doi:10.1093/scan/nsr005 SCAN (2012) 7, 393–402

The neural basis of intuitive and counterintuitive moral judgment

Guy Kahane,[1,2,*] Katja Wiech,[3,4,*] Nicholas Shackel,[2,5] Miguel Farias,[6] Julian Savulescu,[1,2] and Irene Tracey[3,4]
[1]Oxford Centre for Neuroethics, [2]Oxford Uehiro Centre for Practical Ethics, Faculty of Philosophy, University of Oxford, Suite 8, Littlegate House, St Ebbes Street, Oxford OX1 1PT, UK, [3]Nuffield Department of Anaesthetics, [4]Department of Clinical Neurology, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, [5]Department of Philosophy, University of Cardiff, Colum Drive, Cathays, Cardiff, CF10 3EU and [6]Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, UK

Neuroimaging studies on moral decision-making have thus far largely focused on differences between moral judgments with opposing utilitarian (well-being maximizing) and deontological (duty-based) content. However, these studies have investigated moral dilemmas involving extreme situations, and did not control for two distinct dimensions of moral judgment: whether or not it is intuitive (immediately compelling to most people) and whether it is utilitarian or deontological in content. By contrasting dilemmas where utilitarian judgments are counterintuitive with dilemmas in which they are intuitive, we were able to use functional magnetic resonance imaging to identify the neural correlates of intuitive and counterintuitive judgments across a range of moral situations. Irrespective of content (utilitarian/deontological), counterintuitive moral judgments were associated with greater difficulty and with activation in the rostral anterior cingulate cortex, suggesting that such judgments may involve emotional conflict; intuitive judgments were linked to activation in the visual and premotor cortex. In addition, we obtained evidence that neural differences in moral judgment in such dilemmas are largely due to whether they are intuitive and not, as previously assumed, to differences between utilitarian and deontological judgments. Our findings therefore do not support theories that have generally associated utilitarian and deontological judgments with distinct neural systems.

Keywords: neuroimaging; moral judgment; decision-making; functional magnetic resonance imaging

INTRODUCTION

Is it morally permissible to kill a stranger by pushing him onto the track of a runaway trolley in order to save the lives of five others? To sacrifice one life to save five is to act in line with utilitarianism, the view that we should maximize aggregate well-being, regardless of the means employed (Singer, 2005). By contrast, deontological ethical views such as Kant's ethics hold that we must obey certain duties even when this leads to a worse outcome. Many deontologists thus think that it would be wrong to kill the stranger (Kamm, 2000).

Recent neuroimaging studies of moral-decision making have focused on such extreme moral dilemmas (Greene *et al.*, 2001, 2004). Utilitarian (well-being maximizing) judgments were found to be associated with longer response times (RT) and with increased activation in the dorsolateral prefrontal cortex (DLPFC) and inferior parietal lobe, areas implicated in deliberative processing; deontological judgments were associated with greater activation in areas related to affective processing, such as the ventromedial prefrontal cortex, the superior temporal sulcus and the amygdala. These differences in neural activation have been interpreted to reflect distinct neural sub-systems that underlie utilitarian and deontological moral judgments not only in the context of such extreme dilemmas, but quite generally (Greene, 2008).

However, this general theoretical proposal requires further investigation, given that dilemmas involving extreme harm to others are only one kind of moral context in which utilitarian and deontological judgments conflict. Moreover, such extreme moral dilemmas are distinctive in an important way. When asked whether to push a stranger to save five, a large majority chooses the deontological option, a decision that appears to be based on immediate intuitions (Cushman *et al.*, 2006), in line with extensive psychological evidence that moral judgments are often made in this automatic way (Haidt, 2001). Utilitarian judgments in such dilemmas are often highly counterintuitive because they conflict with a stringent duty not to harm. Utilitarian choices, however, can also conflict with less stringent duties, such as duties not to lie or break promises (Ross, 1930/2002). In such cases, it's often the *deontological* choice that appears strongly counterintuitive, as in Kant's notorious contention that lying is forbidden even to prevent murder (Kant, 1797/1966). Most people believe that we are permitted to break a promise or lie if this is necessary to prevent great harm to others.

Received 7 April 2010; Accepted 16 January 2011
Advance Access publication 18 March 2011
The authors are grateful to Amy Bilderbeck for help with data collection. This work was funded by the Wellcome Trust (WT087208MF and WT086041MA) and Fundacao Bial, Portugal (64/06).
*These authors contributed equally to this work.
Correspondence should be addressed to Dr Guy Kahane, Oxford Uehiro Centre for Practical Ethics, Littlegate House, St Ebbe's Street, Oxford OX1 1PT, UK. E-mail: guy.kahane@philosophy.ox.ac.uk

In such situations it is rather the *utilitarian* choice to maximize well-being that is intuitive.

Prior research has therefore not distinguished two distinct variables: the content of a moral judgment—whether it is utilitarian or not—and how intuitive or immediately compelling this judgment is to most people. Thus the reported differences between utilitarian and deontological judgments might be due to the greater intuitiveness of deontological choices in the moral dilemmas previously examined, rather than to a general division between utilitarian and deontological modes of moral judgment. Prior neuroimaging studies therefore offer only limited measures of the neural processes that might generally underlie deontological and utilitarian judgments. More importantly, they offer only limited measures of the processes that might generally underlie intuitive and counterintuitive judgments. Consequently, this key division in the psychology of moral decision-making has not yet been directly investigated at the neural level.

Using functional magnetic resonance imaging (fMRI) in healthy volunteers, we investigated the neural bases of intuitive and counterintuitive moral judgments across different types of moral situations, while controlling for their content. We used a selection of the extreme dilemmas used in prior studies which were controlled for content and intuitiveness, as well as new dilemmas involving different contexts, where intuitiveness and content were reversed ('Materials and Methods' section). This design allowed us to investigate not only the neural correlates of intuitive and counterintuitive moral judgments, but also the neural correlates of utilitarian and deontological judgments across a range of moral contexts, when intuitiveness is controlled.

We hypothesized that counterintuitive judgments are associated with controlled processing regardless of their content. If the content of a judgment is more critical than its intuitiveness, we would expect similar brain activations for the same content (e.g., utilitarian judgment), irrespective of intuitiveness. At the behavioural level, an increased cognitive effort during utilitarian judgments should be reflected in higher difficulty ratings and longer RTs. By contrast, if 'intuitiveness' is the critical factor as we hypothesize, then, irrespective of the content of judgments (utilitarian *vs* deontological), similar neural activations should be observed for judgments of the same degree of intuitiveness (intuitive *vs* counterintuitive), and counterintuitive judgments should be associated with longer RTs and higher difficulty ratings.

MATERIALS AND METHODS

Subjects

Sixteen healthy, right-handed subjects (9 females, mean age: 29.25 years, range: 21–41) participated in the study. The volunteers were pre-assessed to exclude those with a previous history of neurological or psychiatric illness. All subjects gave informed consent, and the study was approved by the local Research Ethics Committee.

Experimental procedures

To study the differential effect of the content (deontological/utilitarian) and the intuitiveness (intuitive/counterintuitive) of the judgment, we used two different sets of dilemmas: scenarios where the utilitarian option is intuitive (UI dilemmas) and scenarios where the deontological judgment is intuitive (DI dilemmas; for criteria of classification, see 'Stimuli' section). Depending on the decision of the participant, trials were subsequently classified as (i) DI_U: DI dilemma, utilitarian decision, (ii) DI_D: DI dilemma, deontological decision, (iii) UI_U: UI dilemma, utilitarian decision and (iv) UI_D: UI dilemma, deontological decision.

The experiment was divided into four sessions, each lasting for ~10 min. The order of presentation of DI and UI dilemmas was randomized throughout. Each dilemma was presented as text through a series of three screens. The first two described a dilemma, and the third suggested a possible solution. After reading the third screen, subjects responded by pressing one of two buttons indicating whether they agreed with the suggested solution ('yes' or 'no'). For half the subjects, the left button was used for 'yes', for the other half pressing the left button indicated disagreement with the suggested solution ('no'). Participants were instructed to read the text and press the button as soon as they had made their decision. No visual feedback was given upon decision.

On arrival, participants were provided with written task instructions and gave their informed consent. They were told that the purpose of the study was to investigate decision-making in moral situations. They were assured that the study was not a test of moral integrity. Subsequently, all participants filled in personality questionnaires (data not shown here). Once they were positioned in the MR scanner, participants were familiarized with the presentation of the dilemmas, the response box and the rating procedure (see below). A test paradigm was run with two example dilemmas to acquaint subjects with the structure of the experiment. The dilemmas were displayed on a black screen (white letters, font: Arial) located above the feet of the subjects. A test image was presented on the screen prior to scanning to ensure that the image was in focus and the participant could comfortably read the text. At the end of each dilemma subjects were prompted to rate its difficulty using a Numerical Rating Scale ranging from 0 (='not difficult at all') to 100 (='very difficult'). Participants were given 6 s for the rating. At the end of each trial, subjects were instructed to fixate a white cross that was displayed in the centre of the screen for 12 s (baseline). The time between presentation of the third part of the dilemma and the button press indicating the subject's decision were recorded as RT (in ms).

Stimuli

We used scenarios where one of a range of moral duties (e.g. not to lie or kill) conflicts with choosing the outcome with

the greater aggregate well-being, in line with utilitarianism. For simplicity, we refer to the latter option as 'utilitarian' and the former as 'deontological', although 'utilitarian' choices in this sense needn't imply an overall utilitarian outlook (Kahane and Shackel, 2008; 2010). The scenarios included a selection of 'personal' dilemmas previously used by Greene *et al.* (2001, 2004) as well as new dilemmas (Supplementary Data). In order to classify the scenarios into DI and UI dilemmas, all scenarios were pre-assessed by 18 independent judges who reported their unreflective response to each dilemma. On this basis, 8 dilemmas for which 12 or more judges chose the deontological option were classified as 'deontological intuitive' (DI), and 10 dilemmas for which 12 or more judges chose the utilitarian option as 'utilitarian intuitive' (UI). As expected, most DI dilemmas were scenarios previously used by Greene *et al.* (2001, 2004) where the better consequence required violating a duty not to harm (five out of eight DI dilemmas; Supplementary Data), and most UI dilemmas involved a conflict between the better consequence and other duties (e.g. not to lie).

Image acquisition

A 3 T scanner (Oxford Magnet Technology, Oxford, UK) was used to acquire T2*-weighted echoplanar images (repetition time: 2.38 s, echo time: 30 ms; flip angle: 90°; matrix: 64×64; field of view: 192×192 mm^2; slice thickness: 3 mm) with BOLD contrast.

Data analysis

The numbers of intuitive and counterintuitive judgments were compared separately for both types of dilemmas using paired *t*-tests. Pearson correlation coefficients were calculated for the correlation between (i) the number of counterintuitive judgments in UI and DI dilemmas, (ii) the number of utilitarian judgments in both types of dilemmas. For RT and difficulty ratings we used an ANOVA to analyse the difference between (i) utilitarian and deontological judgments, (ii) intuitive and counterintuitive judgments and (iii) between UI and DI dilemmas (all main effects). Furthermore, RT and difficulty ratings were analyzed separately for each type of dilemma using paired *t*-tests.

Pre-processing and statistical analysis of the fMRI data were carried out using SPM5 (www.fil.ion.ucl.ac.uk/spm). The first five image volumes of each session were discarded to account for T1 relaxation effects. The remaining volumes were realigned to the sixth volume to correct for head motion before statistical analysis. The EPI images were spatially normalized (Friston *et al.*, 1995) to the template of the Montréal Neurological Institute (MNI; Evans *et al.*, 1993). The normalized EPI-images were smoothed using an 8-mm full-width at half maximum (FWHM) Gaussian kernel, temporally high-pass filtered (cut-off 128 s) and corrected for temporal autocorrelations using first-order autoregressive modelling. For each subject, contrast images were calculated for each of the four possible outcomes (i.e. UI_U, UI_D, DI_U, DI_D). Given that all information about the dilemmas was available with the presentation of the second screen and in order to capture the early phase of decision-making, decision-related activity was modelled as events from 4 s prior to the presentation of the question (third screen) until the button press indicating the decision. The remaining time until the end of the third screen as well as the time of the first and second screen were modelled as regressors-of-no-interest. First level contrasts were taken to the second level for the group data analysis using a flexible factorial design within a random effects model.

Data analysis on the group level was divided into two stages. First, we investigated all three main effects to identify brain regions generally associated with the content (utilitarian/deontological) and intuitiveness of the decision (intuitive/counterintuitive) as well as with the type of dilemma (DI/UI). Brain regions activated during intuitive and counterintuitive moral judgments were identified comparing both types of decisions irrespective of their utilitarian or deontological content (analysis A; Figure 1A). Likewise, utilitarian and deontological moral judgments were compared pooled across intuitive and counterintuitive decisions (analysis B, Figure 1B). Finally, brain responses to the two types of dilemmas were compared, irrespective of the decision made (analysis C, Figure 1C). Note that this last analysis is statistically identical to the interaction analysis between content and intuitiveness.

At the second stage of the analysis, we tested whether neural differences between the utilitarian and deontological judgments in DI dilemmas were due to intuitiveness or content (Figure 1, analyses D and E). To this end, we compared one judgment to the contrary judgment within DI dilemmas, e.g. DI_U > DI_D (analysis D). The results of these comparisons are ambiguous given that the two options differ both in content and in intuitiveness. We therefore performed additional analyses where we compared the same judgment (e.g. DI_U) to the two options in the other type of dilemma [i.e. 'DI_U > UI_U' (analysis D1) and 'DI_U > UI_D' (analysis D2)]. In order to test for similarities between these additional analyses and the original comparison, we used the result of the original comparison (DI_U *vs* DI_D) as an inclusive mask ($P < 0.05$, uncorrected) in the two subsequent analyses (D1 and D2). As our DI dilemmas are closest to dilemmas previously used (e.g. Greene *et al.*, 2004), we only report below findings on this type of dilemma; for analyses of UI dilemmas see Supplementary Tables S10–S13). For consistency reasons, a global threshold was set at $P < 0.001$ uncorrected for all analyses with a minimum cluster extent of five continguous voxels. Activation clusters surviving a more conservative threshold of $P < 0.05$ FWE-corrected are marked with an asterisk in the Supplementary Tables. All coordinates are given in MNI space.

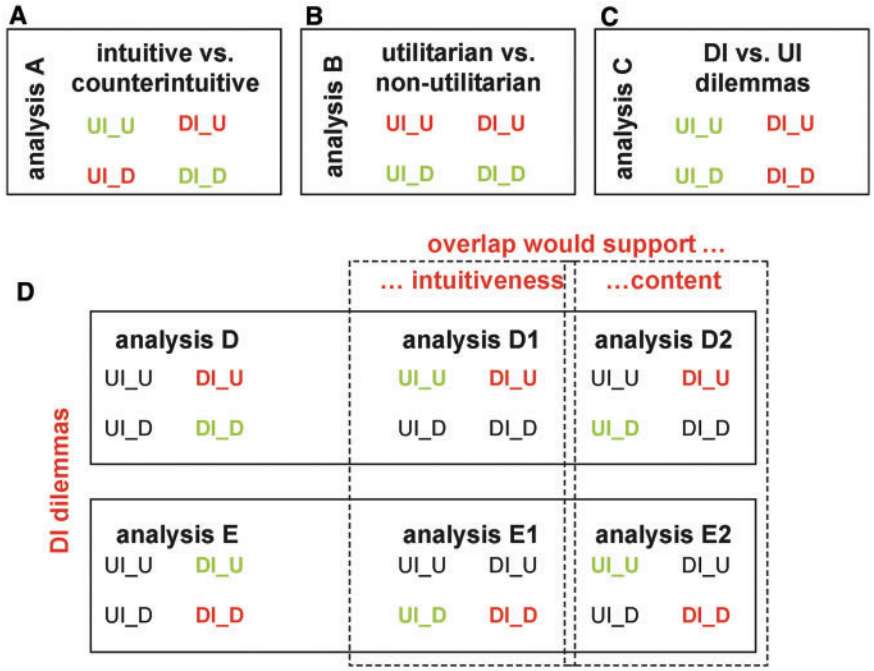

Fig. 1 Overview of fMRI data analysis. (**A**) Brain responses to utilitarian moral judgments (UI_U and DI_U) were compared to responses to deontological moral judgments (UI_D and DI_D). (**B**) Comparison of intuitive (UI_U and DI_D) *vs* counterintuitive moral judgments (UI_D and DI_U). (**C**) Comparison of moral judgments in DI dilemmas (DI_D and DI_U) *vs* judgments in UI dilemmas (UI_U and UI_D). (**D**) Comparison of single conditions. In analysis D, utilitarian judgments in DI dilemmas were compared to (i) deontological judgments in DI dilemmas (DI_U *vs* DI_D; analysis D), (ii) utilitarian judgments in UI dilemmas (DI_U *vs* UI_U; analysis D1) and (iii) deontological judgments in UI dilemmas (DI_U *vs* UI_D; analysis D2). Analysis E (deontological judgments in DI dilemmas) follows a parallel form. The dilemma that is substracted is marked in green, the dilemma that is subtracted from is marked in red.

RESULTS

Behavioural data

Decisions

In both types of dilemmas, participants chose the intuitive option more often than the counterintuitive option [UI dilemmas: $t(15) = 5.81$, $P < 0.001$; DI dilemmas: $t(15) = -4.16$, $P = 0.001$; Figure 2A]. The number of counterintuitive judgments was not correlated between categories ($r = 0.14$, $P = 0.606$). Likewise, the number of utilitarian judgments in UI dilemmas was not correlated with the number of utilitarian judgments in DI dilemmas ($r = -0.14$, $P = 0.599$).

RT and difficulty ratings

For RT and difficulty ratings, we first tested whether utilitarian judgments took longer and were perceived as more difficult. A comparison of utilitarian and deontological moral judgments across dilemmas revealed no significant difference in difficulty rating [$F(1,11) = 0.06$, $P = 0.811$; Figure 2B] or RT [$F(1,11) = 0.314$, $P = 0.586$; Figure 2C]. In contrast, a significant difference in perceived difficulty was observed between intuitive and counterintuitive judgments, the latter being more difficult [$F(1,11) = 24.95$, $P < 0.001$]. RT were not significantly different between intuitive and counterintuitive decisions [$F(1,11) = 0.272$, $P = 0.612$]. However, given that decisions in DI dilemmas took longer [$F(1,11) = 7.627$, $P = 0.018$] and received higher difficulty ratings than in UI dilemmas [$F(1,11) = 18.917$, $P = 0.001$], we performed additional analyses on both measures separately for both types of dilemmas. RTs were not significantly different between both options [DI dilemmas: $t(13) = 0.029$, $P = 0.977$; UI dilemmas: $t(13) = -0.550$, $P = 0.592$]. However, in DI dilemmas the counterintuitive utilitarian decision was perceived as more difficult [$t(13) = 2.564$, $P = 0.024$], whereas in UI dilemmas the counterintuitive deontological judgment got higher difficulty ratings [$t(13) = -2.747$, $P = 0.017$].

Neuroimaging data: effects of intuitiveness, content and type of dilemma

Intuitive vs counterintuitive moral judgments

The contrast 'intuitive > counterintuitive decisions' revealed significant effects in the visual cortex, left premotor cortex, bilateral mid temporal lobe (extending into the right temporal pole) and left lateral orbitofrontal cortex (OFC; Figure 3A and Supplementary Table S1). The reverse contrast ('counterintuitive > intuitive decisions') showed significant effects in the rostral anterior cingulate cortex (rACC) extending into the dorsal part of the ACC, right secondary somatosensory cortex (SII) extending into the primary somatosensory cortex (SI) and posterior insula, bilateral mid insula extending into the temporal lobe, right ventrolateral prefrontal cortex (VLPFC) and lateral OFC (Figure 3B and Supplementary Table S2).

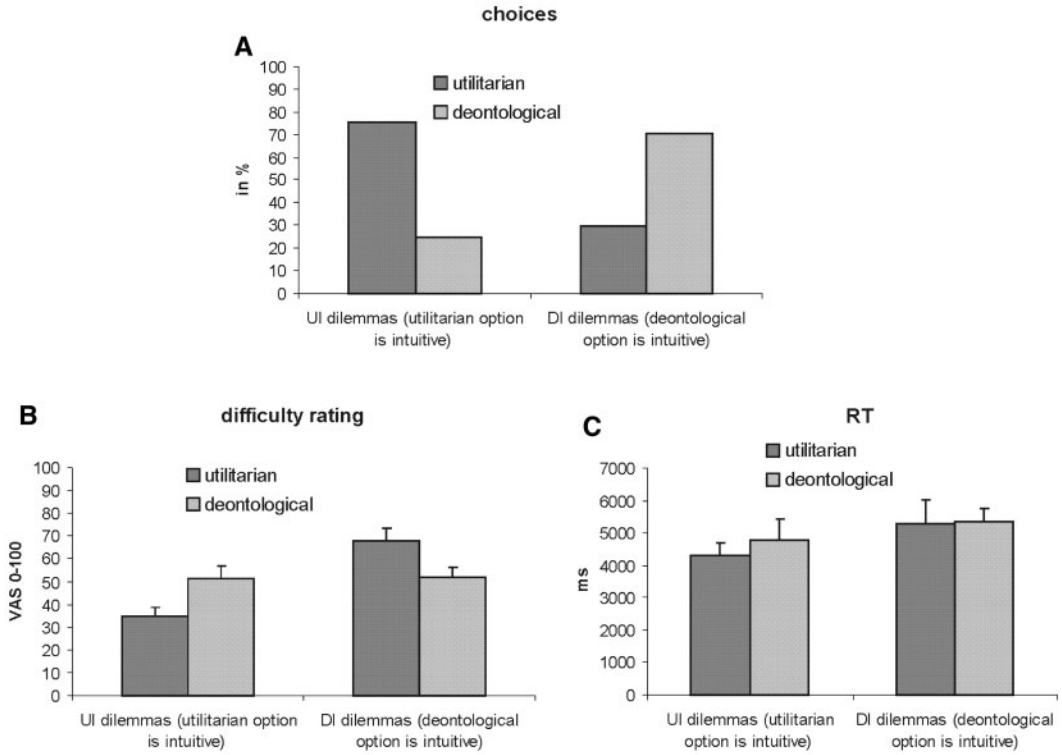

Fig. 2 Behavioral data. (**A**) Relative number of utilitarian and non-utilitarian judgments (averaged across subjects) in DI dilemmas where the deontological option was considered intuitive and UI dilemmas where the utilitarian option was considered intuitive. Participants chose the intuitive option significantly more often than the counterintuitive option in both types of dilemmas ($P \leq 0.001$). (**B**) Difficulty rating for utilitarian and deontological judgments in DI and UI dilemmas averaged across subjects. In both types of dilemmas, counterintuitive judgments were rated as more difficult compared to intuitive judgments ($P < 0.05$). (**C**) Response times for utilitarian and non-utilitarian judgments in DI and UI dilemmas averaged across subjects. Significantly longer response times were found for DI than for UI dilemmas but not for counterintuitive compared to intuitive judgments. Error bars show standard errors.

Utilitarian vs deontological moral judgments

Brain responses in deontological and utilitarian judgments were next compared irrespective of dilemma type. Utilitarian judgments showed no specific significant activation, whereas deontological judgments were characterized by an increased activation in the posterior cingulate cortex (PCC) and right temporo-parietal junction (TPJ; Figure 4, Supplementary Table S3).

Moral judgments in DI vs UI dilemmas

Compared to UI dilemmas, DI dilemmas exhibited stronger activation in the right DLPFC extending into VLPFC, the right TPJ and the occipital lobe (Figure 5; Supplementary Table S4). UI dilemmas did not show any specific significant activation relative to DI dilemmas.

Neuroimaging data: content vs intuitiveness

DI dilemmas: utilitarian > deontological moral judgments (DI_U > DI_D; analysis D)

This analysis revealed increased activation in the right mid insula, lateral OFC on both sides extending into the VLPFC on the right side, in rACC, right SII and left superior temporal lobe (Figure 6, analysis D; Supplementary Table S5). The activations identified in this contrast were used as an inclusive mask for two subsequent analyses. Comparing DI_U with UI_D (both counterintuitive, different content) revealed overlapping activations in the visual cortex only (Figure 6, analysis D1; Supplementary Table S6). In contrast, the comparison between DI_U and UI_U (different intuitiveness, both utilitarian judgments) showed an overlap with result D in the rACC, right VLPFC and SII, as well as the visual cortex and cerebellum (Figure 6, analysis D2; Supplementary Table S7). Finally, ROI analyses on regions previously reported for utilitarian judgment using similar dilemmas (Greene *et al.*, 2004; see Supplementary Methods) revealed no significant activation.

DI dilemmas: deontological > utilitarian moral judgments (DI_D > DI_U; analysis E)

In DI dilemmas, intuitive deontological judgments were accompanied by increased activation in the visual cortex, bilateral temporal lobe covering more posterior parts on the left side and more anterior parts including the temporal pole on the left side. Additional activation was observed in the left premotor and supplementary motor regions as well as in lateral OFC on both sides (Figure 6, analysis E;

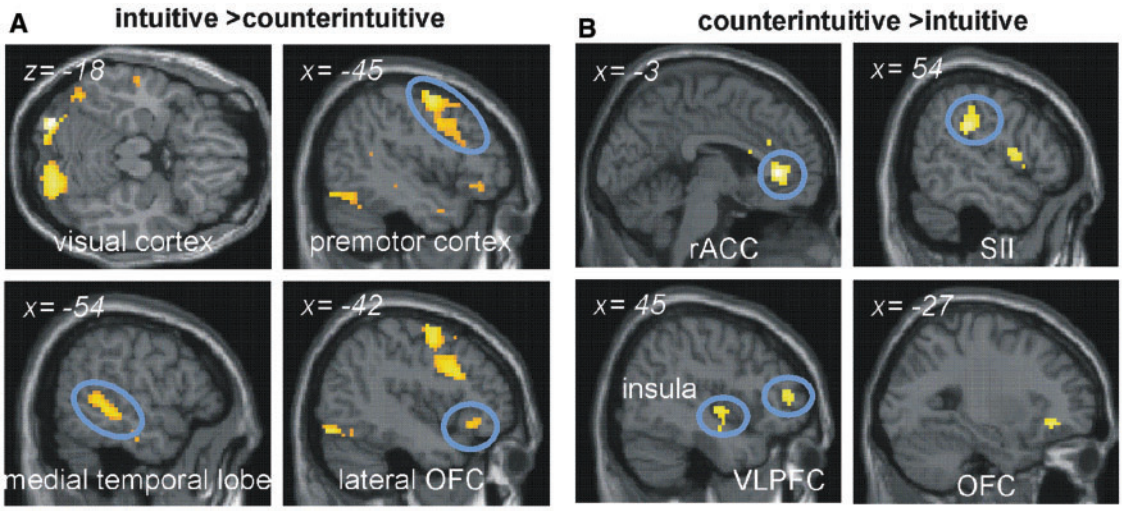

Fig. 3 Comparison of brain responses to intuitive and counterintuitive moral judgments. (**A**) Intuitive moral judgments were associated with increased activation in the visual, premotor and orbitofrontal cortex and the temporal lobe. (**B**) During counterintuitive moral judgments, increased activation was observed in the dorsal and rostral ACC, SII, insula, VLPFC and OFC.

Supplementary Table S8). The comparison of DI_D with UI_U (both intuitive, different content) revealed no significant overlap with the result of analysis E (Figure 6, analysis E1). In contrast, the comparison of DI_D with UI_D (different intuitiveness, same content) showed significant overlap with the result of analysis E in the visual cortex, left premotor cortex and bilateral OFC (Figure 6, analysis E2; Supplementary Table S9).

DISCUSSION

Our study aimed to identify the behavioural and neural correlates of intuitive and counterintuitive judgments, when content is controlled, and the correlates of deontological and utilitarian judgments, when intuitiveness is controlled, allowing us to disentangle the distinct contributions made by intuitiveness and content to the processes involved in responses to moral dilemmas.

Previous neuroimaging studies reported that utilitarian judgments in dilemmas involving extreme harm were associated with activation in the DLPFC and parietal lobe (Greene *et al.*, 2004). This finding has been taken as evidence that utilitarian judgment is generally driven by controlled processing (Greene, 2008). The behavioural and neural data we obtained suggest instead that differences between utilitarian and deontological judgments in dilemmas involving extreme harm largely reflect differences in intuitiveness rather than in content.

Overall, counterintuitive judgments were perceived as more difficult than intuitive judgments, whereas there was no significant difference in perceived difficulty between utilitarian and deontological judgments. At the neural level, counterintuitive and intuitive decisions analysed across the two types of dilemmas were characterized by robust activation in extended networks, as discussed below. In contrast,

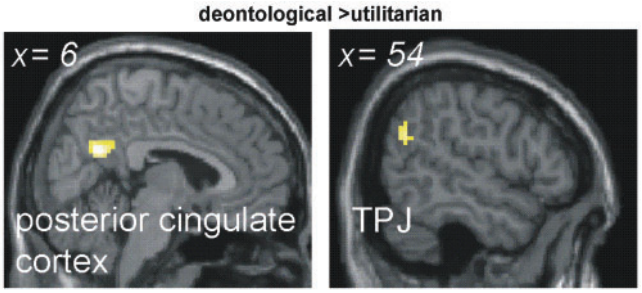

Fig. 4 Comparison of brain responses to deontological and utilitarian moral judgments. Deontological moral judgments led to increased activation in the PCC and the right TPJ. No significant activation was found for the comparison 'utilitarian > deontological moral judgments'.

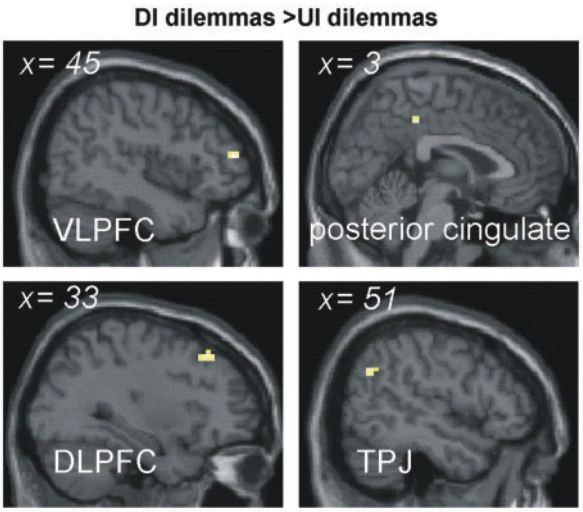

Fig. 5 Comparison of brain responses to moral judgments in DI and UI dilemmas. During moral judgments in DI dilemmas, increased activation was found in the right VLPFC and DLPFC, PCC, right TPJ. No significant activation was found for the comparison 'UI dilemmas > DI dilemmas'.

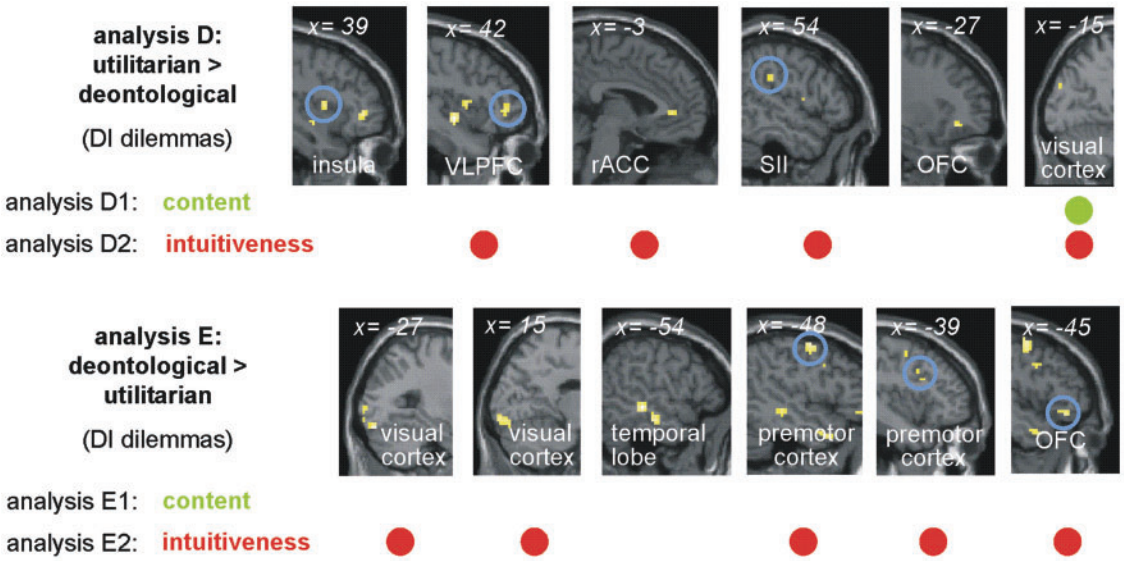

Fig. 6 Analysis of the role of intuitiveness and content in judgments of DI dilemmas. (**A**) Analysis D (DI_U > DI_D): compared to deontological moral judgments in DI dilemmas, utilitarian judgments were associated with increased activation in the right insula, VLPFC, SII, left OFC, rACC and visual cortex. (D1) Of these regions, only the visual cortex was also activated in the comparison of DI_U with deontological judgments in UI dilemmas (indicated by green dots). (D2) In contrast, overlap with the results of analysis D was found in the VLPFC, rACC, SII and visual cortex when DI_U was compared with utilitarian judgments in UI dilemmas (indicated by red dots). Analysis E (DI_D > DI_U): compared to utilitarian moral judgments in DI dilemmas, deontological judgments were associated with increased activation in the visual cortex, bilateral temporal lobe, left premotor and right orbitofrontal cortex. (E1) Of these regions, none showed increased activation when DI_D was compared with utilitarian judgments in UI dilemmas (indicated by green dots). (E2) In contrast, overlap was found in the visual, premotor and orbitofrontal cortex when DI_D was compared with deontological judgments in UI dilemmas (indicated by red dots).

when neural responses were analysed according to their content (i.e. pooling across intuitive and counterintuitive decisions), deontological judgments showed increased activation in the PCC and right TPJ, but not in brain regions previously associated with deontological decisions (Greene *et al.* 2001, 2004). Utilitarian judgments did not exhibit any specific significant activations.

To further investigate whether neural differences were due to intuitiveness rather than content of the judgment, we performed the additional analyses D–G (Figure 6 and Supplementary Figures S1 and S2). When we controlled for content, these analyses showed considerable overlap for intuitiveness. In contrast, when we controlled for intuitiveness, only little—if any—overlap was found for content. Our results thus speak against the influential interpretation of previous neuroimaging studies as supporting a general association between deontological judgment and automatic processing, and between utilitarian judgment and controlled processing.

Importantly, similar results were obtained even when we considered only the contrast between utilitarian and deontological judgments in DI dilemmas (Figure 6), a category of dilemmas that strongly overlaps with that used in previous studies. In contrast to the results reported by Greene *et al.* (2004), we found that utilitarian judgments in such dilemmas were associated with activation in the right mid insula, lateral OFC, right VLPFC, rACC, right SII and left superior temporal lobe (Figure 6). Furthermore, region-of interest analyses of the previously reported locations in the DLPFC and parietal lobe (Greene *et al.*, 2004) revealed no significant result. This divergence from previously reported findings is not entirely unexpected given that we used only a selection of previously used dilemmas that were controlled for intuitiveness and content (Supplementary Data), and given that behavioural studies of 'personal' dilemmas that used better controlled stimuli (Greene *et al.*, 2008; Moore *et al.*, 2008) failed to fully replicate the behavioral findings reported in Greene *et al.*, 2001, 2004. In addition, our analyses show that the neural differences observed between utilitarian and deontological judgments in DI dilemmas were almost entirely due to differences in intuitiveness rather than content, in line with our hypothesis. Our findings thus suggest that even in the context of the extreme moral dilemmas previously studied, the neural activations associated with utilitarian judgments might be due to their counter-intuitiveness, not their content.

The neural bases of intuitive and counterintuitive moral judgments

Although recent research has established a key role to intuition in moral judgment (Haidt, 2001), the biological underpinnings of moral intuitions, and of moral judgments that go against intuition, have not yet been previously studied. Our findings shed light on the neural processes that underlie such judgments, and provide partial support for the hypothesized association between intuitive judgment and

automatic processing, and between counterintuitive judgment and controlled processing.

Despite substantial differences in content between the different types of dilemmas, similar patterns of neural activation were observed for intuitive compared to counterintuitive judgment, and for the reverse comparison, within each category of dilemma, suggesting that common neural processes underlie intuitive and counterintuitive judgments regardless of content. This is a significant finding given that different types of moral scenarios are likely to elicit different kinds of intuitions or emotional responses, and it cannot be assumed *a priori* that common neural processes would underlie moral intuitions in different contexts.

Intuitive moral judgments

Judgments were classified as intuitive if chosen by a large majority of independent judges who reported their immediate, unreflective moral response ('Materials and Methods' section and Supplementary Results). It is likely that such judgments were driven by moral intuitions—immediate focused responses disposing people to make certain kinds of moral judgments (Haidt, 2001; Hauser, 2006). However, although intuitive judgments were easier to make than counterintuitive ones, as we predicted, they were not associated with shorter RTs. This last finding is in line with recent studies which failed to replicate the previously reported RT differences in moral judgment when better controlled stimuli were used (Moore *et al.*, 2008) or found such differences only in the context of cognitive load (Greene *et al.*, 2008).

It is currently under debate whether affective processes play a key role in intuitive moral judgments (Haidt, 2001; Hauser, 2006; Valdesolo and Steno, 2006; Koenigs *et al.*, 2007; Moll and de Oliveira-Souza, 2007; Greene, 2008; Huebner *et al.*, 2009). If intuitive judgments are driven by affective responses, they should be associated with increased activation in emotion-related brain areas such as amygdala, OFC, nucleus accumbens or ventromedial prefrontal cortex (Sanfey and Chang, 2008). However, we found no increased activation in these areas during intuitive moral judgments (Figure 3A), not even within DI dilemmas, the category of dilemmas that strongly overlaps with the dilemmas previously studied.

Instead, a heightened signal level was observed in the visual and left premotor cortex (Figures 3 and 6). Since this result was unexpected, further research is needed to clarify the function of these regions in intuitive judgments. One possibility is that this activation reflects greater imaginative and empathetic engagement with the dilemmas, in line with evidence showing that the visual cortex is activated not only during visual perception but also during visual imagery (O'Craven and Kanwisher, 2000; Lambert *et al.*, 2004), and that this activation correlates with the vividness of the imagery (Cui *et al.*, 2007). Premotor cortex activation has been associated with emotional empathy (Nummenmaa *et al.*, 2008) and with empathy as a form of 'emotional perspective-taking' (Lamm *et al.*, 2007). Since these areas were not previously noted as central to moral cognition (Moll *et al.*, 2005), it seems plausible that the observed neural activity reflects not the processes directly underlying intuitive judgments, but the processes that trigger them by making aspects of a moral situation more salient. Our findings thus indicate a possible role for affective processing in triggering intuitive moral judgments, but they do not provide direct support for the view that intuitive moral judgments are generally based in emotion.

Counterintuitive moral judgment

At the neural level, high-level controlled processes such as problem-solving and planning have consistently been shown to engage the ACC and dorsolateral prefrontal cortex as well as the posterior parietal cortex (Sanfey and Chang, 2008). Of these structures, only the ACC was significantly activated during counterintuitive moral judgment (Figure 3B). However, the ACC activation found in the present study was mainly located in the rostral subdivision of the ACC whereas activations related to controlled cognitive processing are commonly found more dorsally (Beckmann *et al.*, 2009).

There are several possible roles that the rACC might play in generating counterintuitive judgment. The rACC has been implicated in the calculation of costs and benefits (Rudebeck *et al.*, 2006; Walton *et al.*, 2007) and emotional conflict resolution (Etkin *et al.*, 2006). The rACC activation we observed might reflect the conflict experienced when subjects overcame affect-laden intuitions (Greene *et al.*, 2004). However, activation in the rACC gyrus matching the cluster found here has recently been shown to reflect the valuation of social information in primates (Rudebeck *et al.*, 2006) and humans (Behrens *et al.*, 2008). Increased rACC activation has been tied to representation of what others think about us (Amodio and Frith, 2006), as well as to guilt, an emotional response to the belief that one has violated moral standards (Zahn *et al.*, 2009). Thus it is also possible that participants were aware that their choice goes against the socially dominant moral view and could be perceived negatively by others.

The association between counterintuitive judgments and greater perceived difficulty and rACC activation partially supports the hypothesis that counterintuitive judgments involve controlled processing. However, given that counterintuitive judgments were not associated with longer RT or activation in areas implicated in higher-level deliberative processing, it remains unclear whether they involve conscious moral reasoning.

Prior dual process models of moral judgment have presented controlled processing as generating utilitarian moral conclusions through explicit reasoning, and that these counterintuitive conclusions then overcome more intuitive deontological responses (Greene *et al.*, 2004). Since in UI dilemmas the counterintuitive conclusion was deontological

in content, it is unlikely that controlled processing in moral decision-making is generally associated with utilitarian reasoning. Instead, in UI dilemmas such processing might involve the application of explicit moral rules (e.g. 'do not lie'). It is also possible, however, that controlled processing does not generate novel moral responses through explicit reasoning but instead resolves conflict between pre-existing competing moral intuitions, for example, by deciding between concern for others' welfare and an aversion to lying.

Deontological and utilitarian judgments are specific to a moral context

As reported above, our findings do not support a general association between deontological judgments and automatic processing, and between utilitarian judgments and controlled processing. In line with this, we did not find significant shared activations between utilitarian judgments across categories. However, we did find that activation in the PCC and TPJ was associated with deontological compared to utilitarian judgments when these were pooled across DI and UI dilemmas (Figure 4, Supplementary Table S3). Such activation was also the only judgment-specific brain activation that could not be explained by intuitiveness. This finding is especially interesting given that the dilemmas we used involved a range of different duties, ranging from constraints on killing to duties concerning promising and fairness. PCC activation has previously been observed in moral processing (Greene *et al.*, 2001, 2004; Moll *et al.*, 2002), and implicated in autobiographical memory recall and self-relevant emotional processing (Summerfield *et al.*, 2009) and in tasks which require adopting the first-person perspective (Vogeley *et al.*, 2004). The TPJ has been implicated in theory of mind tasks (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005), and in tasks involving self-awareness and agency (Decety and Lamm, 2007). Although scenarios concerning lying in UI dilemmas are likely to have engaged theory of mind capacities, it is not likely that they drive this effect given that deontological judgments within UI dilemmas were not associated with greater TPJ activation compared to utilitarian ones (Supplementary Table S10). Although the observed TPJ activity might nevertheless reflect the central role of intention in determining permissibility in deontological ethics, the association between deontological judgments and increased activation in PCC and TPJ is also intriguingly in line with an influential understanding of such judgments as involving concern with one's own agency and its emotional significance (Williams, 1973), suggesting a possible connection between deontological judgment and affective processing. However, further research is needed to clarify the role of the PCC and TPJ in deontological judgment, and to determine whether they are also implicated in other forms of deontological judgment.

Importantly, a tendency to utilitarian or deontological judgments within one category (DI or UI dilemmas) did not correlate with such a tendency in the other, suggesting that the moral judgments of non-philosophers are not based in explicit moral theories such as utilitarianism or Kantian ethics (Kahane and Shackel, 2010). Rather, they appear to be case-dependent, so that individuals can be strongly disposed to make utilitarian judgments in one type of moral context but not in another. This suggestion is supported by studies of patients with VMPFC lesions which show an abnormally frequent tendency to make utilitarian judgments in personal dilemmas (Koenigs *et al.*, 2007), but also an abnormally frequent tendency to make vindictive responses in the Ultimatum game (Koenigs and Tranel, 2007), arguably an abnormal deontological response (Moll and de Oliveira-Souza, 2007; Kahane and Shackel, 2010).

CONCLUSION

A central strand of research into moral decision-making has focused on dilemmas involving extreme life and death situations. On the basis of fMRI studies of such dilemmas, a general account of the neural mechanisms underlying utilitarian and deontological moral judgments has been proposed (Greene *et al.*, 2004, 2008). By using a wider range of dilemmas and controlling for the distinct contribution of content and intuitiveness, we obtained evidence suggesting that behavioural and neural differences in responses to such dilemmas are largely due to differences in intuitiveness, not to general differences between utilitarian and deontological judgment. Our findings suggest that the distinction between intuitive and counterintuitive judgments is a more fundamental division in moral decision-making, and thus highlight the importance of distinguishing the processes generally implicated in intuitive and counterintuitive moral judgments from the content of such judgments in particular contexts. Indeed, a better understanding of the processes that generate common intuitive responses to moral situations, and of the capacities that nevertheless allow some individuals to arrive at highly counterintuitive conclusions, could shed new light on the sources of pervasive forms of moral consensus and disagreement.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

REFERENCES

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

Beckmann, M., Johansen-Berg, H., Rushworth, M.F. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *Journal of Neuroscience*, *29*, 1175–90.

Behrens, T.E., Hunt, L.T., Woolrich, M.W., Rushworth, M.F. (2008). Associative learning of social value. *Nature*, *456*, 245–9.

Cui, X., Jeter, C.B., Yang, D., Montague, P.R., Eagleman, D.M. (2007). Vividness of mental imagery: individual variability can be measured objectively. *Vision Reseach*, *47*(4), 474–8.

Cushman, F., Young, L., Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Sciences*, *17*(12), 1082–9.

Decety, J., Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist*, *13*(6), 580–93.

Etkin, A., Egner, T., Peraza, D.M., Kandel, E.R., Hirsch, J. (2006). Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, *51*, 871–82.

Evans, A.C., Collins, D.L., Millst, S.R., Brown, E.D., Kelly, R.L., Peters, T.M. (1993). 3D statistical neuroanatomical models form 305 MRI volumes. *Proceedings of IEEE-Nuclear Science Symposium and Medical Imaging*, *1*, 1813–7.

Friston, K.J., Ashburner, J., Frith, C.D., Poline, J.B., Heather, J.D., Frackowiak, R.S.J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *2*, 1–25.

Greene, J.D. (2008). The secret joke of Kant's soul. In: Sinnott-Armstrong, W., editor. *Moral Psychology: The Neuroscience of Morality*. Cambridge: MIT Press, pp. 35–79.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–54.

Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–8.

Hauser, M. (2006). *Moral Minds*. New York: Harper Collins.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Reviews*, *108*(4), 814–34.

Huebner, B., Dwyer, S., Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6.

Kamm, F.M. (2000). Nonconsequentialism. In: Hugh, L., editor. *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell.

Kahane, G., Shackel, N. (2008). Do abnormal responses show utilitarian bias? *Nature*, *452*(7185), E5.

Kahane, G., Shackel, N. (2010). Methodological issues in the neuroscience of moral judgment. *Mind and Language*, *25*(5), 561–82.

Kant, I. (1797/1966). On a supposed right to lie from philanthropy. In: Gregor, M.J., editor. *Practical Philosophy*. Cambridge: Cambridge University Press, pp. 611–15.

Koenigs, M., Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *Journal of Neuroscience*, *27*, 951–5.

Koenigs, M., Liane, Y., Ralph, A., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908–11.

Lambert, S., Sampaio, E., Mauss, Y., Scheiber, C. (2004). Blindness and brain plasticity: contribution of mental imagery? An fMRI study. *Brain Research - Cognitive Brain Research*, *20*(1), 1–11.

Lamm, C., Batson, C.D., Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, *19*(1), 42–58.

Moll, J., de Oliveira-Souza, R., Eslinger, P.J., et al. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, *22*(7), 2730–6.

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.

Moll, J., de Oliveira-Souza, R. (2007). Moral judgments, emotions, and the utilitarian brain. *Trends in Cognitive Sciences*, *11*, 319–21.

Moore, A.B., Clark, B.A., Kane, M.J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Sciences*, *19*, 549–57.

Nummenmaa, L., Hirvonen, J., Parkkola, R., Hietanen, J.K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *NeuroImage*, *43*(3), 571–80.

O'Craven, K.M., Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–23.

Ross, W.D. (1930/2002). *The Right and the Good*. Oxford: Oxford University Press.

Rudebeck, P.H., Buckley, M.J., Walton, M.E., Rushworth, M.F. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science*, *3*(5791), 1310–2.

Sanfey, A.G., Chang, L.J. (2008). Multiple systems in decision making. *Annals of the NY Academy of Science*, *1128*, 53–62.

Saxe, R., Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *NeuroImage*, *19*(4), 1835–42.

Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–9.

Singer, P. (2005). Ethics and intuitions. *Journal of Ethics*, *9*, 331–52.

Summerfield, J.J., Hassabis, D., Maguire, E.A. (2009). Cortical midline involvement in autobiographical memory. *NeuroImage*, *44*(3), 1188–200.

Valdesolo, P., DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*, 476–7.

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., Fink, G.R. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience*, *16*(5), 817–27.

Walton, M.E., Rudebeck, P.H., Bannerman, D.M., Rushworth, M.F. (2007). Calculating the cost of acting in frontal cortex. *Annals of the NY Academy of Science*, *1104*, 340–56.

Williams, B.A.O. (1973). A critique of utilitarianism. In: Smart, J.J.C., Williams, B.A.O., editors. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.

Zahn, R., Moll, J., Paiva, M., et al. (2009). The neural basis of human social values: evidence from functional MRI. *Cerebral Cortex*, *19*, 276–83.