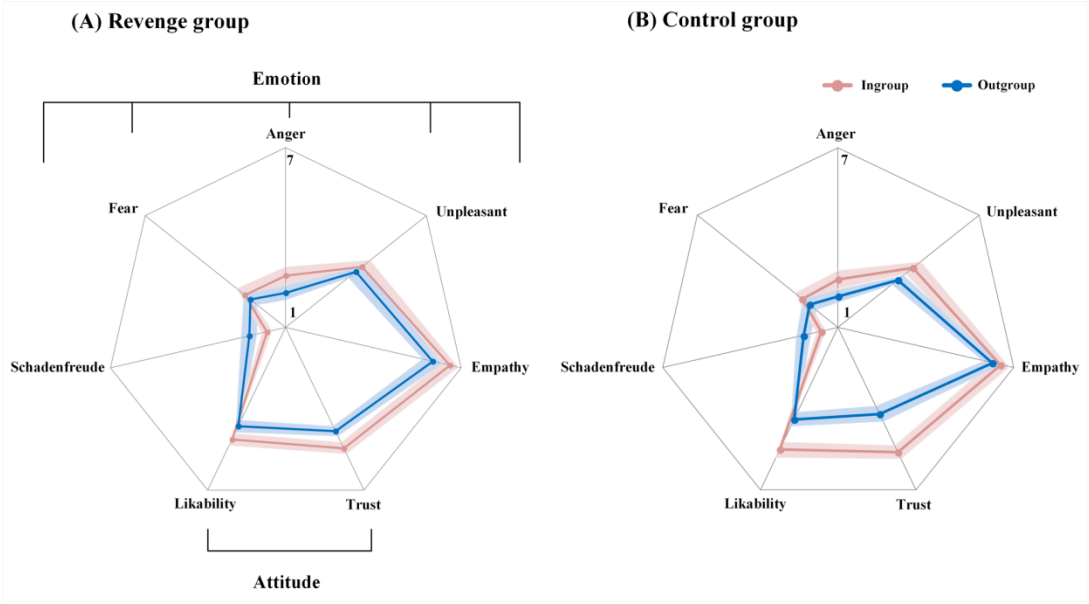

Figure 2 - Figure supplement 1. Illustration of behavioral results. (A) The results of attitude and emotion ratings (from 1 to 7) related to ingroup and outgroup targets in the Revenge group. (B) The results of attitude and emotion ratings (from 1 to 7) related to ingroup and outgroup targets in the Control group.

Previous research has shown that people tend to view their ingroup members as victims and outgroup members as perpetrators during intergroup conflicts (*Ross and Ward, 1995; Lickel et al., 2006*). Accordingly, we conducted another manipulation check to assess whether the participants in the Revenge group tended to remember their ingroup members as being the victim of painful shocks during the competitive game to a greater degree than participants in the Control group. After fMRI scanning, participants were asked to recall how often Involved_Ingroup and Involved_Outgroup targets received painful shocks after losing the game. The analysis of variance (ANOVA) of self-report of frequencies of perceived painful shocks with Intergroup Relationship (ingroup vs. outgroup) as a within-subjects variable and Group (Revenge vs. Control group) as a between-subjects variable revealed a significant main effect of Intergroup Relationship ($F(1,78) = 5.65$, $p = 0.020$, $\eta^2_p = .068$). Participants reported more painful shocks received by ingroup than outgroup members even though Involved_Ingroup and Involved_Outgroup targets actually received the same amount of painful shocks. There was also a significant interaction of Intergroup Relationship x Group ($F(1,78) = 19.21$, $p < 0.001$, $\eta^2_p = .198$). Simple effect analyses revealed that the Control group reported the similar levels of painful shocks that were delivered to Involved_Ingroup and Involved_Outgroup targets (0.49±0.13 vs. 0.51±0.12; $t(39) = -1.58$, $p = 0.123$, Cohen's d = 0.25), whereas the Revenge group reported significantly more painful shocks that were delivered to Involved_Ingroup than to Involved_Outgroup targets (0.56±0.14 vs. 0.48±0.15; $t(39) = 4.39$, $p < 0.001$; Cohen's d = 0.69, Figure 2B). The results are consistent with previous

findings (*Ross and Ward, 1995; Lickel et al., 2006*) and suggest that our manipulations motivated participants of the Revenge (vs. Control) group to more frequently view the involved ingroup member as a victim during intergroup conflict.

Increased endogenous OT in Revenge than Control groups

To test the hypothesis that endogenous OT in humans is sensitive to intergroup conflict, we collected saliva from each participant at three points in time (see Figure 1A). If the oxytocinergic system is activated during intergroup conflict in humans, similar to that in Chimpanzees (*Samuni et al., 2017*), the Revenge (vs. Control) group should show a greater level of OT in saliva at Time 1 after the initial witness of intergroup conflict, and furthermore, this effect may increase even more at Time 2 after the participants had witnessed the whole procedure of conflict. We conducted an ANOVA of salivary OT levels with Group (Revenge vs. Control) as a between-subjects variable, and Time (Time 1, 2, and 3) as a within-subjects variable. Because previous research has shown evidence for associations between the administration of OT and ingroup biases in emotions and attitudes (*De Dreu et al., 2011; Sheng et al., 2013*), the ANOVA included ingroup biases in feelings of closeness and other emotions and attitudes as covariates. The results showed a significant effect of Group ($F(1,67) = 22.66$, $p < 0.001$, $\eta^2_p = 0.253$) and a significant interaction of Group × Timing ($F(2,134) = 4.04$, $p = 0.020$, $\eta^2_p = 0.057$, Figure 3A, see Table 1 for results of simple effect analyses). These results indicate two important consequences of group conflict: Revenge (vs. Control) group showed higher endogenous OT levels *immediately after* the initial conflict was observed, and OT levels continue to rise in response to later intergroup conflict in the revenge condition.

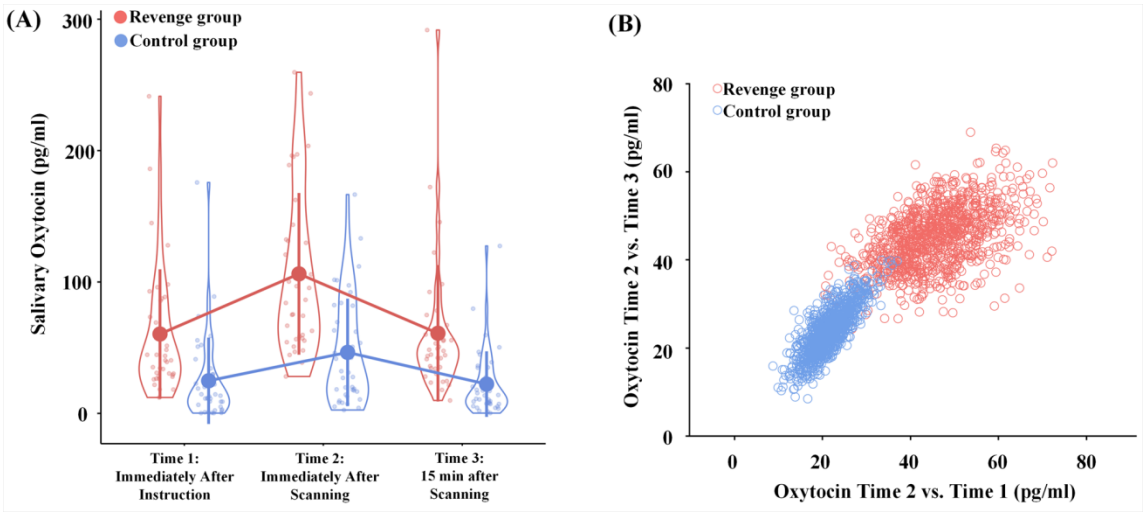

Figure 3. Effects of intergroup conflict on endogenous OT. (A) Salivary OT levels at three points in time of the experimental procedure. Shown are group means (big dots), SD (bars), measures of each individual (small dots), and distribution (violin shape). (B) Results of bootstrapping analyses. Increased OT levels were calculated by subtracting Time-1 and Time-3 measures from Time-2 measure for the two bootstrapped samples, respectively.

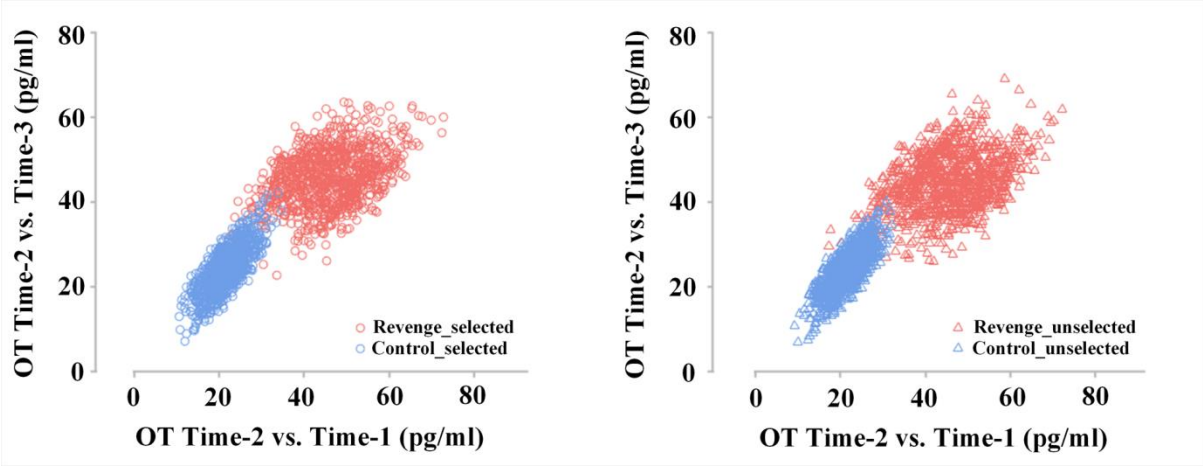

Figure 3 - Figure supplement 1. Results of the modified bootstrapping analysis. The left panel shows the distinct patterns of OT increase of the selected samples. The right panel shows the distinct patterns of OT increase of the unselected samples. Increased OT levels

were calculated by subtracting Time-1 and Time-3 measures from Time-2 measure for each participant. Thereafter, from this data set of each group, half participants were randomly selected without replacement as the selected sample, leaving the other participants in the unselected sample. The means of the selected and unselected samples were then calculated and plotted as one of the points (x, y) in a panel with the horizontal (x) and vertical (y) axes showing OT increases at Time-2 relative to Time-1 and Time-3, respectively. The same procedure was repeated for 1,000 times to estimate the population means and variations for the selected and unselected samples of each participating group. To confirm the separation of the Revenge and Control groups in both the selected and unselected samples, we calculated the Euclidean distance between the selected samples of the Revenge and Control groups and between unselected samples of the Revenge and Control groups, respectively. The mean distance between the Revenge and Control groups in the selected sample is 31.51 with the 95% confidence interval of [9.92, 53.41] which does not include 0. Similarly, the mean distance between the Revenge and Control groups in the unselected sample is 32.83 with the 95% confidence interval of [10.08, 54.90] which does not include 0 either. These results indicate that the selected and unselected samples similarly showed distinct patterns of OT increase between the Revenge and Control groups. Furthermore, to test the similarity between the selected and the unselected samples of each participant group, we calculated the Euclidean distance between the selected and unselected sample in the Revenge and Control groups, respectively. The mean distance between the selected and unselected samples in the Revenge group was -0.75 with the 95% confidence interval of [-39.18, 35.70] which includes 0. The mean distance between the selected and unselected samples in the Control group was 0.57 with the 95% confidence interval of [-21.66, 24.25] which also includes 0. The results suggest similar patterns of OT increases in the selected and the unselected samples of each participant group.

Table 1. Salivary OT levels (pg/ml) across three time points and the results of simple effect analyses

	Revenge (Mean±SD)	Control (Mean±SD)	F	p	η^2_p
OT Time 1	60.40±49.25	24.68±32.46	10.54	= 0.002	.136
OT Time 2	106.19±61.52	47.42±40.83	22.90	< 0.001	.255
OT Time 3	60.96±51.92	21.72±24.87	16.47	< 0.001	.197

To further illustrate the greater increase of OT levels after additional experiences of witnessing intergroup in the Revenge (vs. Control) group, we adopted a standard bootstrapping procedure (*Davison and Hinkley, 1997*) to more examine the difference in increased OT levels between Revenge and Control groups. Specifically, we conducted a bootstrapping analysis to illustrate a greater increase of OT level from Time-1 to Time-2 and a greater decrease of OT level from Time-2 to Time-3 in the Revenge compared to Control group. To do this, we calculated increased OT levels by subtracting measures at Time-1 and Time-3 from Time-2 for each participant. Thereafter, a bootstrapped data set in each group was nonparametrically resampled with replacement (i.e., a participant could be selected more than once). The mean of this bootstrapped sample was then calculated and plotted as one of the points (x, y) in a panel with the horizontal (x) and vertical (y) axes showing OT increases at Time 2 relative to that at Time 1 and Time 3, respectively. The same procedure was repeated for n=1,000 times to estimate the population means and variations for each participating group. As shown in Figure 3B, the bootstrapped sample mean points from the Revenge group fall mostly to the upper right of the 2D plot. To confirm the separation of the two bootstrapped samples, we calculated the Euclidean distance between the two samples. The mean distance between the two samples is 32.29 with the 95% confidence interval of [11.55, 54.61]. We conducted another modified bootstrapping analysis to assess whether the increased OT level related to experiences of conflict obtained from half participants randomly selected from each group can be replicated in the unselected participants. The results suggest

similar group differences in increased OT levels related to witnessing conflict in the selected and unselected samples (see Figure 3 - Figure supplement 1).

Together, these results support our hypothesis that endogenous OT in humans increases during intergroup conflict. Specifically, the level of endogenous OT seemed to start rising immediately after participants initially witnessed intergroup conflict (i.e., at Time 1). Moreover, the OT level increased further after fMRI scanning (i.e., at Time 2) during which the participants had more experiences of intergroup conflict. These findings are consistent with the observations in Chimpanzees (*Samuni et al., 2017*) and provide empirical evidence that intergroup conflict in primates including humans is associated with increased levels of endogenous OT.

Brain responses to perceived pain in Revenge and Control groups

In our design, an increase in brain activity to perceived painful (vs. neutral) expressions is a precondition for examining of revenge-related functional role of the association between endogenous OT and neural responses in either the empathy or theory-of-mind networks. In addition, based on previous findings of ingroup favoritism in empathic neural responses (*Xu et al., 2009; Hein et al., 2010; Cikara et al., 2011; Sheng and Han, 2012; Han, 2018*), we expected greater neural responses to ingroup than outgroup pain if our group manipulations were successful. The presence of ingroup favoritism in empathic neural responses provides a precondition for further analyses of OT associations with empathic neural responses to ingroup and outgroup pain separately. Therefore, we first examined participants' neural responses to perceived pain in others by conducting a whole brain analysis that collapsed all targets and all participants. Similar to previous findings (*Fan et al., 2011; Lamm et al., 2011; Shamay-Tsoory, 2011*), whole-brain analyses of the contrast of painful vs. neutral expressions revealed activations in both the empathy network including the anterior cingulate and bilateral AI/inferior frontal gyrus (IFG) and the theory-of-mind network, including the mPFC,

left TPJ, and right temporal pole (TP) (all activations were identified by combining a voxel-level threshold of $p < .001$ and a cluster-level threshold of $p < .05$, FWE corrected, Figure 4A, Supplementary file 4, see Figure 4 - Figure supplement 1, Figure 4 - Figure supplement 2, Supplementary file 5, Supplementary file 6 for the results of the Revenge and Control groups, separately).

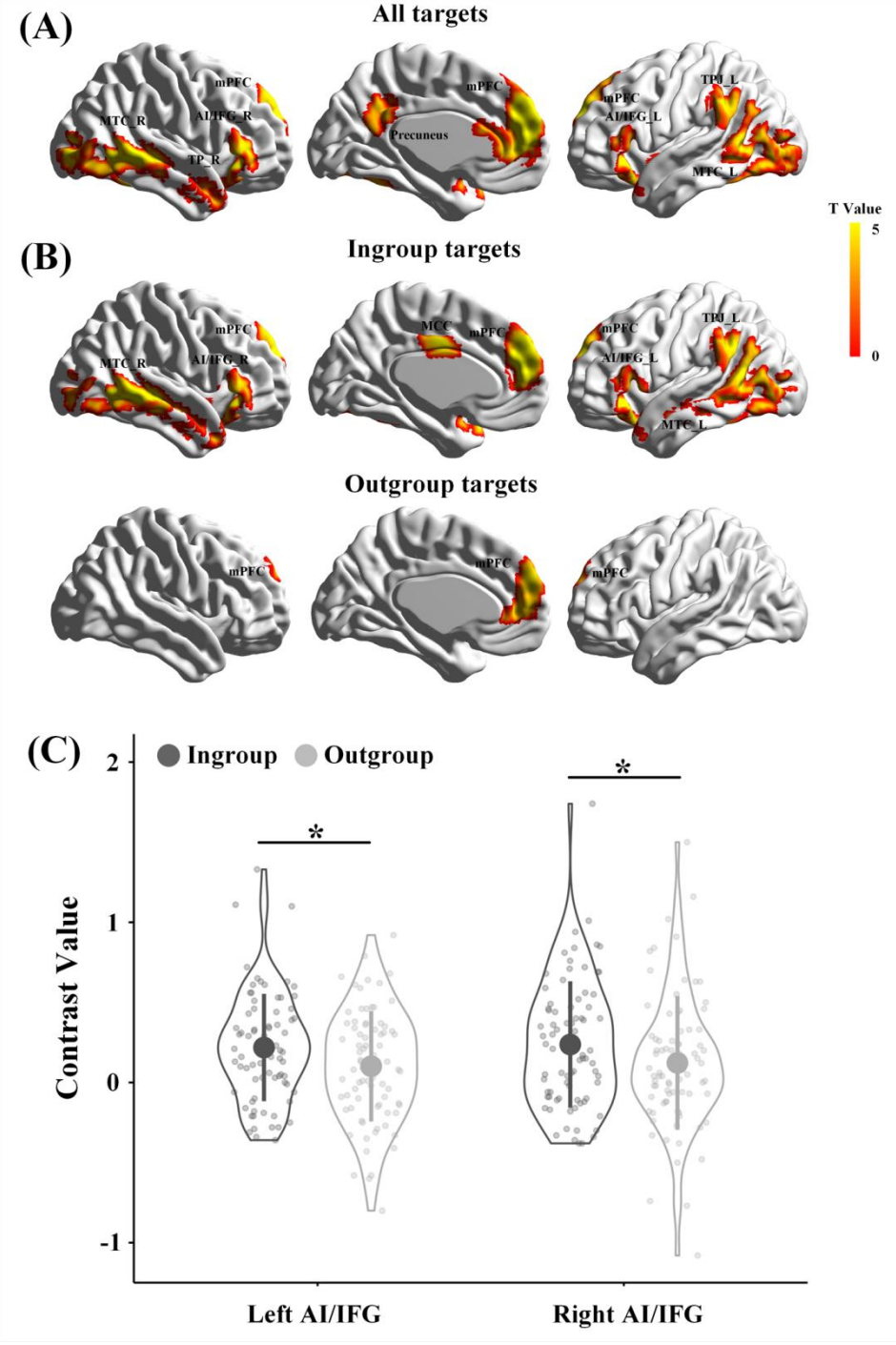

Figure 4. (A) Illustrations of brain responses to painful vs. neutral expressions of all targets

perceived during scanning across all participants from Revenge and Control groups and collapsing involved and uninvolved targets. (B) Illustrations of brain responses to painful vs. neutral expressions of ingroup targets and outgroup targets across all participants. (C) The contrast values of neural responses to painful (vs. neutral) expressions in the left and right AI/IFG. Shown are group means (big dots), SD (bars), measures of each individual (small dots), and distribution (violin shape). mPFC=medial prefrontal cortex; MCC=midcingulate cortex; MTC=middle temporal cortex; AI/IFG=anterior insula and inferior frontal cortex; TPJ=temporoparietal junction; TP=temporal pole.

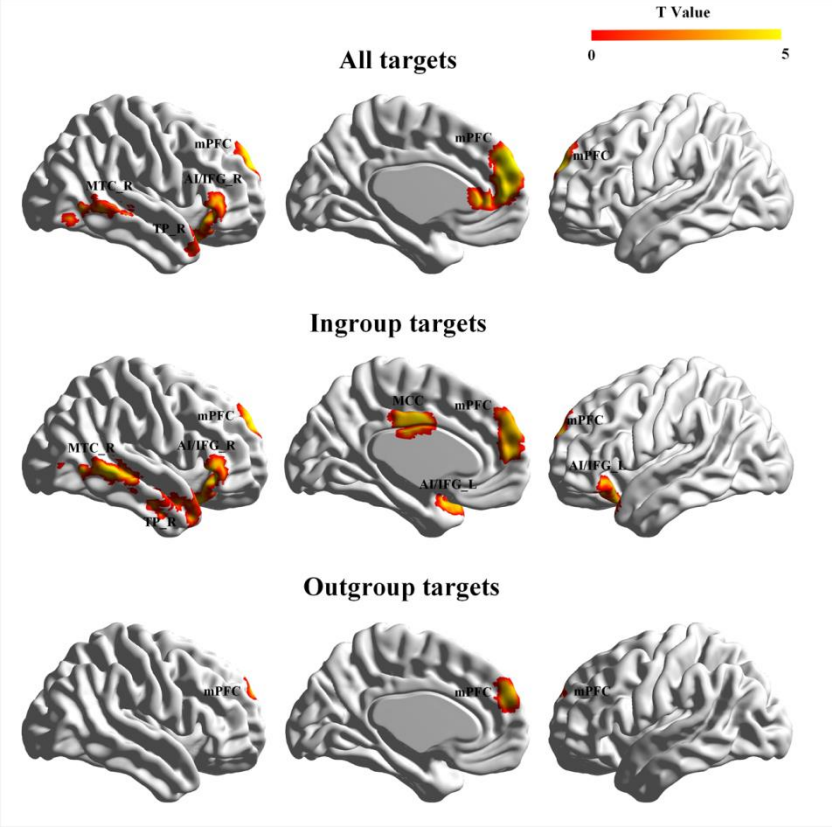

Figure 4 - Figure supplement 1. Illustrations of brain responses to painful vs. neutral expressions across participants and collapsed involved and uninvolved targets in the Revenge group (a voxel level threshold $p < 0.001$, uncorrected and a cluster level threshold of $p<0.05$, FWE corrected). Activations to all targets, ingroup targets, and outgroup targets are shown separately. mPFC=medial prefrontal cortex; MCC=midcingulate cortex; MTC=middle

temporal cortex; AI/IFG=anterior insula and inferior frontal cortex; TP=temporal pole.

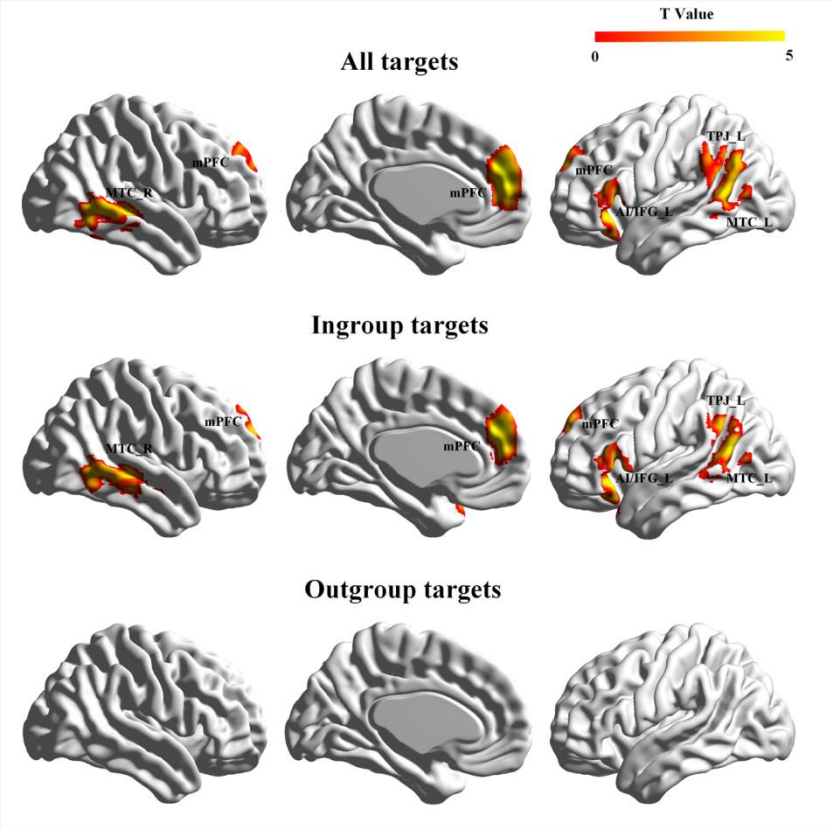

Figure 4 - Figure supplement 2. Illustration of brain responses to painful vs. neutral expressions across participants and collapsed involved and uninvolved targets in the Control group (a voxel level threshold $p < 0.001$, uncorrected and a cluster level threshold of $p<0.05$, FWE corrected). Activations to all targets, ingroup targets, and outgroup targets are shown separately. mPFC=medial prefrontal cortex; MCC=midcingulate cortex; MTC=middle temporal cortex; AI/IFG=anterior insula and inferior frontal cortex; TPJ=temporoparietal junction.

Separate whole-brain analyses that collapsed all participants in the Revenge and Control groups identified activations in the mPFC, aMCC, bilateral AI/IFG, and left TPJ in response to ingroup targets' pain but only in the mPFC in response to outgroup targets' pain (combined a voxel level threshold $p < 0.001$ and a cluster level threshold $p < 0.05$, FWE corrected, Figure 4B, Supplementary file 4). To further examine the ingroup favoritism in neural

responses in these brain regions, we conducted region-of-interest (ROI) analyses of neural responses to others' pain in the brain regions identified in the whole-brain analyses. ROI were defined as spheres with 5-mm radius centered at the peak activation using a leave-one-out method by collapsing all participants. The leave-one-out method identified the bilateral AI/IFG, mPFC, and left TPJ in response to painful vs. neutral expressions at the combined voxel level threshold $p < 0.001$ and cluster level threshold $p < 0.05$, FWE corrected. The contrast values (painful vs. neutral expressions) were extracted from each ROI and subject to ANOVAs with Relationship (Ingroup vs. Outgroup) and Involvement (Involved vs. Uninvolved) as within-subjects variables and Group (Revenge vs. Control group) as a between-subjects variable. The results confirmed greater neural responses to ingroup than outgroup targets' pain in the empathy network including the left IFG/AI (0.22±0.34 vs. 0.10±0.34, $F(1,78) = 5.41$, $p = 0.036$, $\eta^2_p = 0.065$, all results of ROI analyses were FDR corrected) and right IFG/AI (0.24±0.39 vs. 0.12±0.42, $F(1,78) = 4.56$, $p = 0.036$, $\eta^2_p = 0.055$, Figure 4C) but not in the theory-of-mind network (mPFC: 0.33±0.45 vs. 0.20±0.44, $F(1,78) = 3.92$, $p = 0.102$, $\eta^2_p = 0.048$; left TPJ (0.21±0.46 vs. 0.13±0.43, $F(1,78) = 1.45$, $p = 0.232$, $\eta^2_p = 0.018$). The effect of increased neural responses to ingroup than outgroup targets' pain did not differ significantly between Revenge and Control groups (left IFG/AI: $F(1, 78) = 0.09$, $p = 0.770$, $\eta^2_p = 0.001$; right IFG/AI: $F(1, 78) = 0.13$, $p = 0.716$, $\eta^2_p = 0.002$) and between Involved and Uninvolved targets (left IFG/AI: $F(1,78) = 2.71$, $p = 0.104$, $\eta^2_p = 0.034$; right IFG/AI: $F(1,78) = 2.68$, $p = 0.105$, $\eta^2_p = 0.033$).

These results replicate the previous neuroimaging findings of activations in the empathy and theory-of-mind networks in response to perceived pain in others (*Fan et al., 2011; Lamm et al., 2011; Shamay-Tsoory, 2011*) and enhanced neural responses to ingroup than outgroup pain (*Xu et al., 2009; Hein et al., 2010; Cikara et al., 2011; Sheng and Han, 2012; Han, 2018*). These results provide bases for further tests of the association between endogenous

OT and brain responses to perceived pain in others. Importantly, the results provide no evidence for difference in ingroup favoritism in empathic neural responses between the Revenge and Control groups. Therefore, any possible contribution of ingroup biases in brain activity to group differences in endogenous OT and associations between endogenous OT and brain responses to others' pain was reduced to a minimum degree.

Endogenous OT predicts mPFC activity in response to ingroup pain

If the association between endogenous OT to brain responses to ingroup pain serves as a neurobiological correlate of revenge propensity during intergroup conflict, endogenous OT after the initial intergroup conflict at Time 1 should predict subsequent brain responses to ingroup pain, which may then further predict revenge propensity. Accordingly, we first conducted a whole-brain regression analysis to examine whether OT levels at Time 1 predicts the brain responses to perceived ingroup pain. As discussed below, this analysis revealed an association between the mPFC activity and OT level at Time 1 in the Revenge group. In order to then estimate whether the OT-mPFC association was specific to OT levels at time 1, we conducted a second whole-brain regression analysis to examine brain responses to ingroup pain that were associated with endogenous OT measured after fMRI scanning at Time 2. Whole-brain analyses were used so as to not bias OT association with a specific network (e.g., the empathy or theory-of-mind network).

In the first whole-brain regression analysis, the OT level at Time 1 was entered into a general linear model as predictors of brain responses to painful (vs. neutral) expressions of each target. The results showed that, for Revenge (but not Control) group, OT level at Time 1 reliably predicted the mPFC activity in response to Involved_Ingroup target's pain (combined a voxel level threshold $p < 0.001$ and a cluster level threshold $p < 0.05$, FWE corrected, Figure 5A). We conducted ROI-based moderation analyses to further confirm the group differences in the coupling between the OT level at Time 1 and mPFC activity. The mPFC

activity to Involved_Ingroup targets' painful (vs. neutral) expression was extracted from the ROI defined in the leave-one-out whole-brain analysis of the contrast of painful vs. neutral expressions using a threshold that combined a voxel level threshold $p < 0.001$ and a cluster level threshold $p < 0.05$, FWE corrected. The mPFC activity to Involved_Ingroup targets' pain was entered as the independent variable, Group (Revenge vs. Control) was entered as the moderator, and ingroup biases in closeness, emotion and attitudes were entered as covariates for their possible contributions to the association between OT levels and mPFC activity during intergroup conflict. The moderation analysis including the covariates showed that the interaction between mPFC activity and Group accounted for a significant proportion of variance in the OT level at Time 1 ($\Delta R^2 = 0.05$, $\Delta F(1,65) = 5.70$, $p = 0.03$, Figure 5B and 5C, see Supplementary file 7 for statistical details). The results suggest that the association between endogenous OT and mPFC activity was specific to Revenge group.

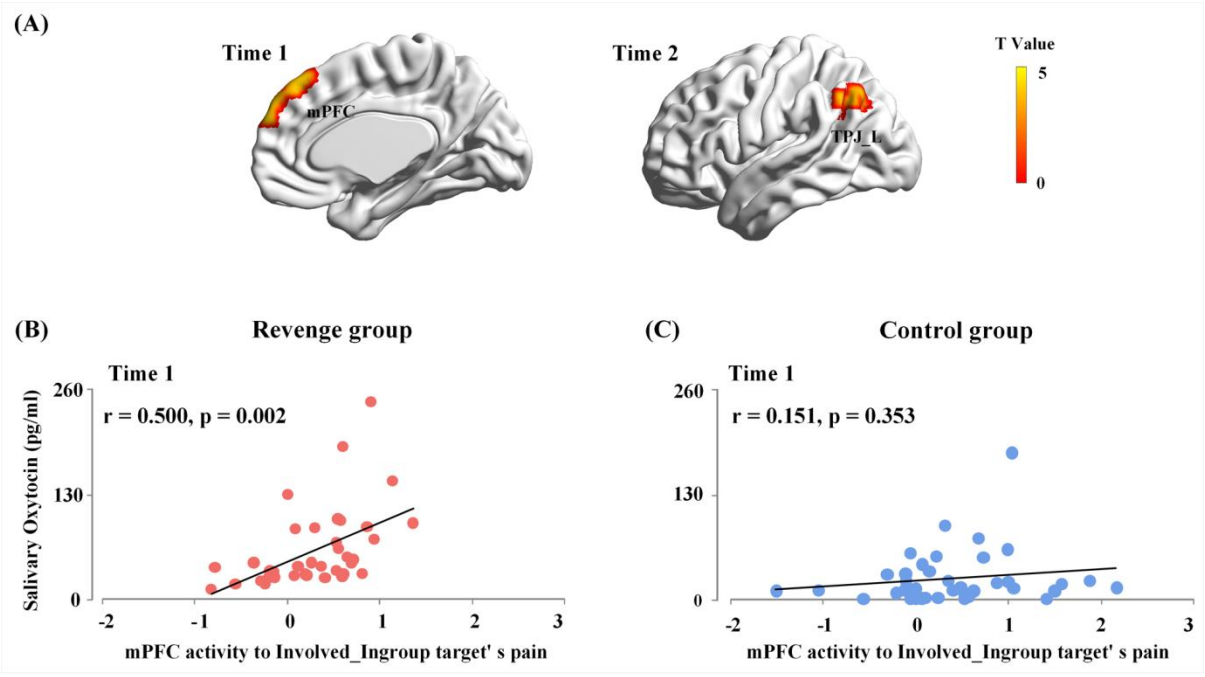

Figure 5. Associations between endogenous OT and brain activity in response to others' suffering. (A) The mPFC activity to Involved_Ingroup target's pain associated with the endogenous OT at Time 1 in Revenge group. The OT level at Time 2 reliably predicted the

left TPJ activities in response to Involved_Ingroup target's pain in Revenge group. A voxel-level threshold of $p < .001$ and a cluster-level threshold of $p < .05$, FWE corrected was used to identify and to visualize brain activations. (B) The associations between endogenous OT-levels at Time 1 with the mPFC activity to Involved_Ingroup target's pain for Revenge group. (C) No significant correlation between endogenous OT-levels and the mPFC activity to Involved_Ingroup target's pain was found for the Control group. Note: the results of the moderation analysis indicate a significant group difference in the association between endogenous OT-levels at Time 1 and the mPFC activity to Involved_Ingroup target's pain.

Furthermore, to test whether an ingroup member was directly involved in the conflict was critical for the association between endogenous OT and mPFC activity to ingroup pain, we conducted an additional ROI-based moderation analysis to examine the differential coupling between OT level at Time 1 and mPFC activity in response to Involved_Ingroup vs. Uninvolved_Ingroup targets' pain in the Revenge group. The moderation analysis used the repeated measures of mPFC activities towards Invovled_Ingroup vs. Uninvolved_Ingroup targets as the moderator. The results showed that the endogenous OT level at Time 1 accounts for a significant amount of variance of the difference between mPFC activities towards Involved_Ingroup vs. Uninvolved_Ingroup targets ($R = 0.15$, $F(1, 75) = 12.99$, $p < 0.001$). The results suggest a stronger coupling between endogenous OT and mPFC activity in response to Involved_Ingroup (compared to Uninvolved_Ingroup) targets' pain during intergroup conflict.

In the second whole-brain regression analysis, OT level at Time 2 was entered into a general linear model as predictors of brain responses to painful (vs. neutral) expressions of each target. The results only showed that OT level at Time 2 was significantly associated with the left TPJ activity in response to Involved_Ingroup target's pain (combined a voxel level

560 threshold $p < 0.001$ and a cluster level threshold $p < 0.05$, FWE corrected, Figure 5A).
561 However, ROI-based moderation analyses failed to confirm any significant differences in the
562 coupling between the OT level at Time 2 and left TPJ activity between the Revenge and
563 Control groups. Thus, the results provide no evidence for revenge specific association
564 between brain responses to ingroup pain and further changes of endogenous OT.
565 Together, these results suggest that intergroup conflict enhanced the link between
566 endogenous OT measured after the initial experience of ingroup conflict and mPFC activity
567 in response to perceived pain of the ingroup member who was directly involved in conflict
568 with an outgroup member. These results provide bases for further examination of the
569 functional role of the mPFC activity in mediating the association between endogenous OT
570 after initial experience and revenge propensity.
571 ***Association between mPFC activity and revenge propensity***
572 Because only the mPFC activity in response to ingroup members' pain was coupled
573 with endogenous OT level at Time 1, we conducted an ROI analysis to examine the
574 associations between the mPFC activity to ingroup members' pain and tendencies to
575 punishment outgroup members in both the Revenge and Control group. The contrast values
576 of painful vs. neutral expressions of Involved_Ingroup targets were extracted from an ROI (a
577 sphere with 5-mm radius) centered at the mPFC activation (using a leave-one-out method by
578 collapsing participants from the two subject groups). The results of correlation analyses
579 showed that, for Revenge (but not Control) group, the mPFC activity in response to
580 Involved_Ingroup targets' pain positively predicted punishment tendencies toward both
581 Involved_Outgroup and Uninvolved_Outgroup targets ($r = 0.35$ and 0.42; $p = 0.026$ and
582 0.014, FDR corrected, Figure 6A). The results suggest that individuals with stronger mPFC
583 activity in response to ingroup pain tended to apply more painful shocks to outgroup
584 members regardless whether they were directly involved in the conflict. This finding provides

585 a potential neural basis for understanding how conflicts between two individuals spread
586 across the two groups that the two individuals are affiliated (*Gelfand et al*., *2012; Lee et al*.,
587 *2013*). In addition, the finding of the association between mPFC activity and revenge
588 propensity provides a basis for the following mediation analysis.
589

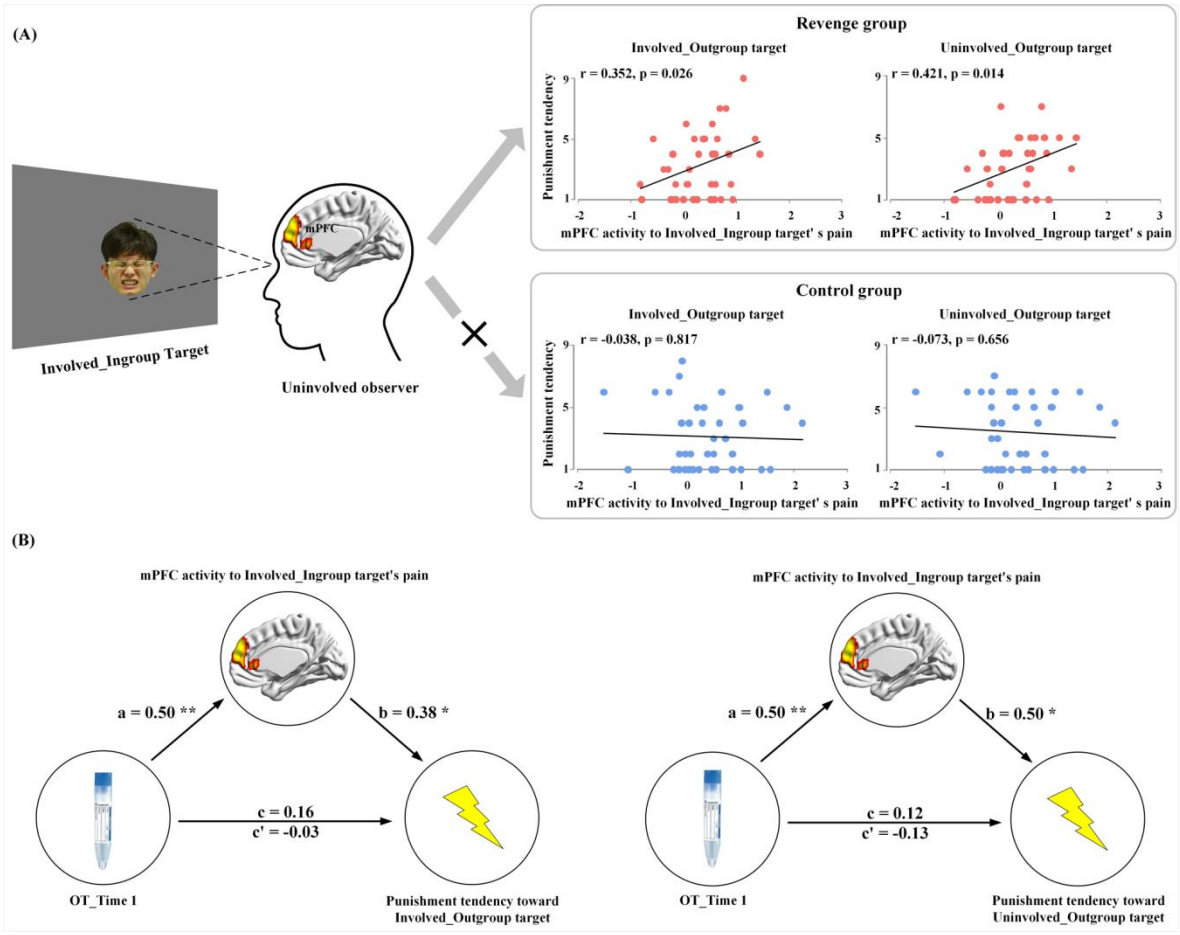

Figure 6. Results of brain-propensity associations and mediation analyses. (A) Brain-propensity associations in Revenge group. The mPFC activity to Involved_Ingroup target's pain in an uninvolved observer from Revenge (but not Control) group predicted his punishment tendencies toward both Involved_Outgroup and Uninvolved_Outgroup targets. (B) The mPFC mediation of endogenous OT and punishment tendencies. The mPFC activity to Involved_Ingroup targets' pain mediates the relationship between the salivary level of endogenous OT at Time 1 and punishment tendencies toward Involved_Outgroup targets (left) and Uninvolved_Outgroup targets (right).

mPFC mediation of association between endogenous OT and outgroup punishment

Finally, we estimated the neurobiological (from endogenous OT to mPFC activity in response to ingroup pain) association of revenge propensity during intergroup conflict by conducting ROI-based mediation analyses in the Revenge group. The analyses focused on the functional role of mPFC activity to ingroup pain in mediating the relationship between endogenous OT measured after the initial intergroup conflict (Time 1) and later punishment tendencies toward outgroup. The first mediation analysis examined whether the mPFC activity to Involved_Ingroup targets' pain mediates the relationship between the OT level at Time 1 and tendency to punish Involved_Outgroup targets. In Step 1 of the mediation model, the regression of the OT level on punishment tendency toward Involved_Outgroup target was not significant ($b = 0.16$, $t(35) = 0.94$, $p = 0.355$) when not considering the mediator (e.g., the mPFC activity). Step 2 showed that the regression of the OT level on the mediator was significant ($b = 0.50$, $t(35) = 3.41$, $p = 0.002$). Step 3 showed that the regression of the mediator on punishment tendency was significant ($b = 0.38$, $t(34) = 2.05$, $p = 0.048$) when controlling for the OT level. Step 4 revealed that the OT level was not a significant predictor of punishment tendency ($b = -0.03$, $t(34) = -0.17$, $p = 0.863$) when controlling for the mediator (Figure 6B, see Supplementary file 8 for statistical details). The indirect effect size was 0.19 with a 95% confidence interval which did not include zero (0.01, 0.44)

The second mediation analysis examined whether the mPFC activity to Involved_Ingroup targets' pain mediates the relationship between the OT level at Time 1 and tendency to punish Uninvolved_Outgroup targets. In step 1 of the mediation model, the regression of the OT level on punishment tendency toward the Uninvolved_Outgroup target was not significant ($b = 0.12$, $t(35) = .70$, $p = 0.488$) when not considering the mediator (e.g., the mPFC activity). Step 2 showed that the regression of the OT level on the mediator was significant ($b = 0.50$, $t(35) = 3.41$, $p = 0.002$). Step 3 showed that the regression of the

mediator on punishment tendency was significant (b = 0.50, t(34) = 2.80, p = 0.008) when controlling for the OT level. Step 4 revealed that the OT level was not a significant predictor of punishment tendency (b = -0.13, t(34) = -0.74, p = 0.467) when controlling for the mediator (Figure 6B, see Supplementary file 9 for statistical details). The indirect effect size was 0.25 with a 95% confidence interval which did not include zero (0.08, 0.46). These results indicate that the mPFC activity in response to ingroup pain caused by an outgroup mediates the association between endogenous OT measured after initially witnessing intergroup conflict and tendencies to retaliate upon outgroup members regardless whether they directly brought physical harm to ingroup members. The results of these mediation analyses provide additional evidence for the endogenous-OT/mPFC association as a neurobiological correlate of revenge propensity during intergroup conflict.

Discussion

Revenge behavior is universal and highly costly. Accordingly, understanding its neurobiological bases is of critical theoretical and practical importance. Revenge behavior during intergroup conflict engages multiple psychological processes from perceiving ingroup suffering to making aggressive decisions toward outgroups. Our research focused on how neurobiological responses to ingroup pain are related to revenge decisions during intergroup conflict. Although multiple motives are involved in vengeful behaviors during intergroup conflict, we compared salivary OT and brain activity in the Revenge and Control groups. This design allowed us to isolate the neurobiological responses to perceived ingroup suffering due to physical harm caused by an outgroup member and its association with seeking revenge through physical harm upon outgroup members.

While previous research has reported endogenous OT reactivity during intergroup conflict in wild chimpanzees (*Samuni et al., 2017*), we showed the first evidence for increased levels of endogenous OT in humans during intergroup conflict that involves

physical harm to ingroups caused by an outgroup. The salivary OT level in humans is affected by affiliative contact and is interrelated with the plasma OT level (*Feldman et al., 2011*). Our results complement the previous research on the effect of intranasal OT administration on ingroup cooperation and outgroup defensive competition in economic games (*De Dreu et al., 2010; De Dreu et al., 2011*). Importantly, our findings suggest that endogenous OT is an influential physiological mechanism in humans that gets activated in response to intergroup conflicts that involves physical harm between ingroup and outgroup members. It is worth noting that, similar to the finding in chimpanzees (*Samuni et al., 2017*), our results suggest increased endogenous OT occurs after initially witnessing intergroup conflict (e.g., at Time 1). It appears that, in both humans and chimpanzees, the oxytocinergic system quickly responds to intergroup conflict. In addition, the endogenous OT increased more after further witnessing intergroup conflict (e.g., at Time 2) and dropped when intergroup conflict had ended (e.g., at Time 3). These results illustrate dynamic changes of endogenous OT that occur across the whole procedure of an intergroup conflict.

Our results also revealed intermediate brain mechanisms linking endogenous OT reactivity to perceived intergroup conflict and revenge propensity during intergroup conflict. Unlike previous research that focused on increased neural activity following aggressive decisions (*Seymour et al., 2007; Krämer et al., 2007; Krueger and Hoffman, 2016; Chester and DeWall, 2016*), our work showed evidence that the mPFC activity in response to ingroup pain predicted the propensity of subsequent revenge behavior during intergroup conflict. The mPFC is well-known for its functional role in representing mental states (*Amodio and Frith, 2006*), social emotion (*Harris and Fiske, 2007; Mathur et al., 2010*), and group identity (*Volz et al., 2009; Molenberghs and Morrison, 2014*). Our results further revealed that the neurobiological association between endogenous OT to mPFC activity in response to Involved_Ingroup targets' suffering is related to propensity to punish outgroup members

675 regardless of whether the outgroup members were directly involved in the conflict. This
676 cross-group brain-propensity association is different from previous findings of a within-group
677 brain-propensity association in an intergroup context without direct conflict (i.e., neural
678 responses to ingroup pain predict tendencies to help ingroup members, or neural responses to
679 outgroup pain predict tendencies not to help outgroup members) (*Hein et al., 2010; Cikara et*
680 *al., 2011; Mathur et al., 2010*). Our results cast a new perspective on the neural
681 underpinnings that drive decisions to apply physical harm toward outgroups during
682 intergroup conflict. More generally, the cross-group brain-propensity association suggests a
683 potential neural mechanism underlying the *contagion* of revenge behavior, and may help to
684 understand why disputes between two individuals can escalate across groups and across time
685 (*Gelfand et al, 2012*).

686 Our findings make fundamental contributions to the intergroup conflict literature and
687 in particular, the neurobiological associations of ingroup love and outgroup hate. Intergroup
688 conflict plays a substantial role in the evolution of both aggressiveness against outgroup and
689 cooperativeness towards ingroups (*Rusch, 2014*). While previous studies have demonstrated
690 the role of OT and mPFC activity in altruistic decisions favoring the ingroup (*De Dreu et al.,*
691 *2010; Mathur et al., 2010; De Dreu et al., 2011*), our results suggest that the context of
692 intergroup conflict may shift the key function of the oxytocinergic system from mediating
693 ingroup love (i.e., a desire to help the ingroup) to facilitating outgroup hate (an aggressive
694 motivation to hurt the outgroup). Such variation in the social function of OT may assist
695 individuals to adapt to changing social contexts (*Shamay-Tsoory and Abu-Akel, 2016; Ma et*
696 *al., 2016*). Unlike previous research that focused on increased activity in the reward system
697 as a consequence of aggressive decisions (*Krämer et al., 2007; Chester and DeWall, 2016*),
698 our findings highlight a neurobiological association from endogenous OT to the mPFC that
699 occurs prior but is linked to revenge propensity during intergroup conflicts. Our results open

a new avenue toward understanding the neurobiological mechanisms mediating aggression-related hormones and social decisions related to intergroup hostility and provide a neuroscientific account of revenge motives during intergroup conflict.

The revenge propensity shown in the Revenge group cannot simply be explained by OT-induced negative emotions. Increasing evidence suggests that the oxytocinergic system is involved in modulating multiple social emotions that are either positively (e.g., empathy, *Sheng et al., 2013*) or negatively (e.g., schadenfreude, *Shamay-Tsoory et al., 2009*) related to social behaviors. Our participants from both the Revenge and Control groups reported greater schadenfreude when viewing outgroup than ingroup suffering. Schadenfreude has been linked to striatum activation induced by misfortunes happening to envied persons (*Takahashi et al., 2009*). While previous research has shown that viewing outgroup pain can activate the nucleus accumbens (NAcc) and greater NAcc activity predicted less motives to help outgroup members (*Hein et al., 2010; Luo et al., 2015*), neither the Revenge group nor the Control group in our work showed activations in the reward system when viewing outgroup members' suffering. Thus negative emotion such as schadenfreude may play a minimal role in modulating revenge propensity in our experimental settings.

Although our measures of endogenous OT and brain activity suggest that the association between endogenous OT to mPFC activity in response to perceived ingroup suffering is related to revenge tendencies, we noted that the latter were estimated with self-reports. It is unclear whether such measures were actually correlated with revengeful behavior. To test this, we conducted an independent behavioral experiment in a new sample (see Supplementary files 10 and 11). The experimental procedures were the same as those in our fMRI experiment except that Phase 3 was modified in the following way. While viewing Involved_Ingroup and Involved_Outgroup targets who played the competitive game with each other (Revenge group) or with a computer (Control group), participants were

occasionally (in 4 trials, 2 on Involved_Ingroup targets and 2 on Involved_Outgroup targets) asked to make punishment decisions by giving the intensity of electric shocks on a Likert Scale (1 = not painful at all, 9 = extremely painful) which were believed to result in painful or non-painful electric shocks to the targets. To enhance participants' beliefs about the experimental setting, they only viewed Involved_Ingroup and Involved_Outgroup targets during the game and they made punishment decisions simultaneously but in different rooms. Moreover, after each punishment decision, a feedback face with either painful or neutral expression, depending on a participant's decision, was presented to inform the consequence of his decision. The results showed evidence that the measures of punishment tendencies were positively correlated with the measures of actual punishment decisions toward Involved_Outgroup target in both Revenge and Control groups ($r=0.75$ and 0.70, FDR corrected $ps < 0.001$). These results indicate that our measures of punishment tendencies can, to a certain degree, reflect individuals' punishment decisions with real consequences.

Our findings also raise new questions about the role of other brain regions in the process of revenge. For example, recent research has shown that, in a scenario in which an observer punishes transgressors due to social norm violation (i.e., third-party punishment), the willingness to punish severely was associated with increased amygdala activity (*Stallen et al., 2018*), possibly reflecting encoding of affective arousal associated with harm done to someone else (*Buckholtz and Marois, 2012; Krueger and Hoffman, 2016*). The current work, however, did not find evidence for an association between amygdala activity and punishment tendencies during intergroup conflict. It is possible that punishment decisions toward outgroup in the context of intergroup conflict are justified as revenge that reduces ingroup suffering and thus bring less negative arousal. Future research is needed to support this speculation.

Our work also expands the literature to examine neural responses implicated in

vicarious revenge which occurs when a person punishes an outgroup member who is not one of the direct causal agents in the original attack against an ingroup member (*Lickel et al*., *2006; Gelfand et al*., *2012; Lee et al*., *2013*). Neither the agent of retaliation nor the target of retribution are directly involved in the original conflict during vicarious retribution, similar to punishment toward Uninvolved-Outgroup targets in our work. We showed that the mPFC activity in response to ingroup pain similarly predicted punishment tendencies toward Involved-Outgroup target and Uninvolved-Outgroup target. The mPFC activity also mediated the relationships between endogenous OT and tendency to punish Involved_Outgroup targets as well as between endogenous OT and tendency to punish Uninvolved_Outgroup targets. Thus, our findings suggest a neurobiological correlate of punishment tendency during intergroup conflict that does not differentiate between direct and indirect vicarious retribution. This is possibly due to that outgroup members are perceived as a unified and coherent entity and share the same blameworthy qualities during intergroup conflict (*McConnell et al*., *1997: Crawford et al*., *2002; Lee et al*., *2013*).

In conclusion, by integrating a neural-behavioral paradigm with fMRI, we showed evidence that intergroup conflict is associated with increased salivary levels of OT in humans which further predicted stronger mPFC activity in response to ingroup suffering caused by an outgroup member. Moreover, the mPFC activity mediates the association between endogenous OT and propensity to seek revenge by giving painful electric shocks to outgroup. Our findings highlight the coupling of the OT system and the mPFC as a neurobiological correlate of revenge propensity during intergroup conflict. Our paradigm can be applied to other samples (e.g., females) and cultures (e.g., the U.S. where individualism is dominant) to advance our understanding of the neurobiological underpinnings of revenge propensity and behavior during intergroup conflict.

Finally, because other motivations also drive revenge behavior in intergroup contexts,

including feeling threat to group pride (*Turner and Tajfel, 1986*), empathy for the harmed ingroup members (*Smith et al., 1999; Davis, 2018*), and normative pressure to avenge the ingroup (*Deutsch and Gerard, 1955*), future research should examine different motivations driving revenge and concomitant emotions that become activated in a host of revenge situations.

METHODS

Participants

Our fMRI experiment recruited 44 male Chinese university students for Revenge Group (mean age±SD = 23.27±2.76 yrs) and 44 male Chinese university students for Control Group (mean age±SD = 23.89±2.16 yrs). Four participants from each group were excluded from fMRI data analyses due to their excessive head movements during scanning, leaving 40 participants in each group being included for data analyses (Revenge Group: mean age±SD = 23.20±2.78 yrs; Control Group: mean age±SD = 23.70±2.03 yrs). Our behavioral experiment recruited independent samples of 40 male Chinese university students for Revenge Group (mean age±SD = 22.50±2.75 yrs) and 39 male Chinese university students for Control Group (mean age±SD = 21.56±2.23 yrs, one participant from Control group dropped out and was substituted by an additional confederate). Demographic information and psychological traits of Revenge and Control groups are shown in Supplementary file 1. All participants were right-handed, had normal or corrected-to-normal vision and reported no neurological or psychiatric history. All participants were paid for their participation. Informed consent was obtained from all participants prior to the experiment. Experimental protocols were approved by the Research Ethics Committee at the School of Psychological and Cognitive Sciences (#2015-12-04), Peking University, complying with the Declaration of Helsinki. The images used in Figures 1 and 6 are photographs of the confederates and the consent to publish was obtained.

The sample size was estimated using G*Power (Faul et al., 2009). Because we aimed to assess the association between brain responses to perceived ingroup members' suffering and retaliation upon outgroup members, the first power analysis estimated the sample size that allowed detection of reliable brain activities in response to others' pain (e.g., the contrast of painful vs. neutral expressions). Based on the previous fMRI study of empathy for pain (*Han et al., 2017*), the effect size of brain activities in response to others' suffering (including aMCC, bilateral AI and bilateral SII) was between 0.39 and 0.84 (a middle effect size). Based on G*Power estimation, a sample size of 34 participants for each group was required to obtain a middle effect size of 0.5 with an error probability of 0.05 and power of 0.80 in paired t-tests (two-tails).

Because there is no previous research allowing us to conduct an experience-based estimation of the effect size of Revenge/Control group difference in salivary OT level, we selected a middle effect size of 0.25 for sample size estimation. To test the difference in endogenous OT levels between Revenge and Control groups, we planned to conduct an ANOVA with Time (Time 1, 2, 3) as a within-subjects variable and Group (Revenge vs. Control) as a between-subjects variable. To detect a significant main effect of Group required a total sample size of 86 with an error probability of 0.05 and power of 0.8, given the correlation among repeated measures (0.5). To detect a significant interaction between Time and Group required a total sample size of 28 with an error probability of 0.05 and power of 0.8, given the correlation among repeated measures (0.5) and the nonsphericity correction (1).

Behavioral and imaging procedures

On each testing day, 4 participants and 2 confederates were recruited. Each band of 4 participants was alternately assigned to the Revenge or Control group in order to balance the sample size of the two groups. Participants and confederates had not known each other before their participation. The experimental procedure consisted of three phases starting at 9:00 am

on the testing day.

Phase 1: Group formation

Upon arrival at a testing room, three photos were taken from each participant (including confederates). One ID photo with neutral expression was used for estimation of attitudes and judgments of group identity. The ID photo and other two photos with neutral or painful expressions were used during fMRI scanning (Figure 1). The photos were taken by asking participants to show a neutral expression or a painful expression (asking participants to imagine a painful experience) that was characterized with facial movements including brow lowering, orbit tightening, and raising of the upper lip (*Prkachin, 1992*). The photos from all participants were modified to the same size (400 x 600 pixel for ID-photos and 400 x 500 pixel for the two photos with neutral or painful expressions).

All participants including confederates were asked to complete questionnaires to estimate self-esteem (*Rosenberg, 1965*), extroversion-introversion (*Eysenck and Eysenck, 1975*), self-construal (*Singelis, 1994*), individualism/collectivism (*Triandis and Gelfand, 1998*), trait empathy (*Davis, 1983*), and trait aggression (*Buss and Perry, 1992*). Subjective socioeconomic status was assessed using a ladder with 10 rungs (*Kilpatrick and Cantril, 1960*). The participants were informed that they would be divided into two groups based on the results of questionnaire measures, though they were actually randomly assigned to two groups so that there were one confederate and two participants in each group.

Participants from each group were asked to wear T-shirts of the same color (red or blue, Figure 1A). Participant introduced their own names, nicknames, majors, and hobbies to get familiar with each other. Participants then started to play the *Saboteur* card game (http://www.annarbor.com/entertainment/saboteur-card-game-review/). During this game ingroup members played cards to build a tunnel to a destination where gold is located or to block the tunnel to prevent outgroup members from reaching the goal. This game required

ingroup members to cooperate with each other but to interfere with outgroup members so as to reach the destination before the outgroup. The intergroup relationship was built by playing this game for 90 minutes. To check the effectiveness of the group manipulation, after the game, participants were asked to complete a modified version of the Inclusion of Other in the Self Scale (*Aron et al., 1992*) to assess their feelings of closeness between oneself and ingroup members and between oneself and outgroup members. Phase 1 lasted for 160 minutes.

Phase 2: Inducing intergroup conflict

After Phase 1, a participant was led to another test room where the two confederates in representation of each group were supposed to be playing a competitive game. During this game, the two confederates performed the classic Stroop task (*Stroop, 1935*) by responding to colors of words by button press. Participants from the Revenge Group were informed that the two confederates (one from ingroup and one from outgroup) competed with each other to make the most correct responses. After 3 trials, the winner who made more correct responses or responded faster then decided whether to give the rival a painful electric shock (as an index of aggression). A pair of foil electrodes connected to an instrument (DS7A Digitimer) for generation of electric shocks and the left hand of each confederate. The participants witnessed that the confederate who won first pressed a button on the instrument to give a non-painful shock to the loser and the confederate who lost showed a neutral expression. The other confederate who won later, however, chose to give a painful shock to the loser by saying "I am curious about how painful an electric shock can be". The confederate who received the electric shock then showed a painful expression to indicate that he was experiencing painful feelings. These confederates' performances provided a cue of how intergroup conflict was initiated. Participants from the Control Group were informed that each confederate performed the Stroop task on his own. After 3 trials, the confederate who

875 performed worse than a standard (with 30% accuracy and reaction times shorter than 2000
876 ms or 100% accuracy and reaction times shorter than 300 ms) would receive a painful or non-
877 painful electric shock randomly given by a computer. One confederate illustrated receiving a
878 painful shock and the other confederate illustrated receiving a non-painful shock. Phase 2
879 lasted for 15 minutes.
880 *Phase 3: Viewing intergroup conflict and reporting punishment tendencies*
881 Before being transported into the MRI scanner, the participant was informed that the
882 two confederates in the test room (one ingroup member and one outgroup member, named as
883 Involved_Ingroup and Involved_Outgroup targets, respectively) would keep playing the
884 game and the winner would decide whether to give the loser a painful or non-painful shock.
885 The participant would be able to see a photo of the loser's face indicating that he was
886 experience of painful or non-painful feelings inside the scanner. During four fMRI scans, a
887 fixation was first presented with its duration varying among 2, 4, 6, 8 s on each trial (Figure
888 1B). An ID-photo of the Involved_Ingroup or Involved_Outgroup target was then presented
889 for 2 s to indicate the person who lost the game. The participant had to judge whether the ID-
890 photo showed an ingroup or an outgroup member by pressing one of two buttons on a
891 response box (the relationship between left/right buttons and ingroup/outgroup members was
892 counter-balanced across participants). Thereafter, the winner's choice, either a yellow circle
893 to indicate a non-painful shock or a yellow lightning symbol to indicate a painful shock, was
894 presented for 2 s. After a fixation with a duration varying among 2, 4, 6, 8 s, a photo of the
895 loser's face was presented for 2 s to indicate being shocked (a photo with neutral expression
896 indicated receiving a non-painful shock and a photo with painful expression indicated
897 receiving a painful shock). The participant was asked to view the photo without any response.
898 The ID-photos, lightning (and round) symbols and photos with expression were subtended a
899 visual angel of 7.58°×11.35°, 3.79°×3.79°and 7.58°×9.47° (width×height) at a viewing

distance of 80 cm, respectively. Each scan started with a 6 s fixation, and a task instruction was presented for 10s followed by 16 trials. The procedure was programmed so that both Involved_Ingroup and Involved_Outgroup targets lost the game on half of the trials and, when losing the game, received painful shocks on half of the trials and non-painful shocks on the other trials. The trials in which a target received painful or non-painful shocks were presented in a random order.

The participant was also informed that, when Involved_Ingroup and Involved_Outgroup targets took a break during the competitive game, the participants had to perform a task to discriminate an ingroup member and an outgroup member who were not involved in the competitive game (named as Uninvolved_Ingroup and Uninvolved_Outgroup targets, respectively). On each trial, a fixation was first presented with its duration varying among 2, 4, 6, 8 s. An ID-photo of an Uninvolved_Ingroup or Uninvolved_Outgroup target was then presented for 2 s. The participant had to judge whether the ID-photo showed an ingroup or an outgroup member by pressing one of two buttons on a response box. Thereafter, a photo of the target with neutral or painful expression was presented for 2 s (Figure 1B). The participant was asked to view the photo without any response. Similarly, there were four scans during which participants viewed Uninvolved_Ingroup and Uninvolved_Outgroup targets. Each scan started with a 6-s fixation, and task instruction was presented for 10 s followed by 16 trials. The procedure was programmed so that both Uninvolved_Ingroup and Uninvolved_Outgroup targets showed painful expressions on half of the trials and showed neutral expressions on the other trials.

The scanning procedure was divided into two sessions. In each session, there were 2 scans when participants viewed Involved_Ingroup and Involved_Outgroup targets and 2 scans when participants viewed Uninvolved_Ingroup and Uninvolved_Outgroup targets. The order of the 4 scans in each session was counterbalanced across participants. After the first

session, the participants were presented with photos of painful expressions of the four targets (Involved_Ingroup, Involved_Outgroup, Uninvolved_Ingroup, and Uninvolved_Outgroup) and rated their emotions for each target on a Likert Scale (1=not at all, 9=extremely strong) in response to the following questions: "How painful do you think the target was?", "How unpleasant were you when viewing the target's pain?", "How angry were you when viewing the target's pain?", "How fearful were you when viewing the target's pain?", and "How happy were you when viewing the target's pain?". After the second session, the participants were presented with ID photos of the four targets and had to rate their attitudes toward the targets on a Likert Scale (1=not at all, 9=extremely) in response to the following questions: "How much do you trust the target?" and "How much do you like the target?". The participants were also asked to report their punishment tendencies by selecting the intensity of an electric shock they would like to apply to the loser on a Likert Scale (1 = not painful at all, 9 = intolerably painful. The order of the rating tasks was counter-balanced across participants. Phase 3 lasted for 60 minutes for each participant.

Measures of endogenous OT

Participants were asked to not drink alcohol, caffeine, or medication within 24 hours prior to their participation. Participants were asked to rinse their mouths with water immediately after lunch. Saliva was collected from each participant at three points in time. The first collection was conducted at the end of Phase 2 (e.g., after the introduction of intergroup conflict, Time 1). The second collection was conducted immediately after Phase 3 outside the scanner (i.e., after viewing intergroup conflict and reporting punishment tendencies during fMRI scanning, Time 3), and the third collection was conducted 15 minutes later (Time 3). Participants were asked to place a roll of cotton in their mouths and to chew on it for a minute until it became saturated. The roll of cotton was then placed in a Salivette (Sarstedt, Rommelsdorft, Germany). The samples were stored at -20℃ until assayed. OT

levels were assayed using a 96-plate commercial OT-ELISA kit (ADI-900-153A; Enzo Life Science). Measurements were performed in duplicate according to the kit's instructions, similar to the procedure of the previous studies (*Van IJzendoorn et al., 2012; Bhandari et al., 2014; Tsuji et al., 2015*). The optical density of the samples and standards was measured at wavelengths of 405 nm, with correction between 570 and 590 nm. Four parameter logistics curve fitting program was used for the calculation of the concentration of OT in the samples. Due to the failure of OT measurements on a few participants, 37 and 40 participants were left in the Revenge and Control groups, respectively, for all further analyses related to OT levels.

fMRI data acquisition and analysis

Brain images were acquired using a 3.0T GE Signa MR750 scanner (GE Healthcare; Waukesha, WI) with a standard 8 channel head coil. Functional images were acquired by using T2-weighted, gradient-echo, echo-planar imaging (EPI) sequences sensitive to blood oxygenation level dependent (BOLD) signals (64 × 64 × 32 matrix with 3.75 × 3.75 × 5 mm^3 spatial resolution, repetition time = 2000 ms, echo time = 30 ms, flip angle = 90°, field of view = 24 × 24 cm). A high-resolution T1-weighted structural image (512 × 512 × 180 matrix with a spatial resolution of 0.47 × 0.47 × 1.0 mm^3, repetition time = 8.204ms, echo time = 3.22ms, flip angle = 12°) was acquired after the first four scans. Padded clamps were used to minimize head motion and earplugs were used to attenuate scanner noise. The stimuli were projected onto a screen at the head of the magnet bore using Presentation. Participants viewed the screen through a mirror attached to the head coil.

Functional images were preprocessed using SPM8 software (the Wellcome Trust Centre for Neuroimaging, London, UK). Head movements were corrected within each run and six movement parameters (translation; x, y, z and rotation; pitch, roll, yaw) were extracted for further analysis in the statistical model. The functional images were resampled to 3 × 3 × 3 mm^3 voxels, normalized to the Montreal Neurological Institute (MNI) template

space and then spatially smoothed using an isotropic of 8 mm full-width half-maximum (FWHM) Gaussian kernel. Four participants from each group (i.e., Revenge and Control group) were excluded from fMRI data analysis due to that their head movements exceeded 5 mm. Hence, 40 subjects in each group were included in further fMRI data analysis, which was identical for Revenge and Control Groups. In the first general linear model (GLM) brain activations were estimated using 8 onset regressors to identify brain activations in response to painful vs. neutral expressions. These included 1) ID photos of Involved_Ingroup and Uninvolved_Ingroup targets; 2) ID photos of Involved_Outgroup and Uninvolved_Outgroup targets; 3) the symbol of painful shocks for involved targets, or the 2-s fixation for uninvolved targets; 4) the symbol of non-painful shocks for involved targets, or the 2-s fixation for uninvolved targets (Figure 1B); 5) photos of painful expressions of Involved_Ingroup and Uninvolved_Ingroup targets; 6) photos of neutral expressions of Involved_Ingroup and Uninvolved_Ingroup targets; 7) photos of painful expressions of Involved_Outgroup and Uninvolved_Outgroup targets; 8) photos of neutral expressions of Involved_Outgroup and Uninvolved_Outgroup targets.

We conducted separate GLM analyses of involved targets (using the above onset regressors but distinguished the symbols of painful/non-painful shocks for Involved_Ingroup and Involved_Outgroup targets) and uninvolved targets for the whole brain regression analysis of the effect of endogenous OT and region-of-interest (ROI) analyses. The GLMs included the realignment parameters to account for any residual movement-related effect. The voxels showing significant event-related responses to painful vs. neutral expressions were created using a canonical haemodynamic response function (HRF). Here our fMRI analyses locked BOLD responses to the onset of facial expressions rather than symbols of shock decisions because we focused on the relationship between OT-level and empathic neural responses to targets' pain. The onset of a symbol of a shock decision indicated what type of

shocks (painful or nonpainful) the winner was going to give to the loser. The onset of a face with a painful or nonpainful expression indicated the time when the target started to receive a painful or nonpainful shock and an emotional (painful or nonpainful) response was initiated. A whole-brain random effect analysis was then conducted to reveal brain regions that showed reliable responses to painful vs. neutral expressions of all targets. Brain activations were defined using a voxel-level threshold of $p < 0.001$, uncorrected and cluster-level threshold of $p < 0.05$, FWE corrected.

ROI analyses

ROI analyses were conducted to test (1) ingroup favoritism in neural responses to perceived pain, (2) group differences in the association between endogenous OT and mPFC activity in response to Involved_Ingroup target's pain, and (3) the correlation between mPFC activity in response to Involved_Ingroup target's pain and punishment propensity towards outgroup members. ROI analyses were also conducted in the mediation analyses which tested the mediation role of mPFC activity in response to Involved_Ingroup target's pain in the association between endogenous OT and punishment tendency towards outgroup members. To define the coordinates of ROIs independently, a leave-one-out test in which whole-brain analyses of the contrast of painful vs. neutral expressions which collapsed all the targets was conducted using 79 of the 80 participants from the two subject groups using a combined voxel-level threshold of $p < 0.001$, uncorrected and cluster-level threshold of $p < 0.05$, FWE corrected. ROI coordinates were defined at the peak voxel the corresponding brain regions for the left-out participant. ROIs were then defined as a sphere with 5-mm-radius centered at the peak voxel of the seed regions for the left-out participant. The contrast values were extracted using MarsBaR (http://marsbar.sourceforge.net).

In the ROI analyses of the ingroup favoritism in neural responses to perceived pain, ROIs included the bilateral AI/IFG in the empathy network and the left TPJ and mPFC in the

theory-of-mind network. The contrast values of these ROIs were extracted from the whole brain analyses of Ingroup painful vs. neutral expression and Outgroup painful vs. neutral expression. In the ROI analyses of moderation and mediation analyses, the mPFC activity and left TPJ activity in response to Involved_Ingroup targets' pain was extracted from the whole brain analysis of Involved_Ingroup target's painful vs. neutral expression.

Moderation analysis

To examine whether intergroup conflict moderated the associations between endogenous OT at Time 1 and mPFC activity (or left TPJ activity) in response to Ingroup_In target's pain, we performed moderated hierarchical regression analyses. To do this, we first defined ROIs using the leave-one-out method by calculating the contrast of painful vs. neutral expressions of one target by including 79 participants from the two subject groups. ROIs were defined as a sphere with 5-mm-radius centered at the peak coordinate of the mPFC or left TPJ. The contrast values of painful vs. neutral expressions were extracted using MarsBaR (http://marsbar.sourceforge.net) from the left-out participant. We then dummy coded the Group variable (i.e., Revenge and Control groups) as 0 and 1. The Group variable (the moderator), the contrast value of the ROI (the independent variable) and the OT level at Time 1 (the dependent variable) were entered into Hayes's PROCESS macro (Model 1) (*Hayes, 2017*). Additionally, ingroup bias in closeness, emotions and attitudes were entered into the model as covariates. The moderator effect was indicated by a significant interaction effect between the moderator and the independent variable.

To examine whether involvement of the targets (e.g., Involved_Ingroup target vs. Uninvolved_Ingroup target) moderated the association between endogenous OT and mPFC activity in response to Ingroup_In target's pain, we performed another moderation analysis with a repeated measure as the moderator. To do this, we conducted a regression analysis with the endogenous OT level at Time 1 as the independent variable and the difference of the

mPFC activities towards the Involved_Ingroup target and Uninvolved_Ingroup target as the dependent variable. The moderation effect was indicated if the independent variable significantly predicts the dependent variable. The analyses were performed using Montoya's MEMORE macro (Model 2, *Montoya, 2019*).

Mediation analysis

We performed mediation analyses to examine whether the mPFC activity mediates the pathway from the endogenous OT to punishment tendency. To do this, we estimated four regression models: 1) whether the independent variable (OT) significantly accounts for the dependent variable (punishment tendency) when not considering the mediator (e.g., Path c'); 2) whether the independent variable (OT) significantly accounts for the variance of the presumed mediator (mPFC activity) (e.g., Path a); 3) whether the presumed mediator (mPFC activity) significantly accounts for the variance of the dependent variable (punishment tendency) when controlling the independent variable (OT) (e.g., Path b); 4) whether the independent variable (OT) significantly accounts for the variance of the dependent variable (punishment tendency) when controlling the presumed mediator (mPFC activity) (e.g., Path c). To establish the mediation, the path c' is not required to be significant, and the only requirement is that the indirect effect a x b is significant. Given a significant indirect effect, if Path c is insignificant, the mediation is classified as indirect-only mediation which is the strongest full mediation (*Kenny et al., 1998; Zhao et al., 2010*). A bootstrapping method was used to estimate the mediation effect. Bootstrapping is a nonparametric approach to estimate the effect-size and test the hypothesis that is increasingly recommended for many types of analyses, including mediation (*Shrout and Bolger, 2002; MacKinnon et al., 2004*). Rather than imposing questionable distributional assumptions, bootstrapping generates an empirical approximation of the sampling distribution of a statistic by repeated random resampling from the available data, and uses this distribution to calculate p-values and construct confidence

intervals. 5,000 resamples were taken for our analyses. Moreover, this procedure supplies superior confidence intervals (CIs) that are bias-corrected and accelerated (*Preacher et al*., *2007; Preacher and Hayes, 2008b; Preacher and Hayes, 2008a*). The analyses were performed using Hayes's PROCESS macro (Model 4, Hayes, 2017).

Additional information

Funding

Funder	Grant reference number	Author
the National Natural Science Foundation of China	31661143039, 31421003, 31470986	Shihui Han

Ethics

Human subjects: Informed consent was obtained from all participants prior to the experiment. Experimental protocols were approved by the Research Ethics Committee at the School of Psychological and Cognitive Sciences (#2015-12-04), Peking University, complying with the Declaration of Helsinki. The images used in Figures 1 and 6 are photographs of the confederates and the consent to publish was obtained.

Additional files

Supplementary files

- Figure 2 - Source data 1. Source data for Figure 2A.
- Figure 2 - Source data 2. Source data for Figure 2B.

- Figure 3 - Source data 3. Source data for Figure 3A.
- Figure 3 - Source data 4. Source data for Figure 3B.
- Figure 4 - Source data 5. Source data for Figure 4A.
- Figure 4 - Source data 6. Source data for Figure 4B _Ingroup.
- Figure 4 - Source data 7. Source data for Figure 4B _Outgroup.
- Figure 4 - Source data 8. Source data for Figure 4C.
- Figure 5 - Source data 9. Source data for Figure 5A _Time 1.
- Figure 5 - Source data 10. Source data for Figure 5A _Time 2.
- Figure 5 - Source data 11. Source data for Figure 5B.
- Figure 5 - Source data 12. Source data for Figure 5C.
- Figure 6 - Source data 13. Source data for Figure 6A.
- Source code file 1. Scripts for the Bootstrap analysis of OT levels in Figure 3B.
- Source code file 2. Scripts for the whole-brain analysis in Figure 4A.
- Source code file 3. Scripts for the whole-brain analysis in Figure 4B _Ingroup.
- Source code file 4. Scripts for the whole-brain analysis in Figure 4B _Outgroup.
- Source code file 5. Scripts for the whole-brain regression analysis _Time 1.
- Source code file 6. Scripts for the whole-brain regression analysis _Time 2.
- Supplementary file 1. Demographic information and psychological traits of the participants in the fMRI experiment. This file shows the means (SD) and statistics for comparisons between the Revenge and Control groups.
- Supplementary file 2. Results of group manipulation check. This file shows the means (SD) of emotions and attitudes and statistics for comparisons between the Revenge and Control groups.
- Supplementary file 3. Factorial models of emotion and attitude rating items. This files

shows the results of a factorial analysis that tested the discriminant validity of the eight items related to measures of emotions and attitudes. The analysis revealed two factors, which explained 62.60% of total variance. Factor 1 was the emotion factor (explaining 37.65% of variance) which included five items, i.e., empathy (0.653), unpleasant (0.906), anger (0.748), fear (0.837), and schadenfreude (-0.349). Factor 2 was the attitude factor (explaining 24.95% of variance) which included two items, i.e., likability (0.907) and trust (0.918).

- Supplementary file 4. Brain activations elicited by painful vs. neutral expressions across the Revenge and Control groups. This file shows MNI coordinates of activated brain regions, cluster sizes, and Z values.
- Supplementary file 5. Brain activations elicited by painful vs. neutral expressions in the Revenge group. This file shows MNI coordinates of activated brain regions, cluster sizes, and Z values.
- Supplementary file 6. Brain activations elicited by painful vs. neutral expressions in the Control group. This file shows MNI coordinates of activated brain regions, cluster sizes, and Z values.
- Supplementary file 7. The results of the moderation analysis. This file shows the statistical details of the moderation analysis that examined how group identity (Revenge vs. Control group) moderated the relationship between endogenous OT (Time 1) and mPFC activity in response to Involved_Ingroup target's pain.
- Supplementary file 8. The results of the mediation analysis. This file shows the statistical details of the moderation analysis that examined whether the mPFC activity mediated the relationship between endogenous OT (Time 1) and punishment tendencies towards the Involved_Outgroup target.
- Supplementary file 9. The results of the mediation analysis. This file shows the statistical details of the moderation analysis that examined whether the mPFC activity mediated the

relationship between endogenous OT (Time 1) and punishment tendencies towards the Uninvolved_Outgroup target.

- Supplementary file 10. Demographic information and psychological traits of the participants in the new behavioral experiment. This file shows the means (SD) and statistics for comparisons between the Revenge and Control groups.
- Supplementary file 11. Ingroup favoritism in self-report of emotions, attitudes, punishment tendencies, and punishment decisions in the new behavioral experiment. This file shows the means (SD) and statistics for comparisons between the Revenge and Control groups.
- Transparent reporting form.

References

Allport GW, Clark K, Pettigrew T. 1954. *The nature of prejudice*. Reading, MA: Addison-Wesley.

Amodio D, Frith C. 2006. Meeting of minds: The medial frontal cortex and social cognition. *Nature Review Neuroscience* **7**: 268–277. DOI: http://dx.doi.org/10.1038/nrn1884

Aron A, Aron EN, Smollan D. 1992. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology* **63**: 596-612. DOI: http://dx.doi.org/10.1037/0022-3514.63.4.596

Bhandari R, Bakermans-Kranenburg MJ, van der Veen R, Parsons CE, Young KS, Grewen KM, Stein A, Kringelbach ML, van IJzendoorn MH. 2014. Salivary oxytocin mediates the association between emotional maltreatment and responses to emotional infant faces. *Physiology & Behavior* **131**: 123-128. DOI: http://dx.doi.org/10.1016/j.physbeh.2014.04.028

Boccia ML, Petrusz P, Suzuki K, Marson L, Pedersen CA. 2013. Immunohistochemical localization of oxytocin receptors in human brain. *Neuroscience* **253**: 155-164. DOI: http://dx.doi.org/10.1016/j.neuroscience.2013.08.048

Brewer MB. 1999. The psychology of prejudice: Ingroup love and outgroup hate? *Journal of Social Issues* **55**: 429-444. DOI: http://dx.doi.org/10.1111/0022-4537.00126

Buckholtz JW, Marois R. 2012. The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience* **15**: 655-661. DOI: http://dx.doi.org/10.1038/nn.3087

Buss AH, Perry M. 1992. The aggression questionnaire. *Journal of Personality and Social Psychology* **63**: 452-459. DOI: http://dx.doi.org/10.1037/0022-3514.63.3.452

Chester DS, DeWall CN. 2016. The pleasure of revenge: retaliatory aggression arises from a neural imbalance toward reward. *Social Cognitive and Affective Neuroscience* **11**: 1173-1182. DOI: http://dx.doi.org/10.1093/scan/nsv082

Cikara M, Botvinick MM, Fiske ST. 2011. Us versus them: Social identity shapes neural responses to intergroup competition and harm. *Psychological Science* **22**: 306-313. DOI: http://dx.doi.org/10.1177/0956797610397667

Crawford MT, Sherman SJ, Hamilton DL. 2002. Perceived entitativity, stereotype formation, and the interchangeability of group members. *Journal of Personality and Social Psychology* **83**: 1076-1094. DOI: http://dx.doi.org/10.1037/0022-3514.83.5.1076

Davis MH. 1983. The effects of dispositional empathy on emotional reactions and helping: A multidimensional approach. *Journal of Personality* **51**: 167-184. DOI: http://dx.doi.org/10.1111/j.1467-6494.1983.tb00860.x

Davis MH. 2018. *Empathy: A social psychological approach*, Routledge. DOI: http://dx.doi.org/10.4324/9780429493898

Davison AC, Hinkley DV. 1997. *Bootstrap methods and their application*, Cambridge University Press. DOI: http://dx.doi.org/10.1017/CBO9780511802843

De Dreu CK. 2010. Social value orientation moderates ingroup love but not outgroup hate in competitive intergroup conflict. *Group Processes & Intergroup Relations* **13**: 701-713. DOI: http://dx.doi.org/10.1177/1368430210377332

De Dreu CK, Greer LL, Handgraaf MJ, Shalvi S, Van Kleef GA, Baas M, Ten Velden FS, Van Dijk E, Feith SW. 2010. The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science* **328**: 1408-1411. DOI: http://dx.doi.org/10.1126/science.1189047

De Dreu CK, Greer LL, Van Kleef GA, Shalvi S, Handgraaf MJ. 2011. Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences, USA*. **108**: 1262-1266. DOI: http://dx.doi.org/10.1073/pnas.1015316108

Deutsch M, Gerard HB. 1955. A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology* **51**: 629-636. DOI: http://dx.doi.org/10.1037/h0046408

Eckstein M, Becker B, Scheele D, Scholz C, Preckel K, Schlaepfer TE, Grinevich V, Kendrick KM, Maier W, Hurlemann R. 2015. Oxytocin facilitates the extinction of conditioned fear in humans. *Biological Psychiatry* **78**: 194-202. DOI: http://dx.doi.org/10.1016/j.biopsych.2014.10.015

Elshout M, Nelissen R, Van Beest I. 2015. A prototype analysis of vengeance. *Personal Relationships* **22**: 502-523. DOI: http://dx.doi.org/10.1111/pere.12092

Ericksen KP, Horton H. 1992. "Blood Feuds": Cross-Cultural Variations in Kin Group Vengeance. *Behavior Science Research* **26**: 57-85. DOI: http://dx.doi.org/10.1177/106939719202600103

Eysenck HJ, Eysenck SBG. 1975. *Manual of the Eysenck Personality Questionnaire (junior and adult)*, Hodder and Stoughton.

Fan Y, Duncan NW, de Greck M, Northoff G. 2011. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews* **35**: 903-911. DOI: http://dx.doi.org/10.1016/j.neubiorev.2010.10.009

Faul F, Erdfelder E., Buchner, A. & Lang, A. G. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41**,** 1149-1160. DOI: http://dx.doi.org/10.3758/BRM.41.4.1149

Feldman R, Gordon I, Zagoory-Sharon O. 2011. Maternal and paternal plasma, salivary, and urinary oxytocin and parent–infant synchrony: considering stress and affiliation components of human bonding. *Developmental Science* **14**: 752-761. DOI: http://dx.doi.org/10.1111/j.1467-7687.2010.01021.x

Gelfand M, Shteynberg G, Lee T, Lun J, Lyons S, Bell C, Chiao JY, Bruss CB, Dabbagh MA, Aycan A, Abdel-Latif AH, Dagher M, Khashan H, Soomro Z. 2012. The cultural contagion of conflict. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367: 692-703. DOI: https://doi.org/10.1098/rstb.2011.0304

Gimpl G, Fahrenholz F. 2001. The oxytocin receptor system: structure, function, and regulation. *Physiological Reviews* **81**: 629-683. DOI: http://dx.doi.org/10.1152/physrev.2001.81.2.629

Halevy N, Bornstein G, Sagiv L. 2008. "In-group love" and "out-group hate" as motives for individual participation in intergroup conflict: A new game paradigm. *Psychological Science* **19**: 405-411. DOI: http://dx.doi.org/10.1111/j.1467-9280.2008.02100.x

Halevy N, Weisel O, Bornstein G. 2012. "In-group love" and "out-group hate" in repeated interaction between groups. *Journal of Behavioral Decision Making* **25**: 188-195. DOI: http://dx.doi.org/10.1002/bdm.726

Han S. 2018. Neurocognitive basis of racial ingroup bias in empathy. *Trends in Cognitive Sciences* **22**: 400-421. DOI: http://dx.doi.org/10.1016/j.tics.2018.02.013

Han X, He K, Wu B, Shi Z, Liu Y, Luo S, Wei K, Wu X, Han S. 2017. Empathy for pain motivates actions without altruistic effects: evidence of motor dynamics and brain activity. *Social Cognitive and Affective Neuroscience* **12**: 893-901. DOI: http://dx.doi.org/10.1093/scan/nsx016

Harris LT, Fiske ST. 2007. Social groups that elicit disgust are differentially processed in mPFC. *Social Cognitive and Affective Neuroscience* **2**: 45-51. DOI: http://dx.doi.org/10.1093/scan/nsl037

Hayes AF. 2017. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*, Guilford publications.

Hein G, Silani G, Preuschoff K, Batson CD, Singer T. 2010. Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron* **68**: 149-160. DOI: http://dx.doi.org/10.1016/j.neuron.2010.09.003

Jackson JC, Choi VK, Gelfand MJ. 2019. Revenge: A multilevel review and synthesis. *Annual Review of Psychology* **70**: 319-345. DOI: http://dx.doi.org/10.1146/annurev-psych-010418-103305

Kahn DT, Klar Y, Roccas S. 2017. For the sake of the eternal group: perceiving the group as trans-generational and endurance of ingroup suffering. *Personality and Social Psychology Bulletin* **43**: 272-283. DOI: http://dx.doi.org/10.1177/0146167216684123

Kenny DA, Kashy DA, Bolger N. 1998. Data analysis in social psychology. *The Handbook of Social Psychology* **1**: 233-265.

Kilpatrick FP, Cantril H. 1960. Self-anchoring scaling: A measure of individuals' unique reality worlds. *Journal of Individual Psychology* **16**: 158-173.
Kopsaj V. 2016. Blood feud and its impact on the Albanian criminality. *Mediterranean Journal of Social Sciences* **7**: 88-95. DOI: http://dx.doi.org/10.5901/mjss.2016.v7n3s1p88
Krämer UM, Jansma H, Tempelmann C, Münte TF. 2007. Tit-for-tat: the neural basis of reactive aggression. *NeuroImage* **38**: 203-211. DOI: http://dx.doi.org/10.1016/j.neuroimage.2007.07.029
Krueger F, Hoffman M. 2016. The emerging neuroscience of third-party punishment. *Trends in Neurosciences* **39**: 499-501. DOI: http://dx.doi.org/10.1016/j.tins.2016.06.004
Lamm C, Decety J, Singer T. 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage* **54**: 2492-2502. DOI: http://dx.doi.org/10.1016/j.neuroimage.2010.10.014
Lee TL, Gelfand MJ, Shteynberg G. 2013. Culture and the contagion of conflict. In M. Yuki and M. Brewer (Eds.), *Culture and group processes*. New York: Oxford University Press, pp. 241-260.
Levy J, Goldstein A, Influs M, Masalha S, Zagoory-Sharon O, Feldman R. 2016. Adolescents growing up amidst intractable conflict attenuate brain response to pain of outgroup. *Proceedings of the National Academy of Sciences,* **113:** 13696-13701. DOI: https://doi.org/10.1073/pnas.1612903113
Lickel B, Miller N, Stenstrom DM, Denson TF, Schmader T. 2006. Vicarious retribution: The role of collective blame in intergroup aggression. *Personality and Social Psychology Review* **10**: 372-390. DOI: http://dx.doi.org/10.1207/s15327957pspr1004_6
Liu Y, Wu B, Wang X, Li W, Zhang T, Wu X, Han S. 2017. Oxytocin effects on self-referential processing: behavioral and neuroimaging evidence. *Social Cognitive and Affective Neuroscience* **12**: 1845-1858. DOI: http://dx.doi.org/10.1093/scan/nsx116
Luo S, Li B, Ma Y, Zhang W, Rao Y, Han S. 2015. Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity. *NeuroImage* **110**: 22-31. DOI: http://dx.doi.org/10.1016/j.neuroimage.2015.01.042
Ma Y, Shamay-Tsoory S, Han S, Zink CF. 2016. Oxytocin and social adaptation: insights from neuroimaging studies of healthy and clinical populations. *Trends in Cognitive Sciences* **20**: 133-145. DOI: http://dx.doi.org/10.1016/j.tics.2015.10.009
MacDonald K, MacDonald TM. 2010. The peptide that binds: a systematic review of oxytocin and its prosocial effects in humans. *Harvard Review of Psychiatry* **18**: 1-21. DOI: http://dx.doi.org/10.3109/10673220903523615
Mackinnon DP, Lockwood CM, Williams J. 2004. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research* **39**: 99-128. DOI: http://dx.doi.org/10.1207/s15327906mbr3901_4
Mathur VA, Harada T, Lipke T, Chiao JY. 2010. Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage* **51**: 1468-1475. DOI: http://dx.doi.org/10.1016/j.neuroimage.2010.03.025
McConnell AR, Sherman SJ, Hamilton DL. 1997. Target entitativity: Implications for information processing about individual and group targets. *Journal of Personality and Social Psychology* **72**: 750-762. DOI: http://dx.doi.org/10.1037/0022-3514.72.4.750
Molenberghs P, Morrison S. 2014. The role of the medial prefrontal cortex in social categorization. *Social Cognitive and Affective Neuroscience* **9**: 292-296. DOI: http://dx.doi.org/10.1093/scan/nss135
Montoya AK. 2019. Moderation analysis in two-instance repeated measures designs: Probing methods and multiple moderator models. *Behavior Research Methods* **51**: 61-82. DOI: http://dx.doi.org/10.3758/s13428-018-1088-6
Ninan I. 2011. Oxytocin suppresses basal glutamatergic transmission but facilitates activity-dependent synaptic potentiation in the medial prefrontal cortex. *Journal of Neurochemistry* **119**: 324-331. DOI: http://dx.doi.org/10.1111/j.1471-4159.2011.07430.x
Preacher KJ, Hayes AF. 2008a. Contemporary approaches to assessing mediation in communication research. In Hayes AF, Slater MD, Snyder LB (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research*. Thousand Oaks, CA: Sage. pp. 13–54.
Preacher KJ, Hayes AF. 2008b. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods* **40**: 879-891. DOI: http://dx.doi.org/10.3758/BRM.40.3.879
Preacher KJ, Rucker DD, Hayes AF. 2007. Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research* **42**: 185-227. DOI: http://dx.doi.org/10.1080/00273170701341316
Prkachin KM. 1992. The consistency of facial expressions of pain - a comparison across modalities. *Pain* **51**: 297-306.
Raver JL, Barling J. 2008. Workplace aggression and conflict: Constructs, commonalities, and challenges for future inquiry. In De Dreu CK, Gelfand MJ (Eds.), *The Psychology of Conflict and Conflict*

Management in Organizations. Mahawah, NJ: Lawrence Erlbaum Associates, pp. 211–244.
Rosenberg M. 1965. *Society and the adolescent self-image*. Princeton Univ. Press, Princeton. DOI:http://dx.doi.org/10.1515/9781400876136
Ross L, Ward A. 1995. Psychological barriers to dispute resolution. *Advances in Experimental Social Psychology* **27**: 255-304. DOI: http://dx.doi.org/10.1016/S0065-2601(08)60407-4
Rusch H. 2014. The evolutionary interplay of intergroup conflict and altruism in humans: a review of parochial altruism theory and prospects for its extension. *Proceedings of the Royal Society B: Biological Sciences* **281**: 20141539. DOI: http://dx.doi.org/10.1098/rspb.2014.1539
Sabihi S, Durosko NE, Dong SM, Leuner B. 2014. Oxytocin in the prelimbic medial prefrontal cortex reduces anxiety-like behavior in female and male rats. *Psychoneuroendocrinology* **45**: 31-42. DOI: http://dx.doi.org/10.1016/j.psyneuen.2014.03.009
Samuni L, Preis A, Mundry R, Deschner T, Crockford C, Wittig RM. 2017. Oxytocin reactivity during intergroup conflict in wild chimpanzees. *Proceedings of the National Academy of Sciences, USA*. **114**: 268-273. DOI: http://dx.doi.org/10.1073/pnas.1616812114
Schiller B, Baumgartner T, Knoch D. 2014. Intergroup bias in third-party punishment stems from both ingroup favoritism and outgroup discrimination. *Evolution and Human Behavior* **35**: 169-175. DOI: http://dx.doi.org/10.1016/j.evolhumbehav.2013.12.006
Seymour B, Singer T, Dolan R. 2007. The neurobiology of punishment. *Nature Reviews Neuroscience* **8**: 300-311. DOI: http://dx.doi.org/10.1038/nrn2119
Shamay-Tsoory SG. 2011. The neural bases for empathy. *The Neuroscientist* **17**: 18-24. DOI: http://dx.doi.org/10.1177/1073858410379268
Shamay-Tsoory SG, Abu-Akel A. 2016. The social salience hypothesis of oxytocin. *Biological Psychiatry* **79**: 194-202. DOI: http://dx.doi.org/10.1016/j.biopsych.2015.07.020
Shamay-Tsoory SG, Fischer M, Dvash J, Harari H, Perach-Bloom N, Levkovitz Y. 2009. Intranasal administration of oxytocin increases envy and schadenfreude (gloating). *Biological Psychiatry* **66**: 864-870. DOI: http://dx.doi.org/10.1016/j.biopsych.2009.06.009
Sheng F, Han S. 2012. Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. *NeuroImage*, **61:** 786-797. DOI: https://doi.org/10.1016/j.neuroimage.2012.04.028
Sheng F, Liu Y, Zhou B, Zhou W, Han S. 2013. Oxytocin modulates the racial bias in neural responses to others' suffering. *Biological Psychology* **92**: 380-386. DOI: http://dx.doi.org/10.1016/j.biopsycho.2012.11.018
Shrout PE, Bolger N. 2002. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological Methods* **7**: 422-445. DOI: http://dx.doi.org/10.1037/1082-989X.7.4.422
Singelis TM. 1994. The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin* **20**: 580-591. DOI: http://dx.doi.org/10.1177/0146167294205014
Smith ER, Murphy J, Coats S. 1999. Attachment to groups: Theory and management. *Journal of Personality and Social Psychology* **77**: 94-110. DOI: http://dx.doi.org/10.1037/0022-3514.77.1.94
Stallen M, Rossi F, Heijne A, Smidts A, De Dreu CK, Sanfey AG. 2018. Neurobiological mechanisms of responding to injustice. *Journal of Neuroscience* **38**: 2944-2954. DOI: http://dx.doi.org/10.1523/JNEUROSCI.1242-17.2018
Stroop JR. 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* **18**: 643-662. DOI: http://dx.doi.org/10.1037/h0054651
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**: 937-939. DOI: http://dx.doi.org/10.1126/science.1165604
Triandis HC, Gelfand MJ. 1998. Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology* **74**: 118-128. DOI: http://dx.doi.org/10.1037/0022-3514.74.1.118
Tsuji S, Yuhi T, Furuhara K, Ohta S, Shimizu Y, Higashida H. 2015. Salivary oxytocin concentrations in seven boys with autism spectrum disorder received massage from their mothers: a pilot study. *Frontiers in Psychiatry* **6**: 58. DOI: http://dx.doi.org/10.3389/fpsyt.2015.00058
Turner JC, Tajfel H. 1986. The social identity theory of intergroup behavior. *Psychology of Intergroup Relations* **5**: 7-24.
van IJzendoorn MH, Bhandari R, van der Veen R, Grewen KM, Bakermans-Kranenburg MJ. 2012. Elevated salivary levels of oxytocin persist more than 7 h after intranasal administration. *Frontiers in Neuroscience* **6**: 174. DOI: http://dx.doi.org/10.3389/fnins.2012.00174
Volz KG, Kessler T, von Cramon DY. 2009. In-group as part of the self: in-group favoritism is mediated by medial prefrontal cortex activation. *Social Neuroscience* **4**: 244-260. DOI: http://dx.doi.org/10.1080/17470910802553565

Wang D, Yan X, Li M, Ma Y. 2017. Neural substrates underlying the effects of oxytocin: a quantitative meta-analysis of pharmaco-imaging studies. *Social Cognitive and Affective Neuroscience* **12**: 1565-1573. DOI: http://dx.doi.org/10.1093/scan/nsx085

Xu X, Zuo X, Wang X, Han S. 2009. Do you feel my pain? Racial group membership modulates empathic neural responses. *Journal of Neuroscience*, **29:** 8525-8529. DOI:https://doi.org/10.1523/JNEUROSCI.2418-09.2009

Zhang H, Gross J, De Dreu C, Ma Y. 2019. Oxytocin promotes coordinated out-group attack during intergroup conflict in humans. *Elife* **8**: e40698. DOI: http://dx.doi.org/10.7554/eLife.40698

Zhao X, Lynch JG, Chen Q. 2010. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research* **37**: 197-206. DOI: http://dx.doi.org/10.1086/651257