doi:10.1093/scan/nsv027 SCAN (2015) 10, 1397–1404

The neural correlates of justified and unjustified killing: an fMRI study

Pascal Molenberghs,[1]* Claudette Ogilvie,[2]* Winnifred R. Louis,[2] Jean Decety,[3,4] Jessica Bagnall,[2] and Paul G. Bain[5]
[1]School of Psychological Sciences, Monash University, Melbourne, Australia, [2]School of Psychology, The University of Queensland, St Lucia, Australia, [3]Department of Psychology, and [4]Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, IL, USA, and [5]School of Psychology and Counselling, Queensland University of Technology, Brisbane, Australia

Despite moral prohibitions on hurting other humans, some social contexts allow for harmful actions such as killing of others. One example is warfare, where killing enemy soldiers is seen as morally justified. Yet, the neural underpinnings distinguishing between justified and unjustified killing are largely unknown. To improve understanding of the neural processes involved in justified and unjustified killing, participants had to imagine being the perpetrator whilst watching 'first-person perspective' animated videos where they shot enemy soldiers ('justified violence') and innocent civilians ('unjustified violence'). When participants imagined themselves shooting civilians compared with soldiers, greater activation was found in the lateral orbitofrontal cortex (OFC). Regression analysis revealed that the more guilt participants felt about shooting civilians, the greater the response in the lateral OFC. Effective connectivity analyses further revealed an increased coupling between lateral OFC and the temporoparietal junction (TPJ) when shooting civilians. The results show that the neural mechanisms typically implicated with harming others, such as the OFC, become less active when the violence against a particular group is seen as justified. This study therefore provides unique insight into how normal individuals can become aggressors in specific situations.

Keywords: morality; intentional harm; violence; conflict; orbitofrontal cortex

INTRODUCTION

Morality entails notions of rights, fairness and justice, as well as rules regarding how people should treat one another (Killen and Rutland, 2011; Decety and Cowell, 2014; Turiel, 2014). Among the many types of normative judgments, both common sense and academic research recognise the distinction between conventional and moral transgressions (Nucci, 1981). Thus all typically developing individuals are aware that inflicting harm on others is morally wrong (Cooney, 2009). However, harm-doing persists, and in some cases aggression against a person or a particular group can be seen as justified, for example when people are acting in self-defence or in competitive social contexts or conflicts like warfare. How such contexts, where violence seems justified, impact on the psychological and neural processes remain largely unknown. Therefore, this study aims to identify the neural underpinnings that distinguish between justified and non-justified violence.

Converging evidence from multiple sources, using a variety of methods, point to specific neural mechanisms underlying moral cognition. However, no region can be singled out as a uniquely moral centre, and all of these regions are implicated in other functions as well (for reviews see Moll *et al.*, 2005; Raine and Yang, 2006; Young and Dungan, 2012; Pascual *et al.*, 2013). While morality is often assessed with complex moral dilemmas, such as the 'trolley problem' (Kamm, 1989), here we focus instead on the more implicit affective and automatic components which are the antecedents of complex moral reasoning (Haidt, 2001). Specifically, this study examines the neural mechanisms involved when imagining directly harming others and how these are influenced by social contexts.

Harming or killing another person typically involves inflicting pain. Previous neuroimaging research has identified the brain regions involved in perceiving others being harmed or in physical pain (Singer *et al.*, 2004; Decety *et al.*, 2012; Molenberghs *et al.*, 2014a). Areas often implicated in these studies include the anterior cingulate cortex (ACC), anterior insula, medial prefrontal cortex (mPFC) and orbitofrontal cortex (OFC). The ACC and insula are reliably associated with the affective components of pain (Singer *et al.*, 2004; Jackson *et al.*, 2005; Lamm *et al.*, 2011), while the mPFC is typically associated with thinking about the mental states of others or so-called Theory of Mind (Amodio and Frith, 2006; van Overwalle, 2009; Eres and Molenberghs, 2013; Schurz *et al.*, 2014). This overlap between first-hand experience of pain and perceiving pain in others is explained by the fact that the two experiences are salient and hence trigger multimodal cognitive processes involved in detecting and orienting attention towards salient events (Iannetti *et al.*, 2013; Seeley *et al.*, 2007). Importantly, the neural networks implicated in perceiving others in pain are modulated by interpersonal relationships, implicit attitudes and group preferences (Singer *et al.*, 2006; Hein and Singer, 2008; Decety *et al.*, 2009; Xu *et al.*, 2009; Cheng *et al.* 2010; Cikara *et al.*, 2011; Eres and Molenberghs, 2013; Fox *et al.*, 2013; Molenberghs, 2013; Porges and Decety, 2013).

The role of the OFC in affective responses is very much shaped by context. This region, for example, is activated by watching intentional harm but not by watching accidental harm inflicted onto others (Decety and Cacioppo, 2012; Decety *et al.*, 2012). Activation in the OFC is not essential for affective responses *per se* but is critical when meaning has to be given to a certain affective stimulus (Roy *et al.*, 2012). As such, activation in this area is shaped by conceptual information which drives appropriate affective physiological and behavioural responses (Roy *et al.*, 2012). Functional neuroimaging studies have shown that this region plays a central role in contextually dependent moral judgment (Zahn *et al.*, 2011).

Received 10 November 2014; Revised 17 February 2015; Accepted 4 March 2015
Advance Access publication 9 March 2015

*These authors contributed equally to this work.

? The authors thank Hanne M. Watkins and Nick Haslam for constructive comments on an earlier version of the manuscript and Zoie Nott for help with data collection. The work was supported by an ARC Early Career Research Award (DE130100120) and Heart Foundation Future Leader Fellowship (1000458) awarded to P.M., an ARC Discovery Grant (DP130100559) awarded to PM and JD and an ARC Discovery Grant (DP1092490) awarded to WL.

Correspondence should be addressed to Pascal Molenberghs, School of Psychological Sciences, Monash University, Building 17, Clayton Campus, Wellington Road, VIC 3800-4072, Australia. E-mail: pascal.molenberghs@monash.edu.

Therefore, the OFC has an important role in moral cognition (for reviews see: Moll *et al.*, 2005; Blair *et al.*, 2006; Decety *et al.*, 2011; Sobhani and Bechara, 2011; Pascual *et al.*, 2013). Meta-analysis on the OFC have routinely shown that the lateral region of the OFC is related to the evaluation of punishers and may lead to a change in behaviour, whereas the medial region is typically related to monitoring the reward value of reinforcers (Kringelbach and Rolls, 2004; Berridge and Kringelbach, 2013). Severe damage to the OFC can lead to an increase in immoral behaviour (Eslinger and Damasio, 1985; Saver and Damasio, 1991; Sobhani and Bechara, 2011), and individuals scoring high on psychopathy often have anatomical and functional dysfunctions in this area (Best *et al.*, 2002; Antonucci *et al.*, 2006; Anderson and Kiehl, 2012; Decety *et al.*, 2013; Molenberghs *et al.*, 2014b).

Very little is known about social situations in which people are directly responsible for harming others. Previous work has reported differences in activation patterns as well as functional connectivity modulated by whether an observed action is harmful (Decety and Porges, 2011), intentional (Akitsuki and Decety, 2009), and whether perceivers enjoy the violence or not (Porges and Decety, 2013). A recent fMRI study investigated the neural mechanisms involved when people are directly responsible for rewarding or harming ingroup (i.e. students from the same university) and outgroup (i.e. students from a neighbouring university) members (Molenberghs *et al.*, 2014b). Participants gave rewards (i.e. money) or punishments (i.e. electroshocks) to ingroup and outgroup members performing a trivia task while undergoing fMRI. The results showed that when participants rewarded others, greater activation was found in regions typically associated with receiving rewards such as the striatum and medial OFC. These areas became more active when people were rewarding ingroup *vs* outgroup members. In contrast, punishing others led to increased activation in regions typically associated with Theory of Mind including the mPFC and posterior superior temporal sulcus, as well as regions typically associated with moral sensitivity such as the lateral OFC (Molenberghs *et al.*, 2014b). Importantly, these areas were equally active when harming ingroup *vs* outgroup members. This suggests that, at least in situations where there is no strong animosity between groups (i.e. students from neighbouring universities), ingroup bias is more about favouring the ingroup rather than harming the outgroup (Brewer, 1999).

However, group membership extends beyond just ingroup–outgroup comparisons. For example, during war the group membership of the victim (e.g. civilian *vs* soldier) should play an important role when deciding to harm the person or not. If the violence against someone is seen as justified (i.e. killing an enemy soldier when under attack) *vs* unjustified (i.e. killing an innocent civilian), we should feel less guilt and this should lead to a difference in neuronal activation in areas typically associated with moral sensitivity. Moral sensitivity is defined here as the quick detection of a moral situation, which typically happens prior to complex moral reasoning (Haidt, 2001; Robertson *et al.*, 2007; Molenberghs *et al.*, 2014a). This process is specifically relevant in our study where people are exposed to moral and immoral situations in a highly dynamic environment. Moral sensitivity as such is the first stage of ethical decision, which according to the social intuitionist model (Haidt, 2001), is associated with an instant feeling of approval or disapproval when we witness a morally laden action.

Previous fMRI research has shown that an increase in moral sensitivity when watching others being harmed is associated with increased activation in lateral OFC (Decety *et al.*, 2012; Molenberghs *et al.*, 2014a). More complex moral reasoning, on the other hand, is typically associated with Theory of Mind regions such as the mPFC, posterior superior temporal sulcus (STS) and adjacent temporoparietal junction (TPJ) (Young and Dungan, 2012). But what about situations where the individual is directly responsible for causing harm? A sense of agency is a major dimension of the moral experience (Berthoz *et al.*, 2006; Moll *et al.*, 2007; Decety and Porges, 2011). In particular, actively causing harm may lead to more moral emotions such as guilt, which is the feeling we experience when we feel personally responsible for the misfortune caused to others (Moll *et al.*, 2007; Kédia *et al.*, 2008). Understanding the link between harm-doing and guilt is of critical importance because it provides unique insights into why ordinary people are able to commit harmful actions in specific situations.

The aim of this study was to identify the different neural mechanisms involved in being responsible for harming others in justified and unjustified situations. To investigate this, we conducted an fMRI study in which participants imagined being the perpetrator while watching animated video clips (see Experimental Procedures for details) from a first-person perspective of a person shooting enemy soldiers, civilians or nobody (control). Note that in our study, people were not able to decide if they wanted to shoot or not shoot. As such we operationalise 'agents' in this study not as 'having the ability the make a decision to shoot or not', as for example in the studies by Correll *et al.* (2002, 2006), but rather as 'a person who is actively imaging to perform an action'.

Imagining being the agent of prosocial or antisocial actions has been previously used successfully as a paradigm in an MRI environment (Decety and Porges, 2011). It was predicted that participants would feel less guilt when imagining shooting soldiers compared with civilians. We also predicted, when imagining shooting soldiers (when compared with imagining shooting civilians), participants would show less activation in areas typically associated with directly harming others and moral sensitivity, such as the lateral OFC (Molenberghs *et al.*, 2014a, b), given that these actions are typically seen as more justified.

METHODS

Participants

Forty-eight participants (24 males) participated in the fMRI experiment (age range: 18–51 years; $M = 22.5$, s.d. $= 5.3$ years). All participants had normal or corrected-to-normal vision and were tested for MRI safety. All participants signed written informed consent upon arrival and were reimbursed $30 AUD on completion of their participation. The study was approved by the Behavioural & Social Sciences Ethical Review Committee of the University of Queensland.

Materials and procedure

Participants were informed that they would be watching first-person perspective video game clips in which a soldier, civilian or nobody would get shot. Having participants imagine being the actor while watching pre-recorded scenes allowed us to better control differences in duration and visual characteristics between conditions, compared with previous fMRI studies in which participants played the games themselves (King *et al.*, 2006; Mathiak and Weber, 2006; Mathiak *et al.*, 2011; Klasen *et al.*, 2012). The video clips were matched between the three conditions (Soldiers, Civilians and Control) on several visual characteristics such as the amount of movement, brightness and clarity (see Pilot Experiment in Supplementary Material for details). Participants were instructed to imagine themselves as the person performing the action so that their mental simulation of the situation could be used to elicit neural activity that closely parallels the actual situation (Decety and Porges, 2011). The video clips in the three conditions were presented in blocks in which three videos from the same condition were presented sequentially (Figure 1), followed by the question, 'Who did you shoot?' Participants responded by using a

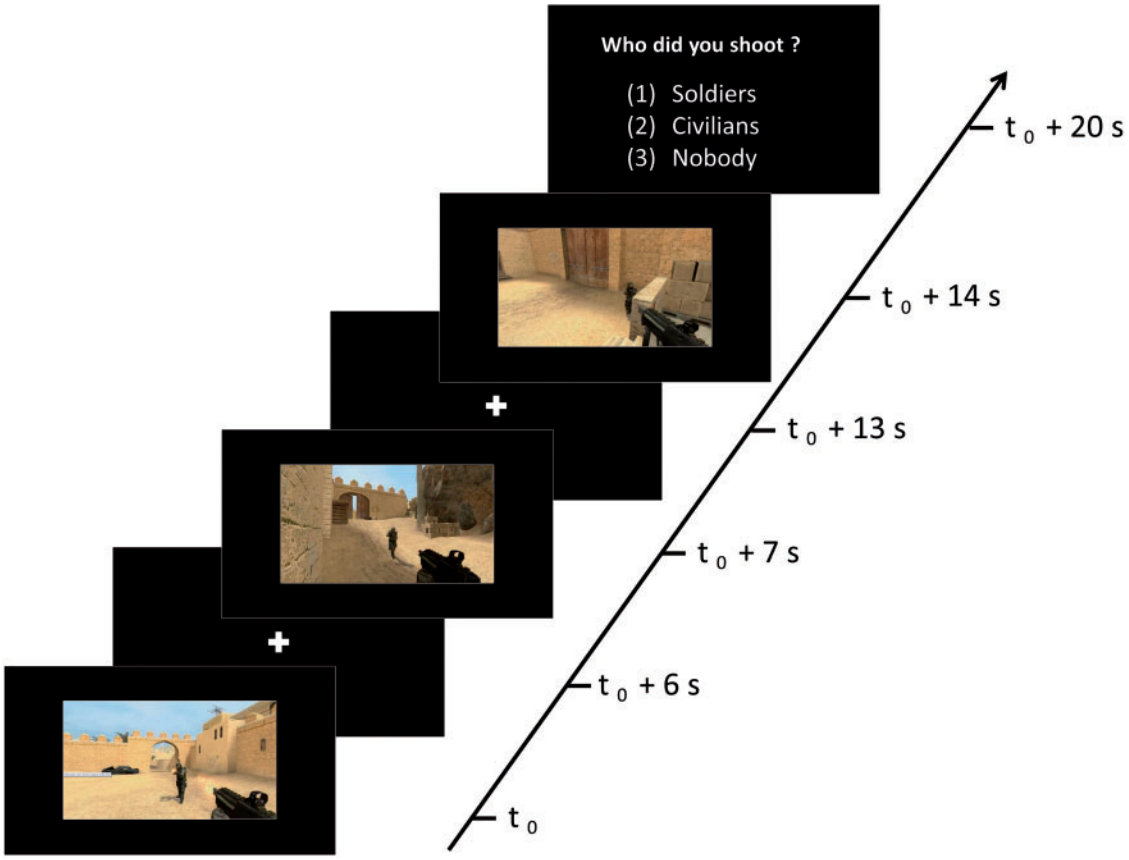

Fig. 1 Schematic representation of an example block from the 'Soldiers' condition. Each block included three video clips of 6 s from the same condition and ended with a 5 s response window.

three button response pad corresponding to the answers, (i) Soldiers, (ii) Civilians and (iii) Nobody. These responses were counterbalanced in their order of appearance across participants. Participants were given a practice run of the experiment on a laptop using E-prime2 (http://www.pstnet.com/eprime.cfm), to ensure they understood what was involved in the task.

The fMRI experiment consisted of five repeated functional runs (5 min each) with a structural scan (5 min), which was conducted between the third and fourth run. At the beginning of each run, participants were presented with the instructions as a reminder. A white fixation dot was then presented on a black screen for 5 s. Each functional run had 12 blocks. During each block, three videos of one condition were presented on a random basis from a list of 18 videos for each of the three conditions. Presenting three videos in sequence from the same condition was used to analyse the fMRI data in a block design, which was chosen as it offers superior statistical power over an event-related design (Aguirre and D'Esposito, 1998). The first two videos in one block were followed by a 1 s delay. The third video was followed by a 5 s response ('Who did you shoot?') window during which reaction and accuracy was recorded. Each block lasted 25 s (Figure 1) and each condition was presented four times, making up the 12 blocks per run (presented in random sequence), ending with a 10 s delay. This was repeated for each of the five 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 with the gradient echo planar imaging (EPI) with the following parameters: repetition time (TR) of 2.5 s, echo time (TE) of 36 ms, flip angle (FA) of 90°. Thirty-six transversal slices with 64×64 voxels at $3\,mm^2$ in-plane resolution and a 10% gap in between the slices covered the whole brain. Whole brain images were generated every 2.5 s and 166 images were acquired during each functional run. The first two images from each functional run were removed to allow for steady-state tissue magnetisation. A three-dimensional high resolution T1-weighted whole brain structural image was acquired after the third run for anatomical reference (TR = 1900, TE = 2.32 ms, FA = 9°, 192 cube matrix, voxel size = $0.9\,mm^3$, slice thickness = 0.9 mm).

fMRI analyses

SPM8 software (http://www.fil.ion.ucl.ac.uk/spm/) run through Matlab (http://www.mathworks.com.au/products/matlab/) was used to analyse the data. To counter any head-movements all the EPI images were realigned to the first scan of each run and a mean image was created. The anatomical image was then co-registered to this mean functional image. To correct for variation in brain size and anatomy between participants, each structural scan was normalised to the MNI T1 standard template (Montreal Neuropsychological Institute) with a voxel size of $1 \times 1 \times 1$ mm using the segmentation procedure. The same segmentation parameters were then also used to normalise all the EPI images to the T1 template with a voxel size of $3 \times 3 \times 3$ mm. This process mathematically transformed each participant's brain image to match the template so that any chosen brain region should refer to the same region across all participants. Before further analysis, all images were smoothed with an isotropic Gaussian kernel of 9 mm.

As part of the first level of analysis, a general linear model was created for each participant. For each participant in each of the three conditions (i.e. Soldiers, Civilians and Control), regions with significant blood oxygen level dependent changes in each voxel were identified using a block design with a duration of 20 s (which corresponds to the presentation of the three videos without the 5 s response) and onsets aligned to the start of each condition.

In the second level of analysis, contrast images for each condition across all participants were included in a factorial design. First, an ANOVA was conducted to identify clusters that were differentially activated between the three conditions (cluster-wise familywise error rate (FWE) of $P < 0.05$; corrected for multiple comparisons for the whole brain with clusters thresholded at $P < 0.001$). These significant regions combined were then used as a mask for subsequent pairwise analyses between the conditions.

To specifically examine the differences between justified and unjustified violence, the contrast Civilians minus Soldiers and its reverse contrast Soldiers minus Civilians were created. These two contrasts isolated the unique activation associated with justified and unjustified violence. The contrasts Civilians minus Control and Soldiers minus Control were also created to compare the shooting conditions to a baseline control condition. Results from the ANOVA and the latter two contrasts are presented in Supplementary Material (Supplementary Figure S1 and Table S1) and interpreted only when further clarification is required, as they were not central to the research question. All pairwise analyses were corrected for multiple comparisons for the size of the mask, using a voxel-level FWE of $P < 0.05$, as a measure of significance.

Guilt

To determine whether our manipulation was effective, we asked participants right after the fMRI experiment to indicate on a 7-point scale (1 = strongly agree to 7 = strongly disagree) how guilty they felt about shooting the soldiers and civilians, respectively: 'I felt guilty about shooting the soldiers/civilians'. Items on the two items were reverse scored so that higher levels indicated 'more guilt'.

Psychophysiological interaction analysis

Two effective connectivity analyses using psychophysiological interaction (PPI) were performed to estimate functional coupling between two sources (seed in left and right OFC) and the rest of the brain for the Civilians minus Soldiers contrast. PPI analysis assesses the hypothesis that activity in one brain region can be explained by an interaction between a cognitive process and activity in another part of the brain. The selection of left and right OFC as the PPI source region was based on its significant involvement in the Civilians minus Soldiers contrast.

Activity within left and right OFC was used as the physiological regressor in the PPI analysis. The individual time series for the left and right OFC was obtained by extracting the first principal component from all raw voxel time series in a sphere (6 mm radius) centred around the peak coordinate identified in the Civilians minus Soldiers contrast. A similar size sphere was used in the PPI analysis for left and right OFC to make the results comparable between the two regions. These time series were mean-corrected and high-pass filtered to remove low-frequency signal drifts. Civilians minus Soldiers was the psychological regressor. The psychological variable used was a vector coding for the specific task (1 for Civilians, −1 for Soldiers) convolved with the haemodynamic response function. A third regressor in the analysis represented the interaction between the first and second regressors. The physiological factor was multiplied with the psychological factor to constitute the interaction term.

PPI analyses were carried out for each subject involving the creation of a design matrix with the interaction term, the psychological factor, and the physiological factor as regressors. Subject-specific contrast images were then entered into two random-effects group analyses. PPI analyses was then conducted to identify if any brain areas showed a significant increase in functional coupling with left and right OFC during Civilians relative to Soldiers. Significant activity for the PPI analyses was defined by a cluster-wise FWE of $P < 0.05$; corrected for multiple comparisons for the whole brain with clusters thresholded at $P < 0.001$.

RESULTS

Behavioural results ('Who did you shoot?') during the fMRI experiment

Reaction time

Mauchly's test of sphericity revealed that the assumption of sphericity had been violated, $\chi^2(2) = 16.93$, $P < 0.001$. Therefore, the degrees of freedom were corrected using Greenhouse-Geisser estimates. A one-way repeated measures ANOVA revealed no significant differences in RT between the Soldiers ($M = 1229$ ms, s.d. = 497 ms), Civilians ($M = 1240$ ms, s.d. = 503 ms) and Control ($M = 1324$ ms, s.d. = 519 ms) conditions, $F(1.53, 71.87) = 3.41$, $P = 0.051$.

Accuracy

Mauchly's test indicated that the assumption of sphericity had not been violated, $\chi^2(2) = 3.33$, $P = 0.19$. A one-way repeated measures ANOVA revealed no significant differences in accuracy between the Soldiers ($M = 98.44\%$, s.d. = 3.12%), Civilians ($M = 97.60\%$, s.d. = 3.99%) and Control ($M = 98.13\%$, s.d. = 3.52%) condition, $F(2, 94) = 0.92$, $P = 0.401$.

Guilt

Consistent with expectations, a paired-samples *t*-test revealed that participants felt more guilt when shooting civilians ($M = 5.77$, s.d. = 1.78) *vs* soldiers ($M = 3.81$, s.d. = 1.91), $t(47) = 8.16$, $P < 0.001$.

fMRI results

Civilians minus Soldiers

When imagining shooting civilians relative to shooting soldiers, significantly more activation was found in bilateral lateral OFC and left fusiform gyrus (Table 1, Figure 2A). To further explore, if the difference in activation in bilateral OFC was influenced by feelings of relative guilt between Civilians *vs* Soldiers, the mean % signal change for all voxels in this bilateral region was extracted for these two conditions.

Table 1 Cluster size and associated peak values for the significant brain regions in the Civilians minus Soldiers and Soldiers minus Civilians contrast

	Cluster size	Peak *P*-value	Peak *Z*-value	MNI Coordinates		
				x	*y*	*z*
Civilians minus Soldiers						
Left lateral OFC	252	0.004	4.32	−33	20	−11
Right lateral OFC	43	0.008	4.17	36	17	−20
Left fusiform gyrus	421	0.032	3.78	−36	−46	−14
Soldiers Minus Civilians						
Precuneus	463	0.001	4.57	9	−49	55
Lingual gyrus	55	0.024	3.87	6	−85	−5

See also Supplementary Table S1 for the main effect of conditions analysis, Civilians minus Control contrast and Soldiers minus Control contrast

A bivariate Pearson correlation revealed that the more guilty participants felt about shooting civilians *vs* soldiers, the higher the % signal change difference was between the two conditions, $r(46) = 0.30$, $P = 0.04$ (Figure 3).

Soldiers minus Civilians

When imagining shooting soldiers relative to shooting civilians, significantly more activation was found in the precuneus and lingual gyrus (Table 1, Figure 2B).

PPI analysis

Effective connectivity analyses showed significant increased coupling between the left OFC and left and right TPJ (left TPJ: −51, −58, 40, Z = 4.13, extent = 257, *P* FWE = 0.002; right TPJ: 45, −55, 31, Z = 4.02, extent = 121, *P* FWE = 0.039; Figure 4) for the Civilians minus Soldiers contrast. The right OFC did not show increased connectivity with any regions for this contrast at the whole brain level. However when using a region of interest approach (6 mm sphere around the left and right TPJ peaks identified in the left OFC PPI analysis), a similar significant increased coupling effect was found (left TPJ: −48, −55, 43, Z = 2.47, *P* FWE = 0.039; right TPJ: 45, −52, 34, Z = 2.89, *P* FWE = 0.043).

DISCUSSION

As expected, participants experienced less guilt when imagining shooting soldiers compared with imagining shooting civilians. Interestingly, mentally simulating the killing of civilians led to increased activation in the lateral OFC, while killing soldiers did not (Figure 2A, Table 1; also see Supplementary Table S1 and Figure S1). In addition, the more guilt participants felt about shooting civilians *vs* soldiers, the greater the activation in the lateral OFC (Figure 3). These results show that being responsible for justified or unjustified violence against others leads to differential feelings of guilt, and that activation in the lateral OFC is directly related with this experience.

Agency is an important factor in the subjective experience of responsibility and previous neuroimaging research has shown that the lateral OFC is an important area when agency is involved in aversive, morally sensitive situations (Moll *et al.*, 2007; Decety and Porges, 2011). Meta-analysis on fMRI studies have consistently associated the lateral OFC with the evaluation of punishers which may lead to a change in ongoing behaviour, whereas the medial OFC is typically related to monitoring the reward value of reinforcers (Kringelbach and Rolls, 2004; Berridge and Kringelbach, 2013). This suggests that the OFC plays an important role in linking certain behaviours and stimuli with either positive (medial OFC) or negative (lateral OFC) value.

A similar distinction in OFC was recently found in a recent fMRI study, in which participants either had to reward other people, which led to increased activation in medial OFC, or punish other people, which led to increased activation in lateral OFC (Molenberghs *et al.*, 2014b). In this study, no overall difference in lateral OFC activation was found when harming members of participants' own university *vs* a

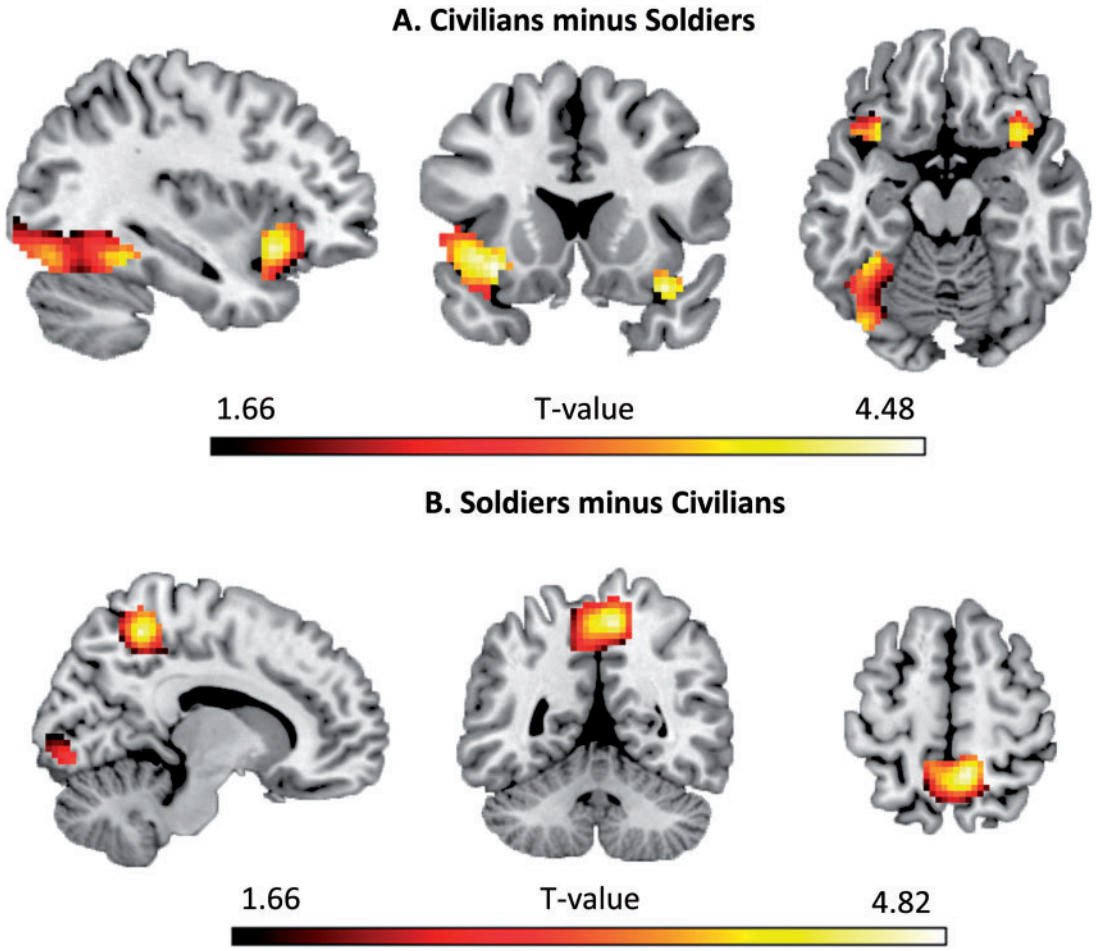

Fig. 2 Significant brain activation differences in the Civilians minus Soldiers contrast (A) and Soldiers minus Civilians contrast (B) displayed on the ch2better template using MRIcron (http://www.mricro.com/mricron). See also Supplementary Table S1 for the main effect of conditions analysis, Civilians minus Control contrast and Soldiers minus Control contrast.

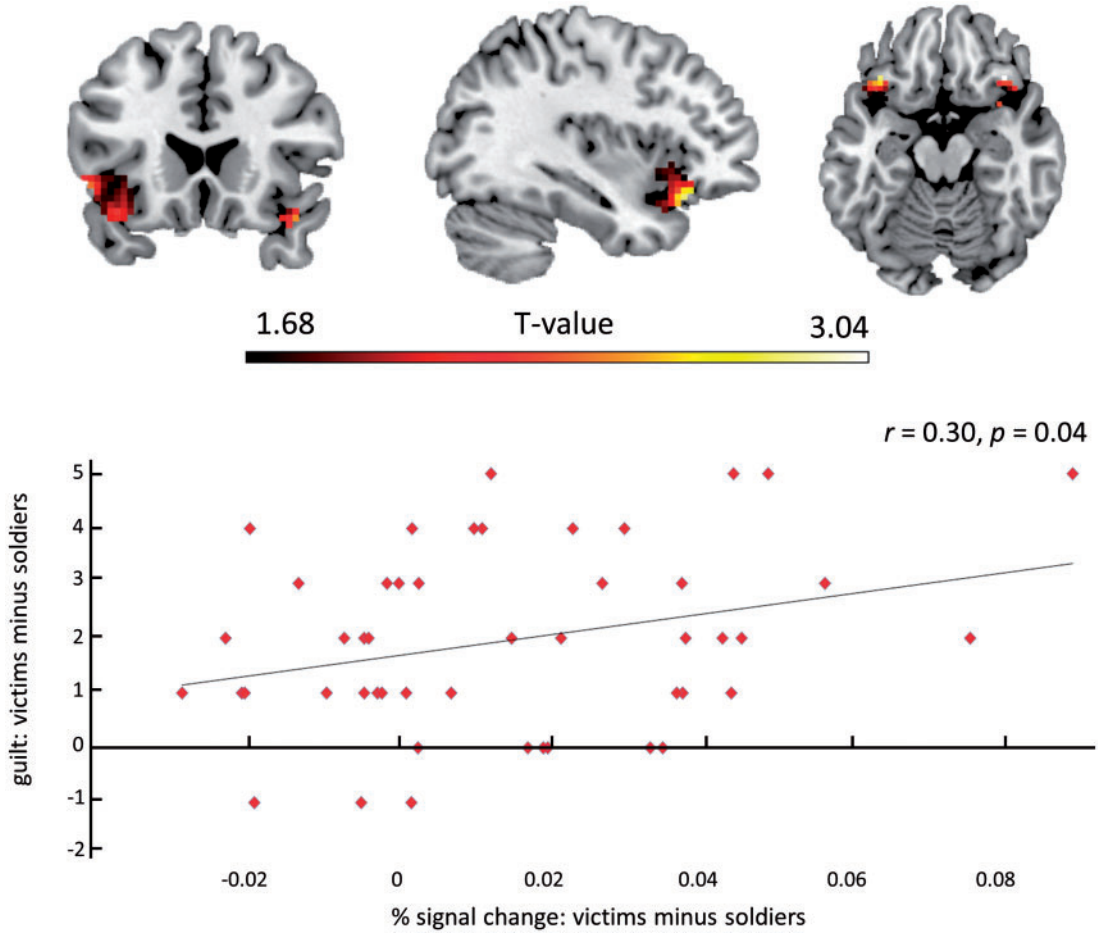

Fig. 3 The more guilt participants felt about shooting the civilians *vs* soldiers, the higher the % signal change in left and right OFC. The same regression in SPM is displayed on the ch2better template using MRIcron (http://www.mricro.com/mricron).

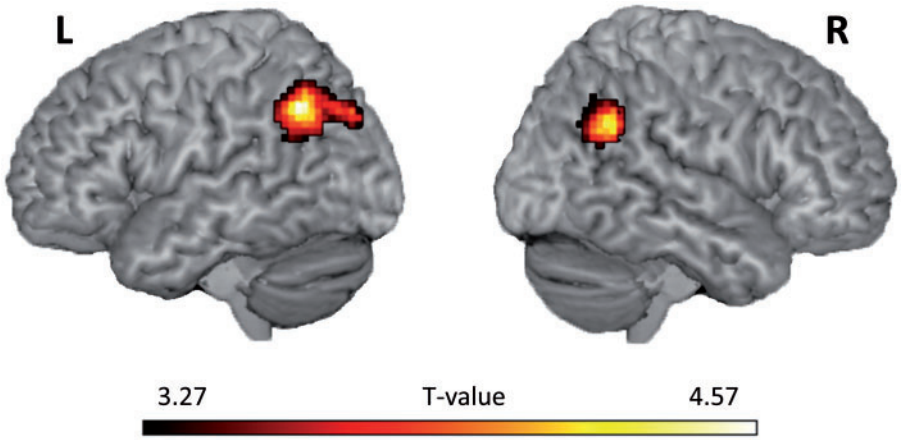

Fig. 4 Psychophysiological analysis. Significant increased connectivity between left OFC and left and right TPJ for the Civilians minus Soldiers contrast displayed on a ch2better rendered template using MRIcron (http://www.mricro.com/mricron).

rival university (Molenberghs *et al.*, 2014b). However, this might not be surprising given that there was no strong animosity between the two groups. Here in a more extreme situation, the group membership of the victim (i.e. enemy soldier *vs* innocent civilian) did have a significant effect on the neural responses associated with intentionally harming others.

The distinctive role of medial and lateral OFC in morality fits well with the model by Janoff-Bulman *et al.* (2009) who describe two types of morality. Prescriptive morality is focused on rewards, such as praise and pride and therefore initiates actions. On the other hand, proscriptive morality is focused on punishments, such as blame and guilt which in turn inhibits actions (i.e. killing civilians is punished through

increased activation in lateral OFC). This is in line with our results, which showed that participants felt guiltier when imagining shooting civilians compared with soldiers and the more guilt they felt, the more activation was found in lateral OFC. The lack of activation in bilateral lateral OFC for shooting soldiers, suggest that the normal associations implicated with harming another human being are not being activated when the target is a soldier. We do not imply that the reduced activation of the lateral OFC for killing soldiers is an active process in which this region is 'actively' inhibited in this condition. Rather we suggest that because the action is seen as justified, there is no need to associate the action with negative reinforcement (i.e. through increased activation the lateral OFC).

It should be noted however that this is just one interpretation of the OFC results. An alternative view would be that people were less inclined to take the perspective from a person who is shooting an innocent civilian compared with a person shooting a soldier. Because the OFC is often activated when people inhibit an aversive or painful sensation (Hooker and Knight, 2006), less activation in this region for the Soldiers condition could be a result of the reduced effort to regulate a person's emotions in this condition. Regardless of the interpretation, our results clearly show a differential role of the lateral OFC in justified and unjustified violence.

Lateral OFC also showed increased connectivity with bilateral TPJ, for the civilians minus soldiers contrast. This suggests that the increase in OFC activity for unjustified violence is subserved by increased coupling with the TPJ. Previous research has found that the TPJ is often involved in lower-level processes associated with the sense of agency and reorienting attention to salient stimuli, as well as higher-level cognitive processing tasks involved in social cognitions, such as empathy, morality and Theory of Mind (Saxe and Wexler, 2005; Decety and Lamm, 2007; Young and Dungan, 2012). The activation of the TPJ during moral judgement tasks is believed to be a result of an increase of inferences of mental states and intentions made by the participant (Young *et al.*, 2007; Young and Saxe, 2009). Disruption of the TPJ by transcranial magnetic stimulation also causes participants to judge attempted harms as less morally forbidden and more morally permissible (Young *et al.*, 2010). This suggests that the TPJ plays an important role in moral behaviour and our results further show that increased connectivity between OFC and TPJ is an important aspect in highlighting immoral behaviour (i.e. killing an innocent civilian).

Finally, imagining shooting civilians compared with soldiers also increased activation in fusiform gyrus, while the opposite contrast increased activation in the precuneus and lingual gyrus. This difference in activation in these visual areas was unexpected given that the video clips were matched on several visual characteristics in our pilot experiment (see Supplementary Material for details). The fusiform gyrus, although not specific for faces *per se* but rather visual expertise in general (Gauthier *et al.*, 1999; Gauthier *et al.*, 2000; Xu, 2005), is typically associated with face processing (Kanwisher *et al.*, 1997) and specifically facial expressions (Ganel *et al.*, 2005). These results thus may suggest that participants were more focused on the faces of the civilians, when killing them. The precuneus on the other hand was more activated when imagining shooting soldiers and this area is typically activated during shifting spatial attention (Molenberghs *et al.*, 2007). This result might explain the increased activation in the early visual cortex (i.e. lingual gyrus), suggesting that participants were more focused on the movement of the soldiers and were refocussing their attention to accurately shoot them.

The present research is not without limitations and we readily acknowledge that imagining harming others during video games is not the same as real life situations. However, considering the limitations of the MRI environment and the need to control visual features between conditions, we believe our paradigm was optimal. Imagining being the agent who is acting in a harmful manner has been used successfully in the past to elicit neural responses involved in morality (Decety and Porges, 2011) and we found similar areas (i.e. lateral OFC) being activated as in our previous fMRI study when people believed they were actually hurting (i.e. giving electroshocks) others directly (Molenberghs *et al.*, 2014b). In addition, the use of video games allowed us to control the visual features of the stimuli (see Supplementary Material for details).

To conclude, imagining unjustified killings (i.e. civilians) resulted in higher levels of guilt, as well as increased activations in lateral OFC compared with justified killings. The data suggest that certain situations in which violence is seen as justified can lead to less activation of the typical brain responses associated with harming another human being. As such, these results have important implications for a better understanding of how ordinary people can override constraints to violent action against particular people in specific situations.

FUNDING

The work was supported by an ARC Early Career Research Award (DE130100120) and Heart Foundation Future Leader Fellowship (1000458) awarded to P.M., an ARC Discovery Grant (DP130100559) awarded to P.M. and J.D. and an ARC Discovery Grant (DP1092490) awarded to W.L.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

CONFLICT OF INTEREST

None declared.

REFERENCES

Aguirre, G., D'Esposito, M. (1998). Experimental design for brain fMRI. In: Bandettini, P.A., Moonen, C., editors. *Functional MRI*. Berlin: Springer Verlag, pp. 369–80.

Akitsuki, Y., Decety, J. (2009). Social context and perceived agency modulate brain activity in the neural circuits underpinning empathy for pain: an event-related fMRI study. *NeuroImage*, *47*, 722–34.

Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*, 268–77.

Anderson, N.E., Kiehl, K.A. (2012). The psychopath magnetized: insights from brain imaging. *Trends in Cognitive Sciences*, *16*, 52–60.

Antonucci, A.S., Gansler, D.A., Tan, S., Bhadelia, R., Patz, S., Fulwiler, C. (2006). Orbitofrontal correlates of aggression and impulsivity in psychiatric patients. *Psychiatry Research*, *147*, 213–20.

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

Berthoz, S., Grezes, J., Armony, J., Passingham, R., Dolan, R. (2006). Affective response to one's own moral violations. *Neuroimage*, *31*, 945–50.

Best, M., Williams, J.M., Coccaro, E.F. (2002). Evidence for a dysfunctional prefrontal circuit in patients with an impulsive aggressive disorder. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 8448–53.

Blair, J., Marsh, A., Finger, E., Blair, K., Luo, J. (2006). Neuro-cognitive systems involved in morality. *Philosophical Explorations*, *9*, 13–27.

Brewer, M.B. (1999). The psychology of prejudice: ingroup love and outgroup hate? *Journal of Social Issues*, *55*, 429–44.

Cheng, Y., Chen, C.Y., Lin, C.P., Chou, K.H., Decety, J. (2010). Love hurts: an fMRI study. *NeuroImage*, *51*, 923–9.

Cikara, M., Botvinick, M.M., Fiske, S.T. (2011). Us versus them social identity shapes neural responses to intergroup competition and harm. *Psychological Science*, *22*, 306–13.

Cooney, M. (2009). *Is killing wrong?: a study in pure sociology*. Charlottesville: University of Virginia Press.

Correll, J., Park, B., Judd, C.M., Wittenbrink, B. (2002). The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*, 1314–29.

Correll, J., Urland, G.R., Ito, T.A. (2006). Event-related potentials and the decision to shoot: the role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*, 120–8.

Decety, J., Cacioppo, S. (2012). The speed of morality: a high-density electrical neuroimaging study. *Journal of Neurophysiology*, *108*, 3068–72.

Decety, J., Chen, C., Harenski, C., Kiehl, K.A. (2013). An fMRI study of affective perspective taking in individuals with psychopathy: imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience*, *7*, 489.

Decety, J., Cowell, J.M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences*, *18*, 337–9.

Decety, J., Echols, S.C., Correll, J. (2009). The blame game: the effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience*, *22*, 985–97.

Decety, J., Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *The Neuroscientist*, *13*, 580–93.

Decety, J., Michalska, K.J., Kinzler, K.D. (2011). The developmental neuroscience of moral sensitivity. *Emotion Review*, *3*, 305–7.

Decety, J., Michalska, K.J., Kinzler, K.D. (2012). The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. *Cerebral Cortex*, *22*, 209–20.

Decety, J., Porges, E.C. (2011). Imagining being the agent of actions that carry different moral consequences: an fMRI study. *Neuropsychologia*, *49*, 2994–3001.

Eres, R., Molenberghs, P. (2013). The influence of group membership on the neural correlates involved in empathy. *Frontiers in Human Neuroscience*, *7*, 176.

Eslinger, P.J., Damasio, A.R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology*, *35*, 1731–41.

Fox, G.R., Sobhani, M., Aziz-Zadeh, L. (2013). Witnessing hateful people in pain modulates brain activity in regions associated with physical pain and reward. *Frontiers in Psychology*, *4*, 472.

Ganel, T., Valyear, K.F., Goshen-Gottstein, Y., Goodale, M.A. (2005). The involvement of the "fusiform face area" in processing facial expression. *Neuropsychologia*, *43*, 1645–54.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*, 191–7.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C. (1999). Activation of the middle fusiform'face area'increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–73.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–34.

Hein, G., Singer, T. (2008). I feel how you feel but not always: the empathic brain and its modulation. *Current Opinion in Neurobiology*, *18*, 153–8.

Hooker, C.I., Knight, R.T. (2006). The role of the lateral orbitofrontal cortex in the inhibitory control of emotion. In: Zald, D.H., Rauch, S.L., editors. *The Orbitofrontal Cortex*. New York: Oxford University Press, pp. 307–324.

Iannetti, G., Salomons, T.V., Moayedi, M., Mouraux, A., Davis, K.D. (2013). Beyond metaphor: contrasting mechanisms of social and physical pain. *Trends in Cognitive Sciences*, *17*, 371–8.

Jackson, P.L., Meltzoff, A.N., Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage*, *24*, 771–9.

Janoff-Bulman, R., Sheikh, S., Hepp, S. (2009). Proscriptive versus prescriptive morality: two faces of moral regulation. *Journal of Personality and Social Psychology*, *96*, 521–37.

Kamm, F.M. (1989). Harming some to save others. *Philosophical Studies*, *57*, 227–60.

Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–11.

Kédia, G., Berthoz, S., Wessa, M., Hilton, D., Martinot, J.-L. (2008). An agent harms a victim: a functional magnetic resonance imaging study on specific moral emotions. *Journal of Cognitive Neuroscience*, *20*, 1788–98.

Killen, M., Rutland, A. (2011). *Children and Social Exclusion: Morality, Prejudice, and Group Identity*. New York: Wiley-Blackwell.

King, J.A., Blair, R.J.R., Mitchell, D.G., Dolan, R.J., Burgess, N. (2006). Doing the right thing: a common neural circuit for appropriate violent or compassionate behavior. *Neuroimage*, *30*, 1069–76.

Klasen, M., Weber, R., Kircher, T.T., Mathiak, K.A., Mathiak, K. (2012). Neural contributions to flow experience during video game playing. *Social, Cognitive and Affective Neuroscience*, *7*, 485–95.

Kringelbach, M.L., Rolls, E.T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*, 341–72.

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*, 2492–502.

Mathiak, K., Weber, R. (2006). Toward brain correlates of natural behavior: fMRI during violent video games. *Humam Brain Mapping*, *27*, 948–56.

Mathiak, K.A., Klasen, M., Weber, R., Ackermann, H., Shergill, S.S., Mathiak, K. (2011). Reward system and temporal pole contributions to affective evaluation during a first person shooter video game. *BMC Neuroscience*, *12*, 66.

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience and Biobehavioral Reviews*, *37*, 1530–6.

Molenberghs, P., Gapp, J., Wang, B., Louis, W.R., Decety, J. (2014a). Increased moral sensitivity for outgroup perpetrators harming ingroup members. *Cerebral Cortex,* doi: 10.1093/cercor/bhu195.

Molenberghs, P., Bosworth, R., Nott, Z., et al. (2014b). The influence of group membership and individual differences in psychopathy and perspective taking on neural responses when punishing and rewarding others. *Human Brain Mapping*, *10*, 4989–99.

Molenberghs, P., Mesulam, M.M., Peeters, R., Vandenberghe, R.R.C. (2007). Remapping attentional priorities: differential contribution of superior parietal lobule and intraparietal sulcus. *Cerebral Cortex*, *17*, 2703–12.

Moll, J., de Oliveira-Souza, R., Garrido, G.J., et al. (2007). The self as a moral agent: linking the neural bases of social agency and moral sensitivity. *Social Neuroscience*, *2*, 336–52.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*, 799–809.

Nucci, L. (1981). Conceptions of personal issues a domain distinct from moral or societal concepts. *Child Development*, *52*, 114–21.

Pascual, L., Rodrigues, P., Gallardo-Pujol, D. (2013). How does morality work in the brain? A functional and structural perspective of moral behavior. *Frontiers in Integrative Neuroscience*, *7*, 65.

Porges, E.C., Decety, J. (2013). Violence as a source of pleasure or displeasure is associated with specific functional connectivity with the nucleus accumbens. *Frontiers in Human Neuroscience*, *7*, 447.

Raine, A., Yang, Y. (2006). Neural foundations to moral reasoning and antisocial behavior. *Social, Cognitive and Affective Neuroscience*, *1*, 203–13.

Robertson, D., Snarey, J., Ousley, O., et al. (2007). The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, *45*, 755–66.

Roy, M., Shohamy, D., Wager, T.D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*, 147–56.

Saver, J.L., Damasio, A.R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia*, *29*, 1241–9.

Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporoparietal junction. *Neuropsychologia*, *43*, 1391–9.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34.

Seeley, W.W., Menon, V., Schatzberg, A.F., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*, 2349–56.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–62.

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

Sobhani, M., Bechara, A. (2011). A somatic marker perspective of immoral and corrupt behavior. *Social Neuroscience*, *6*, 640–52.

Turiel, E. (2014). Morality: epistemology, development, and social opposition. In: Killen, M., Smetana, J.G., editors. *Handbook of Moral Development,* 2nd edn. New York: Psychology Press, pp. 3–22.

van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*, 829–58.

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*, 8525–9.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, *15*, 1234–42.

Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 6753–8.

Young, L., Cushman, F., Hauser, M., Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 8235–40.

Young, L., Dungan, J. (2012). Where in the brain is morality? Everywhere and maybe nowhere. *Social Neuroscience*, *7*, 1–10.

Young, L., Saxe, R. (2009). An fMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*, 1396–1405.

Zahn, R., de Oliveira-Souza, R., Moll, J. (2011). The neuroscience of moral cognition and emotion. In: Decety, J., Cacioppo, J.T., editors. *The Oxford Handbook of Social Neurosciene*. Oxford: University Press, pp. 477–90.