UNIVERSITY OF EXETER

ORE Open Research Exeter

TITLE

The influence of group membership and individual differences in psychopathy and perspective taking on neural responses when punishing and rewarding others

AUTHORS

Molenberghs, P; Bosworth, R; Nott, Z; et al.

JOURNAL

Human Brain Mapping

DEPOSITED IN ORE

19 November 2015

This version available at

http://hdl.handle.net/10871/18708

COPYRIGHT AND REUSE

Open Research Exeter makes this work available in accordance with publisher policies.

A NOTE ON VERSIONS

The version presented here may differ from the published version. If citing, you are advised to consult the published version for pagination, volume/issue and date of publication

Human Brain Mapping

The influence of group membership and individual dispositions on the neural response when punishing and rewarding others.

Journal:	*Human Brain Mapping*
Manuscript ID:	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Molenberghs, Pascal; University of Queensland, School of Psychology Bosworth, Rebecca; University of Queensland, School of Psychology Nott, Zoie; University of Queensland, School of Psychology Louis, Winnifred; University of Queensland, School of Psychology Smith, Joanne; University of Exeter, School of Psychology Vohs, Kathleen; University of Minnesota, Carlsson School of Management Amiot, Catherine; Université du Québec à Montréal, Département de psychologie Decety, Jean; University of Chicago, Department of Psychology and Department of Psychiatry and Behavioral Neuroscience
Keywords:	fMRI, social neuroscience, group membership, rewarding, punishing, in-group bias, discrimination, psychopathy, Theory of Mind, empathy

SCHOLARONE™
Manuscripts

The influence of group membership and individual dispositions on the neural response when punishing and rewarding others.

Pascal Molenberghs[1], Rebecca Bosworth[1], Zoie Nott[1], Winnifred R. Louis[1], Joanne R. Smith[2], Kathleen D. Vohs[3], Catherine E. Amiot[4], Jean Decety[5,6]

[1] School of Psychology, The University of Queensland, Brisbane, Australia

[2] School of Psychology, University of Exeter, Exeter, UK

[3] Carlsson School of Management, University of Minnesota, United States

[4] Département de psychologie, Université du Québec à Montréal, Montréal, Québec, Canada

[5] Department of Psychology and Center for Cognitive and Social Neuroscience, The University of Chicago, Illinois, United States

[6] Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Illinois, United States

Correspondence should be addressed to p.molenberghs@uq.edu.au.

Tel: +617 3365 6257

Fax: +617 3365 4466

School of Psychology
McElwain Building
The University of Queensland
St Lucia, QLD 4072
Australia

Number of Figures: 3

Keywords: fMRI, social neuroscience, group membership, rewarding, punishing, in-group bias, discrimination, psychopathy, Theory of Mind, empathy

Running Title: Rewarding and hurting in-group and out-group members

For Peer Review

Abstract

Understanding how the neural correlates involved in punishing and rewarding others are influenced by group membership and personality is important in order to gain a better understanding of how complex social problems such as racism and in-group bias develop. In this fMRI study, forty-eight participants gave rewards (money) or punishments (electroshocks) to in-group or out-group members. The present results provide the first empirical comparison of personally delivering rewards and punishments to in-group and out-group members. The results show that when participants reward others, more activation is found in regions typically associated with receiving rewards such as the bilateral putamen and medial orbitofrontal cortex. In addition, more activation is found in these regions, when participants reward in-group members compared to out-group members. However, when participants punish others, more activation is found in regions typically associated with Theory of Mind (medial prefrontal cortex and posterior superior temporal sulcus) and watching others in pain (dorsal anterior cingulate cortex, bilateral anterior insula and lateral orbitofrontal cortex). Activity in these regions did not modulate by group membership. Additional regression analysis also show that people who have poor perspective taking skills and higher levels of psychopathy show less activation in these later regions when punishing others. In sum, these results show that when an individual is personally responsible for delivering rewards and punishments to others, in-group bias is stronger for reward allocation than for punishments: the first neuroscience evidence of this dissociation.

For Peer Review

Introduction

In daily life, individuals in power often have to make decisions that have either positive or negative outcomes to others. Empathy is a necessary ability for these individuals to understand and sympathize with the effects these decisions have on the feelings and emotions of others, even if the recipients of the decisions belong to a different group. The alternative is a world ruled by cold-hearted decisions in which in-group bias thrives. Therefore it is crucial to get a better understanding of the neural underpinnings involved in delivering rewards and punishments directly to others, and investigate how these processes are modulated by personality and group membership. Empathy involves both affective and cognitive components (Bernhardt and Singer, 2012; Decety, 2011; Shamay-Tsoory, 2011) and the neural circuits underpinning these components can be modulated by various factors including individual differences and group membership (Chiao and Mathur, 2010; Eberhardt, 2005; Eres and Molenberghs, 2013; Hein et al., 2010; Hein and Singer, 2008; Ito and Bartholow, 2009; Kubota et al., 2012; Molenberghs, 2013). Yet, to date not much is known about how these feelings of empathy are influenced when participants are directly responsible for the actions that cause pleasure or harm to others. In addition, no study thus far has investigated how this process is influenced by individual dispositions and group membership.

Affective understanding of other's emotions is partially subserved by simulating the emotions we perceive in others onto similar brain regions as when we experience these emotions ourselves (Keysers and Gazzola, 2009; Molenberghs et al., 2012). The sensory components of the emotions are not necessarily simulated, but rather their affective responses (Singer *et al.*, 2004). For example, when experiencing pain, individuals' somatosensory cortex is activated (sensory stimulation) in addition to regions that represent the affective components

For Peer Review

of pain such as the anterior insula and dorsal anterior cingulate cortex (Lloyd *et al.*, 2004; Singer *et al.*, 2004). However, when individuals see others receiving physical pain the sensory components are not necessarily simulated through activation of the somatosensory cortex. Instead, only the affective components of the painful stimulation may be simulated, through activation of the anterior insula and dorsal anterior cingulate cortex (Jackson et al., 2005; Singer et al., 2004). The affective components of pain have been shown previously to be modulated by group membership. For example, when participants were shown pictures of faces being penetrated with a needle, they showed more affective pain sharing (more activation in the dorsal anterior cingulate cortex and anterior insula) when the faces where from members of their own race compared to members from a different race (Xu *et al.*, 2009).

Individuals do not only simulate negative painful emotions with others but also positive emotions such as rewards. When people receive awards themselves, they typically activate the striatum and medial orbitofrontal cortex (Fliessbach et al., 2007; Haruno and Kawato, 2006; Izuma et al., 2008; Knutson et al., 2001; Kringelbach, 2005; McClure et al., 2004). Similar areas are involved when giving donations to charities (Harbaugh et al., 2007; Izuma et al., 2010; Moll et al., 2006). These areas also become more activated depending on how close the individual feels to the group. For example, Telzer and colleagues (2010) conducted an fMRI study in which participants could earn money and either keep it for themselves or give it to their family. They found that participants who identified more with their family showed increased reward system activation when contributing money to their family. However, the question still remains unknown how these reward related neural systems are modulated when giving rewards to in-group versus out-group members.

For Peer Review

Affective sharing of emotions is merely one component which helps us to empathize with the emotions of others. To fully understand what another person is thinking or feeling, one must also reason about the mental states of others. This cognitive component of empathy typically involves brain regions that we associate with Theory of Mind (i.e., the ability to attribute mental states to others) such as the medial prefrontal cortex and right posterior temporal sulcus (Amodio and Frith, 2006; Van Overwalle, 2009). The neural correlates involved in cognitive empathy are also modulated by group membership (Eres and Molenberghs, 2013). For example, an fMRI study by Adams and colleagues (2009) showed that participants were better in decoding the mental state of members from their own culture. This intercultural advantage was associated with increased activation in the posterior superior temporal sulcus. In another fMRI study, Mathur and colleagues (2010) showed scenes of African-Americans and Caucasian-Americans, in which in-group members and out-group members experience emotional pain (e.g., people in the midst of a natural disaster). They found that in-group biases in empathy were correlated with activity in the medial prefrontal cortex in response to in-group relative to out-group pain scenes.

What has been missing in the literature however, are investigations into how group membership and individual dispositions influences neural responses when individuals are responsible for delivering the rewards and punishments to others directly. For example, when giving rewards to in-group members do we activate reward related areas more than when giving rewards to out-group members? Do we activate affective pain regions more when delivering punishments to in-group compared to out-group members? Do we think more about the mental states (more activation in Theory of Mind regions) of others when we hurt other people and how is this influenced by group membership and individual dispositions? These are important questions that need to be answered to get a better fundamental

For Peer Review

understanding of how complex problems such as in-group bias and racism develop (Molenberghs, 2013). To investigate this, we used fMRI and allowed participants to deliver rewards (money) or punishments (electroshocks) to in-group and out-group members in response to their answers on a difficult trivia task (Figure 1). It was predicted that when participants deliver rewards to others, the same areas typically involved in processing rewards, such as the striatum and medial orbitofrontal cortex, would become more active. In addition, we predicted that these regions will become more active when people deliver rewards to in-group members compared to out-group members. When participants deliver shocks to others more activation in regions typically associated with Theory of Mind such as the medial prefrontal cortex and posterior temporal sulcus (Amodio and Frith, 2006) was expected. More activation in affective pain areas such as the anterior insula and dorsal cingulate cortex was also predicted when punishing others.

Based on previous research discussed above (Mathur et al., 2010; Xu et al., 2009), it could be argued that participants would empathize more with the pain of in-group members. On the other hand, behavioural research has consistently shown that when we are responsible, in-group bias is more about preferential treatment of the in-group rather than hostility towards the out-group, especially in situations where there is no direct competition between the two groups (Brewer, 1999; Halevy *et al.*, 2008). Indeed as Brewer (1999) noted, in-group love and out-group hate are not necessarily reciprocally related. Although we are usually happy if we can benefit our own group, we do not necessarily find joy in harming the out-group. Therefore an alternative prediction would be that rewarding others would show an in-group bias (more activation in reward related areas for in-group versus out-group), but hurting out-group members would activate the regions involved in punishing others (affective pain and Theory of Mind areas) to the same degree as hurting in-group members.

For Peer Review

In addition we investigated how individual dispositions such as an ability to see things from another person's perspective as well as psychopathic traits would moderate the effect of delivering pain to others. Previous research has shown that participants who show high levels of antisocial behaviour, show less activation in regions typically associated with empathy, when watching others being hurt or in distress (Decety et al., 2009; Deeley et al., 2006; Lockwood et al., 2013; Marsh et al., 2013; Meffert et al., 2013). However to date it remains unclear how the neural responses involved in empathy are influenced by these differences in individual dispositions when one is responsible for delivering harm to others directly. Although people who score high on the psychopathy scale can be very charming and giving, they have a profound lack of empathizing with the pain they cause to others (Hare, 2011). It was therefore predicted that people who are better at perspective taking and have lower levels of psychopathy would show more activation in the areas associated with punishing others.

Methods

Participants. Forty-eight healthy University of Queensland (UQ) students (24 females, mean age = 22.5 years, SD = 4.9 years) completed the experiment. To minimize the chance that participants were familiar with the deception techniques, we excluded any participants who had completed a Psychology course. Participants were paid $30 for their time. All participants gave written informed consent. The study was approved by the Behavioural & Social Sciences Ethical Review Committee of the University of Queensland.

Team allocation. To make group membership salient, participants were randomly allocated to either the red (24 participants) or green (24 participants) UQ team and asked to wear a

For Peer Review

jumper representing their team colour. Participants were told that members of the other team colour were students from a neighbouring university (Queensland University of Technology; QUT). While wearing the coloured jumper, participants filled in a computerized questionnaire.

Group identification questionnaire. Group identification was assessed by presenting participants with two statements: "I identify myself as a UQ student" and "I identify myself as a QUT student". Participants had to indicate on a 7-point Likert scale (7 = totally agree and 1 = totally disagree) how much they agreed with each statement. Only participants who identified more with UQ than QUT were invited to take part in the second session (fMRI experiment) a couple of days later.

Perspective Taking (PT) questionnaire. PT is a 10-item subscale of the Questionnaire of Cognitive and Affective Empathy (QCAE; (Reniers *et al.*, 2011)). All items are rated on a 4-point Likert scale. Perspective taking refers to the extent to which individuals adopt others perspectives (e.g., 'I can usually appreciate another person's view point, even if I do not agree with it'). Total scores were calculated by adding the scores from the 10-items with lower scores indicating that people are poorer at perspective taking. The total score was then used as a regressor in the fMRI analysis.

Psychopathy questionnaire. The SRP-III is a 64-item questionnaire that measures Psychopathy using a 5-point Likert scale (Paulhus *et al.*, 2012). The SRP contains four subscales assessing Interpersonal Manipulation (IP), Callous Affect (CA), Erratic Lifestyle

For Peer Review

(ELS) and Anti-Social Behaviour (ASB). Each subscale contains 16-items. A total score for psychopathy was calculated by averaging the scores from the four subscales. Higher psychopathy scores indicate higher levels of psychopathy. This score was then used as a regressor in the fMRI analysis.

Functional MRI experiment.

Participants were screened using a MRI safety checklist. Participants were shown a brief demonstration of the experiment on a laptop. The demonstration ensured participants correctly understood how to perform the task. Participants were then taken to meet two confederates (a UQ and a QUT student). The two confederates were either both males (50% of the time) or both females. Each confederate wore either a red or green jumper representing the corresponding institution colour (i.e., the UQ student was wearing green 50% of the time). The experimenter introduced the confederates by their name, institution and jumper colour (e.g., 'This is Peter in the red jumper. Peter is from UQ'). Each confederate was sitting in front of a Dell desktop computer with electrodes attached to their hands. The electrodes were attached to a set of wires that extended into a control box. Participants had to press one of three buttons indicating the delivery of a shock, reward or nothing. It was made known to participants that they should deliver the electroshocks to the students via a hand held remote in the scanner, only if the other students answered incorrectly. Participants were also informed that they were responsible for giving the rewards to the students (0.5 AUD for each correct answer), but only if the students gave the correct answer. In reality the other students were confederates and they did not receive any electroshocks or monetary rewards. Participants were then taken to the MRI scanner. The experimental stimuli were presented on a black background with white, red and green coloured text (see Figure 1).

For Peer Review

The task consisted of six experimental conditions: reward in-group (RI), shock in-group (SI), neutral in-group (NI), reward out-group (RO), shock out-group (SO), and neutral out-group (NO). Each trial (Figure 1) started with a difficult trivia question (4s) presented on a black screen in white text. Below the trivia question, 'Y' (answer = yes) or 'N' (answer = no) was presented in order to represent the correct answer to the question (Figure 1). On a second slide, both the UQ and QUT students' answer (3s) was presented on a black screen with red and green text. Text colour represented the corresponding university institution colour (e.g., UQ represented by red text 50% of the time). Order of answers (UQ first versus QUT first) was counterbalanced. Possible answers ('Y' (yes), 'N' (no) and 'X' (no response)) were pseudorandomized (i.e., one third of the time per confederate). A red and green circle (3s) appeared in the middle of the screen at separate times (i.e., if red was presented left in the previous slide the red dot would appear first) signifying the institution colour and response phase. Participants then presented a response to students in relation to their answers. Each colour was presented first 50% of the time. The order of all trivia questions, correct answer (yes or no), student answer (yes, no or no answer) and circle colour (red or green) was pseudorandomized (i.e., equal amounts across the experiment). The entire task was conducted in 5 repeated fMRI runs. Each run consisted of 12 trials per condition thus there was a total of 60 trials per condition across the 5 runs.

fMRI Image Acquisition. A 3-Tesla Siemens MRI scanner with 32-channel head volume coil was used to obtain the data. Functional images were acquired using gradient-echo planar imaging (EPI) with the following parameters: repetition time (TR) 3s, echo time (TE) 30 ms, flip angle (FA) 90°, 64 × 64 voxels at 3 x 3 in-plane resolution. Whole brain images were acquired every three seconds and a total of 164 were acquired per run. The first four TR

For Peer Review

periods from each functional run (during which a fixation point was presented) were removed to allow for steady-state tissue magnetization. T1-weighted image covering the entire brain was also acquired after the last run and used for anatomical reference (TR = 1900, TE = 2.32 ms, FA = 9°, 192 cubic matrix, voxel size = 0.9 cubic mm, slice thickness = 0.9 mm).

fMRI analysis. All MRI data were analysed with SPM8 Software (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London) run through Matlab (Mathworks Inc, USA). EPI images were realigned to the first scan of each run in order to correct for any head movements. Additionally, realignment corrected for any systematic differences in the images between runs. During coregistration, the anatomical scan was turned to the mean functional scan. The anatomical image was then segmented to the MNI T1 template. All functional and structural images were normalised to the created template using the segment parameters (voxel size 3 x 3 x 3 and 1 x 1 x 1, respectively). This was done in order to correct for brain size variations across participants. Finally, images were smoothed using a Gaussian kernel of 6 mm.

During the first level analysis, a general linear model was created for all participants. For each participant, in each of the six conditions (RI, SI, NI, RO, SO, and NO), an event related design identified the regions with significant BOLD changes in each voxel compared to the baseline. The events were modelled by a canonical hemodynamic response function (with time derivative) time-locked to the onset of each action (shock, reward or nothing). These onsets occurred at the start of the slide representing the green or red circle (Figure 1). To remove any potentially confounding effects of reaction time (RT) and accuracy we only modelled correct trials and included RT as a parametric modulation in our fMRI design. In

For Peer Review

the second level analysis, contrast images for each condition minus baseline across all participants were included in the factorial design. Follow-up t-tests were calculated for each research hypothesis to determine if the differences in brain activation between conditions were significant. First, significant brain regions active for punishing and rewarding were identified by comparing each condition compared to the neutral condition. A cluster-level threshold with a familywise error rate (FWE) of $p < 0.05$ corrected for the whole brain was used to identify significant activation, with clusters defined by a voxel-level probability threshold of $p < 0.001$. Region of interest analyses were then used for all follow up tests. A voxel-level threshold with a family wise error rate (FWE) of $p < 0.05$ corrected for the size of the cluster (defined by a 3 mm box around the peak coordinates using the WFU PickAtlas; http://www.fmri.wfubmc.edu/cms/software) was used to define significant activation for all follow-up analysis. All follow-up analyses were independent from the main contrasts.

For Peer Review

Results

Group identification questionnaire. A paired samples t-test revealed that participants reported that they identified more with UQ (M = 6.73, SD = 0.49) compared to QUT (M = 1.33, SD = 0.69), t (47) = -38.0, p < .001.

fMRI results. As expected, no difference in accuracy (p =.59) or reaction time (p =.12) was found between the six (RI, SI, NI, RO, SO and NO) conditions but crucially different patterns in fMRI activation were found for rewarding and punishing others.

Rewarding others. Significant neural activation was found for the reward (RI, RO) minus neutral (NI, NO) contrast (Figure 2) in the left putamen (-27, 8, 1, Z = 5.30, extent 136, *p* =

.001), right putamen (27, 5, -5, Z = 4.24, extent 77, p = .01) and medial orbitofrontal cortex (mOFC: 0, 23, -5, Z = 5.12, extent 148, p < .001). The same three regions were more active (Figure 2) in the reward in-group (RI) compared to the reward out-group (RO) condition (left putamen: -24, 11, -2; Z = 2.46, p = .04; right putamen: 27, 5, -8, Z = 3.26, p = .0004; mOFC: -3, 26, -2, Z = 2.61, p = .03).

[Insert Figure 2 here]

Punishing others. Significant neural activation was found for the shock (SI, SO) minus neutral (NI, NO) contrast (Figure 3) in the medial prefrontal cortex extending into the dorsal anterior cingulate cortex (mPFC/dACC: -3, 53, 28, Z = 5.73, extent 383, p < .001), left orbitofrontal cortex extending into the left anterior insula (left OFC/AI: -48, 23, -8, Z = 6.07, extent 232, p < .001), right orbitofrontal cortex extending into the right anterior insula (right OFC/AI: 45, 29, -8, Z = 4.69, extent 104, p = .003) and right posterior superior temporal sulcus (right pSTS: 57, -25, -5, Z = 4.38, extent 101, p = .004). No significant difference was found in any of these four regions for the SI minus SO contrast. These fMRI results are consistent with the view that different networks are involved in rewarding and punishing others and that in-group bias is more about in-group favouritism rather than out-group harm.

To further explore how individual dispositions would influence the extent to which participants empathize with the punishment delivered to others, we used the participants' scores on the perspective taking (a lower score means poorer perspective taking) and psychopathy (a higher score means higher levels of psychopathy) questionnaire as regressors. A positive one-tailed Pearson correlation (r = .25; p = .04) was found between % signal change in these four regions combined and the perspective taking score (Figure 3B). Follow-up tests in SPM found a positive trend in all four regions (Figure 3B) between perspective

For Peer Review

taking scores and the shock minus neutral contrast (mPFC/dACC: -3, 56, 25, Z = 2.77, p = .02; left IFG/AI: -45, 23, -8, Z = 2.30, p = .06; right IFG/AI: 48, 26, -8, Z = 2.06, p = .08; right pSTS: 57, -22, -5, Z = 1.93, p = .1). A negative one-tailed Pearson correlation (r = -.31; p = .02) was found between % signal change in these four regions combined and the psychopathy score (Figure 3C). Follow-up tests in SPM found a positive correlation in all four regions between psychopathy scores and the shock minus neutral contrast (mPFC/dAA: -6, 53, 31, Z = 2.53, p = .03; Left IFG/AI: -45, 26, -5; Z = 2.57, p = .03; right IFG: 42, 32, -11, Z = 2.60, p = .03; right pSTS: 57, -25, -2; Z = 2.61; p = .03). These results are in line with the view that people with poor perspective taking skills and higher levels of psychopathy have problems with empathizing with the harm they cause to others.

[Insert Figure 3 here]

Discussion

We measured the neural responses involved in rewarding and punishing others and investigated how they are modulated by group membership and individual dispositions. The results showed more activation in the bilateral putamen and medial orbitofrontal cortex when participants were rewarding others. Activation in the striatum was centered on the putamen, rather than other striatum regions that have also been associated with reward processing such as the caudate nucleus and nucleus accumbens. Specifically, the putamen has been implicated in making the association between a certain stimulus and the right action to receive the reward, while the caudate nucleus and ventral striatum are specifically involved in comparing the predicted and actual award (Haruno and Kawato, 2006). Given that the reward in our study was always the same amount (0.5 AUD) and participants had to learn the correct button

For Peer Review

response for the confederates' correct answer, it is no surprise that activation was centered on the putamen. The medial orbitofrontal cortex activation fits well with previous research where it has been shown to be associated with evaluation processes and value-guided decision-making (Noonan *et al.*, 2012) and monitoring the reward value of different reinforcers (Kringelbach and Rolls, 2004). More activation in the same regions was found when participants rewarded in-group members compared to out-group members. This finding is consistent with previous research showing that greater identification with the family resulted in increased activation in reward related regions when giving money to a family member compared to keeping money for themself (Telzer *et al.*, 2010). However, the present data are the first to show that delivering rewards to others is influenced by group membership. It seems that even if two individuals are unknown to us, we would rather give a reward to an in-group compared to out-group member.

Neural response for punishing others directly revealed a different pattern of brain activation, which is the first time this has been shown in the same study. Here, more activation was found in regions typically associated with Theory of Mind such as the medial prefrontal cortex and right posterior temporal sulcus (Amodio and Frith, 2006). This may indicate that participants were thinking about the mental states of others when delivering the electroshocks to the confederates which is consistent with previous neuroimaging research on third-party punishment (Buckholtz *et al.*, 2008). In addition, increased activity was found in the dorsal anterior cingulate cortex and bilateral anterior insula. This suggests that although participants did not see the confederates in pain, they were probably imagining the pain that the electroshock was causing (Jackson *et al.*, 2006). More activation was also found in the left and right orbitofrontal cortex adjacent to the anterior insula activation. The lateral orbitofrontal cortex has previously been shown to activate when people see others in pain or

For Peer Review

imagine hurting others (Decety and Porges, 2011) and imaging hurting others in the same study led to reduced activation in medial orbitofrontal cortex. The distinction between medial orbitofrontal cortex activation for rewards and lateral orbitofrontal cortex activation for punishments in the current study is also consistent with previous research, which found that linking certain behaviour with a monetary reward led to increased medial orbitofrontal cortex activation, while linking behaviour with monetary losses led to increases in lateral orbitofrontal cortex activation (Kringelbach and Rolls, 2004; O'Doherty et al., 2001)

Interestingly, no effect of group membership was found when punishing others. The same regions were equally activated when punishing in-group members compared to out-group members. Previous research has shown that when we see others in pain, we empathize more with the pain of in-group members than out-group members (Avenanti et al., 2010; Azevedo et al., 2012; Cheon et al., 2011; Hein et al., 2010; Xu et al., 2009). In those studies, however, the participants themselves were not responsible for causing the pain. Indeed, the findings of the current study provide evidence to suggest that we empathise equally with the painful experience delivered to both the in-group and out-group when we are responsible for delivering the pain ourselves, and when there is no direct competition or strong animosity between the two groups (i.e., neighbouring universities). This fits well with a long-held view in social psychology that in-group bias begins with the preferential treatment of the in-group member rather than necessarily involving direct hostility towards the out-group member (Brewer, 1999; Halevy *et al.*, 2008) and that ‘‘aggravating conditions’’ are needed for group members to punish and inflict harmful consequences onto out-group members (Mummendey and Otten, 1998). To our knowledge, this is the first time this dissociation has been shown in neuroimaging research. These results are also in line with the commonly held view that it is socially acceptable to show increased happiness when an in-group member is being rewarded

For Peer Review

but socially unacceptable to not care about the pain of an out-group member, especially when one is responsible for the pain caused. On the other hand, in situations where there is strong competition between the two groups (e.g., rival sporting teams) individuals can get pleasure out of the failures of the opposing team (Cikara *et al.*, 2011). In addition, people can also get pleasure out of punishing others if the other person has violated a social norm (De Quervain *et al.*, 2004).

Furthermore, the dispositional measures showed that participants who were better at perspective taking and had lower levels of psychopathy showed more activation when punishing others in the same areas identified during the punishing minus neutral contrast. Correlation results between empathic dispositions and hemodynamic activations in the past have been mixed (Decety, 2011), which is probably due to a lack of power in many fMRI studies (Yarkoni, 2009). To overcome this problem we included a large sample (N=48) in our fMRI study. Previous research has shown that clinical populations with psychopathic traits compared to healthy controls show less activity in regions involved in empathy when watching video clips of people being hurt (Decety et al., 2009; Deeley et al., 2006; Lockwood et al., 2013; Marsh et al., 2013). Our results further extend this view by showing that in normal populations, participants who show lower levels of perspective taking and higher levels of psychopathy have reduced activation in Theory of Mind and affective pain regions when they are responsible for the pain inflicted to others. This is also the first study that investigated the influence of these individual dispositions, when the participants were responsible for causing the harm.

To conclude, it appears that different neural networks are involved when we are responsible for rewarding or punishing others. When we reward others we activate reward related areas in

For Peer Review

the brain and more so for in-group members than out-group members. On the other hand, when punishing others, we activate pain related areas and show increased activation in areas involved in affective and cognitive empathy. Activation in these areas was modulated by personality differences rather than group membership. Past neuroimaging work on intergroup discrimination has focused on passively observing or witnessing differences in outcomes for in-group and out-group; the present work is among the first to study active discrimination and the first to compare reward and punishment in the same study. These results suggest that in situations where we are directly responsible, in-group favouritism is more about in-group love than out-group hate. The real world reminds us that many groups can and have revelled in causing injury or death to others, e.g. by nationality or religions. However, reassuringly, it appears that everyday groups do not experience the same delight in harm-doing.

Acknowledgements

This work was supported by an ARC Early Career Research Award (DE130100120) awarded to PM, an ARC Discovery Grant (DP130100559) awarded to PM and JD and an ARC Discovery Grant Discovery Grant (DP1092490) awarded to WL, JRS, and KDV.

For Peer Review

References

Adams RB, Rule NO, Franklin RG, Wang E, Stevenson MT, Yoshikawa S, Nomura M, Sato W, Kveraga K, Ambady N. (2009): Cross-cultural Reading the Mind in the Eyes: An fMRI Investigation. Journal of Cognitive Neuroscience 22(1):97-108.

Amodio DM, Frith CD. (2006): Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews Neuroscience 7(4):268-277.

Avenanti A, Sirigu A, Aglioti SM. (2010): Racial Bias Reduces Empathic Sensorimotor Resonance with Other-Race Pain. Current Biology 20(11):1018-1022.

Azevedo RT, Macaluso E, Avenanti A, Santangelo V, Cazzato V, Aglioti SM. (2012): Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. Human Brain Mapping:n/a-n/a.

Bernhardt BC, Singer T. (2012): The neural basis of empathy. Annual Review of Neuroscience 35:1-23.

Brewer MB. (1999): The psychology of prejudice: Ingroup love and outgroup hate? Journal of social issues 55(3):429-444.

Buckholtz JW, Asplund CL, Dux PE, Zald DH, Gore JC, Jones OD, Marois R. (2008): The neural correlates of third-party punishment. Neuron 60(5):930-940.

Cheon BK, Im D-m, Harada T, Kim J-S, Mathur VA, Scimeca JM, Parrish TB, Park HW, Chiao JY. (2011): Cultural influences on neural basis of intergroup empathy. NeuroImage 57(2):642-650.

Chiao JY, Mathur VA. (2010): Intergroup empathy: how does race affect empathic neural responses? Current Biology 20(11):R478-R480.

Cikara M, Botvinick MM, Fiske ST. (2011): Us Versus Them Social Identity Shapes Neural Responses to Intergroup Competition and Harm. Psychological Science 22(3):306-313.

De Quervain DJ-F, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E. (2004): The neural basis of altruistic punishment. Science.

Decety J. (2011): Dissecting the neural mechanisms mediating empathy. Emotion Review 3(1):92-108.

Decety J, Michalska KJ, Akitsuki Y, Lahey BB. (2009): Atypical empathic responses in adolescents with aggressive conduct disorder: a functional MRI investigation. Biological Psychology 80(2):203.

Decety J, Porges EC. (2011): Imagining being the agent of actions that carry different moral consequences: an fMRI study. Neuropsychologia 49(11):2994-3001.

Deeley Q, Daly E, Surguladze S, Tunstall N, Mezey G, Beer D, Ambikapathy A, Robertson D, Giampietro V, Brammer MJ. (2006): Facial emotion processing in criminal psychopathy Preliminary functional magnetic resonance imaging study. The British Journal of Psychiatry 189(6):533-539.

Eberhardt JL. (2005): Imaging race. American Psychologist 60(2):181.

Eres R, Molenberghs P. (2013): The influence of group membership on the neural correlates involved in empathy. Frontiers in Human Neuroscience 7.

Fliessbach K, Weber B, Trautner P, Dohmen T, Sunde U, Elger CE, Falk A. (2007): Social comparison affects reward-related brain activity in the human ventral striatum. Science 318(5854):1305-1308.

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(4):405-411.

Harbaugh WT, Mayr U, Burghart DR. (2007): Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science 316(5831):1622-1625.

For Peer Review

Hare RD. 2011. Without conscience: The disturbing world of the psychopaths among us: Guilford Press.

Haruno M, Kawato M. (2006): Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. Journal of neurophysiology 95(2):948-959.

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(1):149-160.

Hein G, Singer T. (2008): I feel how you feel but not always: the empathic brain and its modulation. Current Opinion in Neurobiology 18(2):153-158.

Ito TA, Bartholow BD. (2009): The neural correlates of race. Trends in Cognitive Sciences 13(12):524-531.

Izuma K, Saito DN, Sadato N. (2008): Processing of social and monetary rewards in the human striatum. Neuron 58(2):284.

Izuma K, Saito DN, Sadato N. (2010): Processing of the incentive for social approval in the ventral striatum during charitable donation. Journal of Cognitive Neuroscience 22(4):621-631.

Jackson PL, Brunet E, Meltzoff AN, Decety J. (2006): Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. Neuropsychologia 44(5):752-761.

Jackson PL, Meltzoff AN, Decety J. (2005): How do we perceive the pain of others? A window into the neural processes involved in empathy. NeuroImage 24(3):771-779.

Keysers C, Gazzola V. (2009): Expanding the mirror: vicarious activity for actions, emotions, and sensations. Current Opinion in Neurobiology 19(6):666-671.

Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. (2001): Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport 12(17):3683-3687.

Kringelbach ML. (2005): The human orbitofrontal cortex: linking reward to hedonic experience. Nature Reviews Neuroscience 6(9):691-702.

Kringelbach ML, Rolls ET. (2004): The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Progress in neurobiology 72(5):341-372.

Kubota JT, Banaji MR, Phelps EA. (2012): The neuroscience of race. Nature neuroscience 15(7):940-948.

Lloyd D, Di Pellegrino G, Roberts N. (2004): Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? Cognitive, Affective, & Behavioral Neuroscience 4(2):270-278.

Lockwood PL, Sebastian CL, McCrory EJ, Hyde ZH, Gu X, De Brito SA, Viding E. (2013): Association of Callous Traits with Reduced Neural Response to Others' Pain in Children with Conduct Problems. Current Biology.

Marsh AA, Finger EC, Fowler KA, Adalio CJ, Jurkowitz IT, Schechter JC, Pine DS, Decety J, Blair R. (2013): Empathic responsiveness in amygdala and anterior cingulate cortex in youths with psychopathic traits. Journal of Child Psychology and Psychiatry.

Mathur VA, Harada T, Lipke T, Chiao JY. (2010): Neural basis of extraordinary empathy and altruistic motivation. NeuroImage 51(4):1468-1475.

McClure SM, Laibson DI, Loewenstein G, Cohen JD. (2004): Separate neural systems value immediate and delayed monetary rewards. Science 306(5695):503-507.

Meffert H, Gazzola V, den Boer JA, Bartels AA, Keysers C. (2013): Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. Brain 136(8):2550-2562.

For Peer Review

Molenberghs P. (2013): The neuroscience of in-group bias. Neuroscience & Biobehavioral Reviews 37:1530-1536.

Molenberghs P, Cunnington R, Mattingley JB. (2012): Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. Neuroscience and biobehavioral reviews 36(1):341-9.

Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J. (2006): Human fronto–mesolimbic networks guide decisions about charitable donation. Proceedings of the National Academy of Sciences 103(42):15623-15628.

Mummendey A, Otten S. (1998): Positive–negative asymmetry in social discrimination. European review of social psychology 9(1):107-143.

Noonan M, Kolling N, Walton M, Rushworth M. (2012): Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement. European Journal of Neuroscience 35(7):997-1010.

O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. (2001): Abstract reward and punishment representations in the human orbitofrontal cortex. Nature neuroscience 4(1):95-102.

Paulhus D, Hemphill J, Hare R. (2012): Manual for the self-report psychopathy scale. Toronto: Multi-health systems.

Reniers RL, Corcoran R, Drake R, Shryane NM, Völlm BA. (2011): The QCAE: A questionnaire of cognitive and affective empathy. Journal of personality assessment 93(1):84-95.

Shamay-Tsoory SG. (2011): The Neural Bases for Empathy. The Neuroscientist 17(1):18-24.

Singer T, Seymour B, O'Doherty J, Kaube H, Dolan RJ, Frith CD. (2004): Empathy for pain involves the affective but not sensory components of pain. Science 303(5661):1157-1162.

Telzer EH, Masten CL, Berkman ET, Lieberman MD, Fuligni AJ. (2010): Gaining while giving: An fMRI study of the rewards of family assistance among White and Latino youth. Social neuroscience 5(5-6):508-518.

Van Overwalle F. (2009): Social Cognition and the Brain: A Meta-Analysis. Human Brain Mapping 30(3):829-858.

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(26):8525-8529.

Yarkoni T. (2009): Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al.(2009). Perspectives on Psychological Science 4(3):294-298.

For Peer Review

Figure 1. Schematic representation of the trivia task used during the fMRI experiment for a green UQ participant. A 'no' question (4s), followed by a 'yes' QUT student response and a 'no' UQ student response (3s). Red circle (3s) representing QUT followed by a green circle (3s) representing UQ. Participants gave either a reward (for a correct answer), a shock (for an incorrect answer) or nothing (for no answer) during each circle presentation. In this case, the QUT confederate would get a shock and the UQ confederate would get a reward.

For Peer Review

Figure 2. Significant left putamen, right putamen and medial oribitofrontal cortex activation for the reward (RI, RO) minus neutral (NI, NO) and reward in-group (RI) minus reward out-group (RO) contrasts. Activations are displayed on a ch2better.nii.gz template using MRIcron.

For Peer Review

Figure 3. A. Significant activation for the shock (SI, SO) minus neutral (NI, NO) contrast. B. Positive correlation (in red) between perspective taking score and shock minus neutral contrast. C. Negative correlation (in blue) between psychopathy score and shock minus neutral contrast. mPFC / dACC = medial prefrontal cortex extending into dorsal anterior cingulate cortex. left OFC / AI = left orbitofrontal cortex extending into left anterior insula. right OFC / AI = right orbitofrontal cortex extending into right anterior insula. right pSTS = right posterior superior temporal sulcus. Activations are displayed on a ch2better.nii.gz template using MRIcron.

For Peer Review

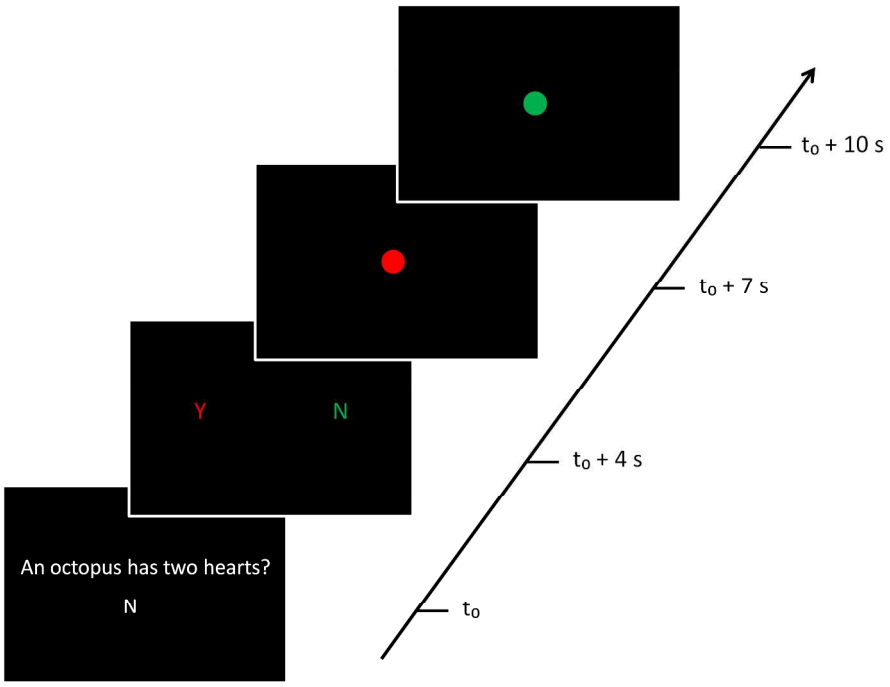

Figure 1. Schematic representation of the trivia task used during the fMRI experiment for a green UQ participant. A 'no' question (4s), followed by a 'yes' QUT student response and a 'no' UQ student response (3s). Red circle (3s) representing QUT followed by a green circle (3s) representing UQ. Participants gave either a reward (for a correct answer), a shock (for an incorrect answer) or nothing (for no answer) during each circle presentation. In this case, the QUT confederate would get a shock and the UQ confederate would get a reward.
254x190mm (300 x 300 DPI)

iew

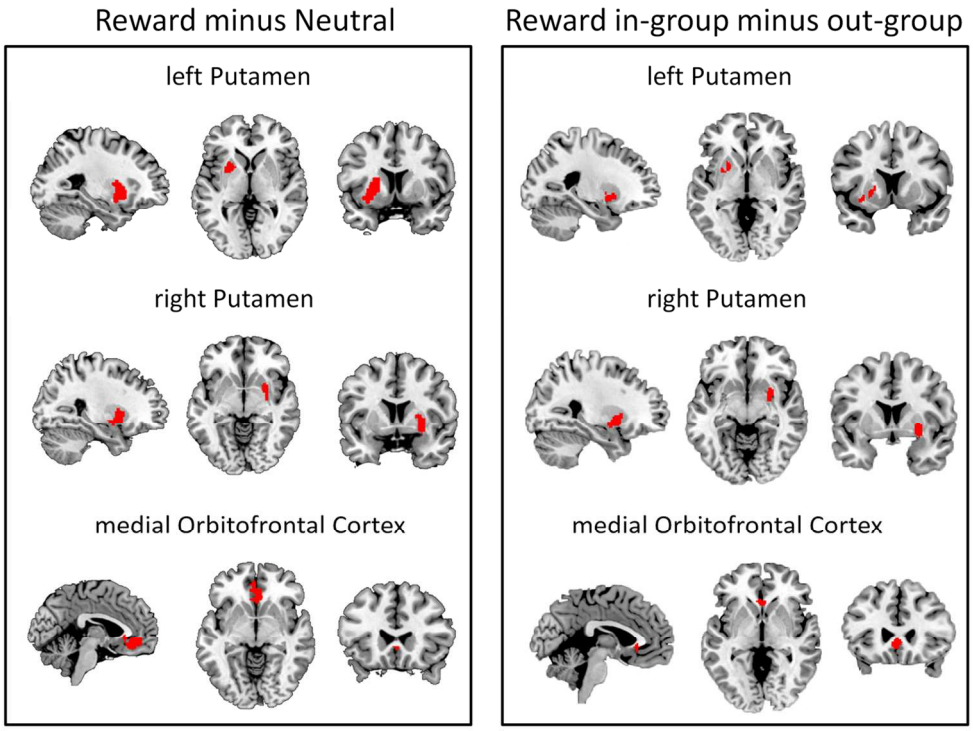

Figure 2. Significant left putamen, right putamen and medial oribitofrontal cortex activation for the reward (RI, RO) minus neutral (NI, NO) and reward in-group (RI) minus reward out-group (RO) contrasts. Activations are displayed on a ch2better.nii.gz template using MRIcron.
254x190mm (150 x 150 DPI)

eview

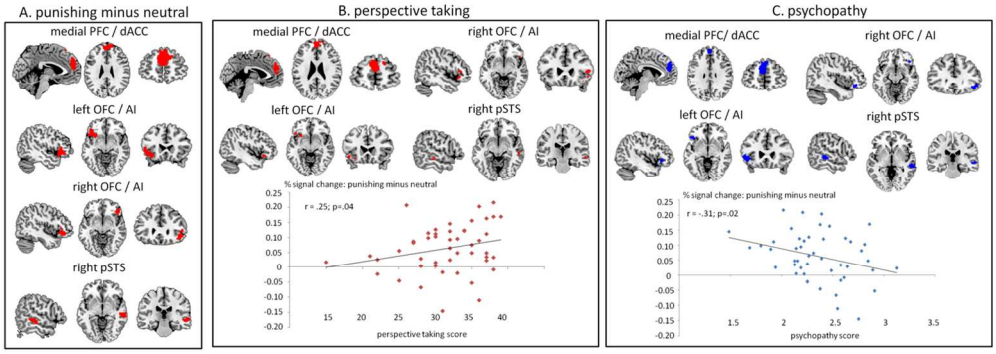

Figure 3. A. Significant activation for the shock (SI, SO) minus neutral (NI, NO) contrast. B. Positive correlation (in red) between perspective taking score and shock minus neutral contrast. C. Negative correlation (in blue) between psychopathy score and shock minus neutral contrast. mPFC / dACC = medial prefrontal cortex extending into dorsal anterior cingulate cortex. left OFC / AI = left orbitofrontal cortex extending into left anterior insula. right OFC / AI = right orbitofrontal cortex extending into right anterior insula. right pSTS = right posterior superior temporal sulcus. Activations are displayed on a ch2better.nii.gz template using MRIcron.
249x88mm (150 x 150 DPI)

eer Review