NIH Public Access
Author Manuscript
Soc Neurosci. Author manuscript; available in PMC 2011 September 21.

Published in final edited form as:
Soc Neurosci. 2011 August ; 6(4): 398–413. doi:10.1080/17470919.2011.559363.

Neural basis of moral verdict and moral deliberation

Jana Schaich Borg[1], **Walter Sinnott-Armstrong**[2], **Vince D. Calhoun**[3,4], and **Kent A. Kiehl**[4,5]

[1]Neuroscience Institute, Stanford University School of Medicine, Stanford, CA, USA

[2]Department of Philosophy and Kenan Institute for Ethics, Duke University, Durham, NC, USA

[3]Departments of Electrical Engineering, Computer Science, and Neurosciences, University of New Mexico, Albuquerque, NM, USA

[4]The Mind Research Network, Albuquerque, NM, USA

[5]Departments of Psychology and Neurosciences, University of New Mexico, Albuquerque, NM, USA

Abstract

How people judge something to be morally right or wrong is a fundamental question of both the sciences and the humanities. Here we aim to identify the neural processes that underlie the specific conclusion that something is morally wrong. To do this, we introduce a novel distinction between “moral deliberation,” or the weighing of moral considerations, and the formation of a “moral verdict,” or the commitment to one moral conclusion. We predict and identify hemodynamic activity in the bilateral anterior insula and basal ganglia that correlates with committing to the moral verdict “this is morally wrong” as opposed to “this is morally not wrong,” a finding that is consistent with research from economic decision-making. Using comparisons of deliberation-locked vs. verdict-locked analyses, we also demonstrate that hemodynamic activity in high-level cortical regions previously implicated in morality—including the ventromedial prefrontal cortex, posterior cingulate, and temporoparietal junction—correlates primarily with moral deliberation as opposed to moral verdicts. These findings provide new insights into what types of processes comprise the enterprise of moral judgment, and in doing so point to a framework for resolving why some clinical patients, including psychopaths, may have intact moral judgment but impaired moral behavior.

Keywords

Morality; Judgment; Anterior insula; Ventromedial prefrontal cortex

Moral judgments are fundamental to human interaction. Whether the goal is to explain social decision-making (Hsu, Anen, & Quartz, 2008), to assess legal culpability (Aharoni, Funk, Sinnott-Armstrong, & Gazzaniga, 2008), or to determine whether morality is based on emotion or reason (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), diverse disciplines converge in asking: What in the brain makes us judge something to be morally wrong?

Here, we introduce a distinction between moral deliberation and moral verdicts. People can detect and deliberate about moral considerations without reaching moral verdicts in the same

© 2011 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

Correspondence should be addressed to: Kent A. Kiehl, The Mind Research Network, 1101 Yale Boulevard, Albuquerque, NM 87131, USA. kkiehl@unm.edu.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

way that juries can observe and deliberate about the evidence in a criminal trial without yet arriving at a verdict about whether or not a defendant is guilty. This analogy illustrates that the neural processes underlying deliberation about a moral issue should be dissociable from the neural processes underlying the commitment to a verdict about a moral issue. This functional magnetic resonance imaging (fMRI) study took some first steps toward investigating whether moral deliberation and moral verdicts are indeed correlated with distinct brain processes.

"Moral deliberation" will be defined as the detection, filtering, and weighing (consciously or unconsciously) of relevant moral principles, heuristics, or concepts that identify morally relevant features and thereby create a "moral context." "Moral verdicts," in contrast, will be defined as (conscious or unconscious) valenced opinions or commitments about what is morally wrong or not wrong, or what one morally ought to do or not to do. The detection, filtering, and weighing processes that comprise moral deliberation represent the integration of many sources of relevant information (evidence, biases, emotions, etc.) over time. A moral verdict, on the other hand, is a discrete conclusion or choice based on interpretation of, or deliberation over, the moral context. "Moral deliberation" and "moral verdict" are similar to the concepts of "decision variable" and "choice" used in fields of perceptual decision-making (Gold & Shadlen, 2007).

To separate these components of a moral judgment, we asked participants to judge as "wrong" or "not wrong" acts that most people take to be immoral (e.g., murdering your friend, stealing money, lying, breaking promises), acts that most people take to be not immoral (e.g., giving to charity, working hard, teaching, recycling), and acts that most people take to be morally controversial (e.g., genetically engineered food, same-sex marriage, abortion, euthanasia). Morally controversial acts were designed to reference more moral concepts and evoke deeper moral deliberation than non-controversial acts, and thus were expected to invoke deeper moral processing and/or deliberation across participants while simultaneously eliciting varied moral verdicts. Taking advantage of this, in response to both controversial and non-controversial items, we compared the cases when a given individual arrived at a verdict that an act was wrong to the cases when that same individual arrived at a verdict that an act was morally not wrong. Moral deliberation and moral verdict, therefore, were manipulated independently.

From previous brain-imaging studies of perceptual and economic decision-making, we hypothesized that verdicts that either controversial or non-controversial acts are morally wrong would correlate with activity in regions of the anterior insula and subcortical basal ganglia. Activity in the anterior insula correlates with rejection of unfair offers (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003), rejection of inequitable allocations (Hsu et al., 2008), decisions not to donate to charity (Moll et al., 2006), decisions not to purchase in a shopping task (Knutson, Rick, Wimmer, Prelec, & Loewenstein, 2007), and verdicts of disbelief (Harris, Sheth, & Cohen, 2008). Moreover, participants taught to increase their left anterior insula activity by biofeedback report increased negative ratings of aversive pictures (Caria, Sitaram, Veit, Begliomini, & Birbaumer, 2010). Since the anterior insula is involved in reporting negative verdicts in economic, social, and emotional contexts, it was a favored candidate region for our hypotheses about regions involved in making negative verdicts in moral contexts. The caudate (a member of the basal ganglia) is active when punishing unfair economic partners (de Quervain et al., 2004), making the basal ganglia a set of promising candidate regions as well. Importantly, we hypothesized that these correlations between activity in the anterior insula and basal ganglia and verdicts that an act is morally wrong would persist regardless of whether the acts were controversial or not controversial, meaning regardless of whether the acts reference more or less moral principles or required more or less moral deliberation. We further hypothesized that the most robust correlations between

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

brain activity and moral verdicts would be in the brain activity immediately preceding a reported judgment, as opposed to the brain activity after a stimulus presentation when the verdict was not yet likely to be reached.

Next, since controversial acts should reference more moral principles and might require more moral deliberation than non-controversial acts, morally controversial stimuli should be more correlated than non-controversial stimuli with activity in brain regions involved in general moral processing. Based on previous studies of moral processing (Eslinger et al., 2009; Finger, Marsh, Kamel, Mitchell, & Blair, 2006; Greene, Nystrom, Engell, & Darley, 2004; Greene et al., 2001; Harenski & Hamann, 2006;Hauke R. Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Moll et al., 2002, 2005; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Schaich Borg, Lieberman, & Kiehl, 2008), we hypothesized that activity in the ventromedial prefrontal cortex (vmPFC), posterior cingulate, and temporoparietal junction (TPJ) would be greater in response to controversial acts than non-controversial acts. Importantly for the distinction between moral deliberation and moral verdict, we predicted that activity in these regions would correlate more with whether an act was controversial or non-controversial than with the verdict that the act was wrong or not wrong. Unlike our predictions for the anterior insula and basal ganglia, we predicted that the most robust correlations between brain activity in the vmPFC, posterior cingulate, and TPJ and moral deliberation would be during the time immediately following a stimulus presentation as opposed to when a verdict was being reported. This is because moral deliberation would predominate before the verdict was reached but would be mostly absent once a verdict was confident enough to be reported.

Here we present data supporting the hypotheses described above about (1) the neural systems mediating a moral verdict that an act is morally wrong, and (2) dissociating moral verdict from moral deliberation. Then we discuss how distinguishing moral deliberation from moral verdicts can account for the previously unknown role of the anterior insula and basal ganglia in moral judgments, can provide a framework for explaining discrepant clinical observations made of antisocial patients, and might impact the legal and philosophical debates over which clinical patients are culpable for their actions and how “emotion” or “reason” contribute to moral decision-making.

MATERIALS AND METHODS

Subjects

Subjects were recruited via advertisements. Twenty-six healthy participants (mean age 32, *SD* = 10.85) provided written, informed, IRB-approved consent at Hartford Hospital and Yale University and were compensated $20/h for their participation. All participants were right-handed on self-report and were able to perform the task successfully during practice sessions prior to scanning.

Task

Participants were shown 50 short statements describing acts that most people judge to be immoral, 50 statements describing acts that most people judge to be not immoral, and 50 statements describing acts that most people judge to be morally controversial. (These statements are provided in Supplementary Information 7.) After presentation of a stimulus, participants pressed one button to indicate if they thought an act was “wrong” and another button to indicate if they thought an act was “not wrong.” Participants had a maximum of 10 s to respond. Participants’ responses were immediately followed by a jittered 1–6-s black screen prior to the onset of the next trial. Stimuli were randomized and counter-balanced

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

across three runs. The length of each run varied across sessions and participants because stimuli presentation was self-paced.

Behavioral ratings

After scanning, participants were asked to rate each experimental stimulus for difficulty (1 = extremely easy, 7 = extremely difficult), emotional valence (1 = extremely unpleasant, 7 = extremely pleasant), and moral content (1 = no moral content, 7 = extreme moral content). An additional 40 participants were asked to rate each experimental stimulus for arousal (1 = not at all emotionally worked up, 7 = extremely emotionally worked up), moral wrongness (1 = not at all morally wrong, 7 = extremely morally wrong), and how likely other people would be to agree with the participant’s judgment (1 = no one would agree, 7 = everyone would agree).

Data acquisition

Participants viewed all experimental stimuli via a mirror on top of the head coil that reflected a screen at the rear entrance of the magnet bore. Stimuli were displayed on the screen, using a computer-controlled projection system and E-prime software (Psychological Software Tools, Pittsburgh, PA, USA). Data were collected in three separate runs on a Siemens Allegra 3T head-dedicated scanner equipped with 40 mT/m gradients and a standard quadrature head coil at the Olin Neuropsychiatry Research Center at the Institute of Living, Hartford, Connecticut. The functional scans were acquired with gradient-echo echo-planar-imaging (scanning parameters: repeat time (TR) = 1.50 s, echo time (TE) = 27 ms, field of view = 24 cm, acquisition matrix = 64 × 64, flip angle = 70°, voxel size = 3.75×3.75×4 mm, gap = 1 mm, 29 slices, ascending acquisition). Six “dummy” scans were performed at the beginning of each functional run to allow for longitudinal equilibrium, and they were discarded before image analysis.

Preprocessing

All preprocessing was performed with Statistical Parametric Mapping 2 (SPM2, Wellcome Department of Imaging Neuroscience). Each of the three sessions was separately realigned to the first scan of the session, using INRIalign (Freire, Roche, & Mangin, 2002). Translation and rotation corrections did not exceed 3 mm and 2.5°, respectively, for any of the participants, and corrections did not exceed 1 mm or 1° for 19 out of 26 participants. After realignment, a mean EPI image was computed for each run and was subsequently matched to the SPM2 EPI template. Data were transformed into standard Montreal Neurological Institute (MNI) space, and this transformation was then applied to all functional images. Data were then spatially smoothed with a 9-mm full width at half-maximum Gaussian kernel, and submitted to a fifth-order infinite impulse response Butterworth low-pass filter of 0.16 Hz to remove any high-frequency noise.

Individual subject statistics

Four conditions of interest were modeled for each subject’s individual responses: *non-controversial wrong*, *non-controversial not wrong*, *controversial wrong*, and *controversial not wrong*. Whether a particular response was classified in the wrong group or the not-wrong group was determined by each individual participant’s response to a stimulus.

Events were time-locked to response onsets, stimulus onsets, or stimuli onsets modeled by a variable boxcar function defined by the duration between the stimulus onset and the response. Regressors were created by convolving the appropriate stimulus function and its temporal derivative with the canonical hemodynamic response. First-order motion parameters obtained from realignment were included as confounds in each participant’s

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

model to remove possible residual task-related motion effects, and a high-pass filter (cutoff period, 128 s) was incorporated to remove noise associated with low-frequency confounds.

Group level statistics

Random effects analyses—The amplitude of the hemodynamic response used in second-level analyses was calculated from condition-specific nonderivative and derivative terms from each subject's first-level analysis (Calhoun, Stevens, Pearlson, & Kiehl, 2004). Images containing these amplitudes were entered into second-level, one-sample *t*-tests for the following primary contrasts: (1) *controversial* (*wrong* + *not wrong*) > *non-controversial* (*wrong* + *not wrong*), (2) *wrong* (*controversial* + *non-controversial*) > not-wrong (*controversial* + *non-controversial*), (3) *non-controversial wrong* > *non-controversial not wrong*, and (4) *controversial wrong* > *controversial not wrong*. Results for contrasts 1–3 are reported if they withstood a false-discovery rate multiple-comparisons correction of at least $p = .01$. Fifty trials per participant were included in contrast 4, *controversial wrong* > *controversial not wrong*, as opposed to the 100 trials included in the *non-controversial wrong* > *non-controversial not wrong* contrast or the 150 trials included in the *controversial* > *non-controversial* contrast. Therefore, only voxels observed in contrasts 1–2 were strongly interpreted in contrasts 3–4, and the threshold criteria for contrast 4 was $p < .01$ uncorrected due to its reduced number of trials.

Data extraction for effect size analyses—*t*-tests allow valid statistical inference about whether or not a null hypothesis can be rejected, but they do not by themselves allow inference about how much different factors contribute to the observed signal. We wanted to explore how much the wrong vs. not-wrong factor, the controversial vs. non-controversial factor, or their interaction could explain the variance in specific voxels of interest. To do this, we identified voxels in our a priori regions of interest (ROIs) that surpassed FDR $p < .01$ multiple-comparisons corrections in one of the whole-brain group *t*-tests. Then, to visualize the relative amounts of variance accounted for by different factors in the present experiment, 6-mm radius (12-mm diameter) spherical ROIs were constructed around the most significant voxel in these a priori regions that surpassed multiple-comparisons corrections. Data were extracted from unsmoothed images, using the Marsbar tool-box (http://marsbar.sourceforge.net/MarsBaR) for SPM2, and transformed into *z*-scores, using the mean of the entire time series for each run. Variance associated with movement parameters was removed, and values were transformed into *z*-scores again (in case any variance had changed). These preprocessed data were then used to calculate the generalized eta squared (Olejnik & Algina, 2003) associated with the controversial vs. non-controversial variable, the wrong vs. not-wrong variable, and their interaction. Effect sizes for all experimental conditions were calculated on the collective activity recorded 3, 4, 5, 6, 7, and 8 TRs after judgments, chosen because they traversed the expected lag of the maximal hemodynamic response to the judgment (Friston, Jezzard, & Turner, 1994).

Note on effect size calculations—The absolute values of our calculated effect sizes will not be comparable to those reported by future studies because (1) signal-to-noise will differ greatly across experiments and scanning centers, and (2) the absolute value of the effect is potentially statistically biased given that some of our effect sizes are being calculated from data that are not fully independent of the data used to identify the most significant voxel (Lieberman, Berkman, & Wager, 2009; Vul, Harris, Winkielman, & Pashler, 2009). However, most importantly for their use here, these eta squared values provide reliable estimates of how much variance was accounted for by the *controversial* vs. *non-controversial* condition, the *wrong* vs. *not wrong* condition, or their interaction in the present experiment, and thus they were used as a supplementary guide for inferring how much a given region was involved in moral deliberation as opposed to making moral

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

verdicts. These effect sizes are informative, because (1) the voxels under examination were chosen by a valid inferential procedure (a *t*-test constrained by a valid multiple-comparisons correction), and (2) visualizing the data this way helps support and explain the results of our *t*-tests in ways that refine predictions for future studies (Kriegeskorte, Lindquist, Nichols, Poldrack, & Vul, 2010). We clarify that unless effect sizes are being calculated in voxels that were identified by an independent group-level *t*-test, effect size visualizations are meant to supplement our group-level *t*-tests, not add extra inferential information. Now that the present study has identified which previously unknown brain regions should be involved in negative moral verdicts, future studies with independent participants can be used to calculate purely unbiased effect sizes for all experimental factors simultaneously to support the valid inferential *t*-tests reported in this study.

RESULTS

Behavior

As predicted, behavioral ratings provided by participants directly after scanning indicated tha, on average, participants rated by self-report the 50 controversial acts (*controversial*) as significantly more difficult to judge (on a scale from 1 to 7, 3.06 for *controversial*, 2.16 for *non-controversial, p* < .0001) than the 100 non-controversial acts (*non-controversial*). Likewise, the group of 40 extra participants reported that people would be significantly less likely to agree with their responses to *controversial* acts than to *non-controversial* acts (on a scale from 1 to 7, for 3.52 for *controversial*, 2.96 for *non-controversial*, *p* < .0001). Scanning participants took the same amount of time to respond to *controversial* acts as compared to *non-controversial* acts (*p* < .233), but slightly longer to judge an act to be *not-wrong* than *wrong* (*p* < .0001, on average 0.16 s more for *controversial* and 0.21 s more for *non-controversial* with a significant interaction of *p* < .0001). The 50 non-controversial acts people judged to be morally wrong (*non-controversial wrong*) were rated as significantly more unpleasant (mean score of 2.10 on a scale from 1 to 7 where 1 = extremely unpleasant and 7 = extremely pleasant) than *controversial* acts (mean score of 3.82), and *controversial* acts were rated as significantly more unpleasant than the 50 non-controversial acts people judged to be not wrong (*non-controversial not wrong*, mean score of 5.83). However, *non-controversial* acts grouped together (including both *wrong* and *not wrong* acts) were not rated as significantly more unpleasant than *controversial* acts (*p* = .61). This reflects the fact that participants judged most *non-controversial wrong* items to be “wrong” and to have negative moral valence, judged most *non-controversial not wrong* items to be “not wrong” and to have positive moral valence, and had varied responses to controversial items which, together, averaged out to have equal moral valence to non-controversial items. All items were rated as having significant moral content by the scanning participants.

fMRI

Random effects analyses were first run with events time-locked to participants’ responses to maximize the chances of detecting activity correlated with moral verdicts. With the exception of the *controversial* > *non-controversial* contrast (see Methods), all results reported here withstood a false-discovery rate, multiple-comparisons correction of at least *p* = .01. *t*-values are provided in referenced Supplementary Information tables.

No significant voxels withstood multiple-comparisons correction when the interaction between controversial vs. non-controversial and wrong vs. not wrong was tested, so the wrong > not wrong results will be discussed primarily as a main effect across both controversial and non-controversial items instead of separately for the controversial items and non-controversial items separately. Some non-significant interactions appeared in the simple effects, however (Supplementary Information1). (“Non-significant interactions”

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

refers to voxels with activity that differentially surpassed multiple-comparisons (FDR corrected $p < .01$) thresholds in the *non-controversial wrong* > *non-controversial not wrong* compared to the *controversial wrong* > *controversial not wrong* simple effect, or the *wrong* > *not wrong* main effect contrast, but that nonetheless failed to surpass the same multiple-comparisons and cluster thresholds for the *(non-controversial wrong – non-controversial not wrong)* > *(controversial wrong – controversial not wrong)* or *(controversial wrong – controversial not wrong)* > *(non-controversial wrong – non-controversial not wrong)* interaction tests.) These non-significant interactions are listed when relevant to a priori hypotheses or when highly notable. When simple effects are not described for an anatomic area, there were no notable differences between the simple effects and the main effects, meaning that brain areas involved in judging an item to be wrong were consistent regardless of whether the item was controversial or non-controversial.

Group-level *t*-tests of moral verdicts: wrong > not wrong and not wrong > wrong

When the brain activity correlated with judging items to be wrong was compared to that associated with judging items to be not wrong across non-controversial and controversial items (*wrong* > *not wrong*), using group-level voxel-wise *t*-tests, robust activity, as predicted, was found in the bilateral anterior insula extending into the inferior frontal gyri and temporal poles and the bilateral basal ganglia, including the globus pallidus, putamen, and caudate head (Figure 1, table in Supplementary Information 1). Activity was also observed in the bilateral amygdala in the *non-controversial wrong* > *non-controversial not wrong* simple effect (Figure 1). These results were consistent across subjects (Supplementary Information 2).

Outside our explicitly predicted areas, highly significant activity was identified in the left lingual gyrus (Brodmann area (BA 19)) extending into the left cuneus (BA 17) in the *wrong* > *not wrong* contrast. Further visual system activity was identified in the left middle and right inferior occipital gyri and left fusiform gyrus (all of which extended bilaterally in the *non-controversial wrong* > *non-controversial not wrong* simple effect). Significant clusters were also identified in the left middle temporal gyrus (BA 21), bilateral middle frontal gyrus (BA 6/8), left superior frontal gyrus (BA 10, right in the *controversial wrong* > *controversial not wrong* simple effect), dorsomedial superior frontal gyrus (BA 9/8 overlapping anterior BA 6, superior BA 10, and anterior cingulate gyrus, BA 32), bilateral hippocampus (extending posteriorally into the parahippocampal gyrus on the left, unilaterally on the right in the *non-controversial wrong* > *non-controversial not wrong* simple effect), bilateral thalamus, and brainstem. Clusters in the bilateral angular gyrus close to the TPJ (BA 39; 67 voxels on the left, 117 voxels on the right) remained significant even after correction for multiple comparisons. These clusters were much smaller and unilaterally in the left hemisphere in the *non-controversial wrong* > *non-controversial not wrong* simple effect and did not withstand multiple-comparisons corrections in the *controversial wrong* > *controversial not wrong* simple effect (although a unilaterally right cluster of $t = 3.25$ and 41 voxels was identified). Importantly, no statistically significant differences in activity were detected in the vmPFC or the posterior cingulate.

Despite their differences in moral valence, no voxels withstood multiple-comparisons correction when the brain activity associated with items judged to be not wrong was compared to that associated with items judged to be wrong (*not wrong* > *wrong*).

In sum, directly comparing moral verdict outcomes facilitated the identification of a wide network of cortical and subcortical structures previously unknown to be involved in negative moral verdicts, including our hypothesized regions of the anterior insula and basal ganglia. As predicted, this network did not appear to include the vmPFC or the posterior cingulate, two areas commonly associated with moral processing in previous studies.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Group-level *t*-tests of moral deliberation: controversial > non-controversial and non-controversial > controversial

When the brain activity associated with controversial items was compared to the brain activity associated with non-controversial items (*controversial* > *non-controversial*, Figure 2, table in Supplementary Information 3), unique and extensive bilateral brain activity was associated with controversial items relative to non-controversial items in superior and posterior regions around the TPJ traversing the supra-marginal gyrus and angular gyrus (BA 39/40). Furthermore, significant activity was observed in the vmPFC and the posterior cingulate. Figure 2 shows that this contrast yielded activity in coordinates reported by previous fMRI studies of morality (see Greene et al. (2001, 2004), Harenski et al. (2006), and Heekeren et al. (2005) for some particularly close matches). These results were consistent (with the exception of statistically non-significant laterality differences in the frontal pole) regardless of whether the controversial items were judged to be wrong or not-wrong (see *controversial wrong* > *non-controversial wrong* and *controversial not wrong* > *non-controversial not wrong*, Supplementary Information 4): There was no significant interaction between controversial vs. non-controversial and wrong vs. not wrong. In addition, activity was observed in the anterior insula, basal ganglia, and left amygdala, similar to the *wrong* > *not wrong* contrast (*t*-scores are compared in table in Supplementary Information 3).

No voxels withstood multiple-comparisons correction when the brain activity associated with non-controversial items was compared to that associated with controversial items.

To summarize, activity in many of the brain regions identified in previous studies of morality— including in particular the vmPFC, posterior cingulate, and areas around the TPJ — correlated with different levels of moral deliberation, but not with the verdict of the deliberation process. In addition, activity in the anterior insula and basal ganglia correlated with comparing levels of moral deliberation in the *controversial* > *non-controversial* contrast as well as with commitment to the moral verdict that an act was wrong in the *wrong* > *not wrong* contrast.

ROI analyses

Due to the overlap in regions identified in the *wrong* > *not wrong* contrast and the *controversial* > *non-controversial* contrast, we quantified how much variance could be accounted for by each of these variables in the regions for which we made a priori hypotheses: the anterior insula and the basal ganglia. We also included the amygdala and the three regions most consistently identified in previous studies of morality: the vmPFC, posterior cingulate, and areas around the TPJ. To achieve this, data were extracted from 6-mm spherical ROIs constructed around the most significant voxel of selected contrasts in these selected anatomic regions (see Supplementary Information 5 for coordinates and Methods for details), and effect sizes across the ROIs were calculated with the generalized eta squared statistic (Olejnik & Algina, 2003). As discussed in the Methods section, this eta squared statistic provided a reliable estimate of how much variance was accounted for by the *controversial* vs. *non-controversial* condition, the *wrong* vs. *not wrong* condition, or their interaction within the present experiment, but they should not all be compared to effect sizes reported in future experiments.

Consistent with the overlap in the group-level *t*-maps but not explicitly addressed in our predictions, considerable variance was accounted for by both the *wrong* > *not-wrong* main effect and the *controversial* > *non-controversial* main effect in the bilateral anterior insula and basal ganglia. More variance was accounted for by the controversial > non-controversial main effect than the wrong > not-wrong main effect, suggesting that activity in these regions

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

may have coded for moral deliberation as well as for moral verdict. Average activity in the insula varied across the cluster but generally increased well above baseline when stimuli were controversial, increased more so when they were judged to be wrong, increased less above baseline when non-controversial items were judged to be wrong, and remained at baseline when they were judged to be not wrong. Average activity in the basal ganglia had a similar pattern, but often dropped below baseline when non-controversial items were judged to be not wrong (Supplementary Information 5). Thus, our predictions that the anterior insula and basal ganglia should be involved in committing to moral verdicts were supported, but, in addition, these brain regions have supplementary roles in moral deliberation.

Outside our explicitly predicted regions, we calculated the effect size of each experimental condition in the left lingual gyrus and left cuneus due to their provocative results in the group-level *t*-tests. Consistent with the *t*-tests, much more variance was accounted for by the main effect of *wrong* > *not-wrong* than *controversial* > *non-controversial*, and the variance accounted for by the main effect of *wrong* > *not-wrong* was greater in these two brain regions than any other region tested. Activity in these regions increased equally above baseline when an item was judged to be wrong, regardless of whether it was controversial or non-controversial, and increased slightly above baseline when an item judged to be wrong was controversial (Supplementary Information 5).

Intriguing results were also found in brain regions implicated in prior studies of morality. Consistent with the results of the group-level *t*-tests, almost all of the variance accounted for by our model in the signal detected from the vmPFC, the posterior cingulate, and bilateral regions around the TPJ was accounted for by the main effect of *controversial* > *non-controversial* contrast. Little or no variance was accounted for by the main effect of the *wrong* > *not-wrong* contrast or the controversial⁄wrong interaction (Figure 3) in these regions. These effect sizes can also be appreciated by plotting the average condition-specific activity across these regions (Supplementary Information 5). In all three regions, average activity increased above baseline only when stimuli were controversial. Although activity in the left and right TPJ was slightly higher when judging a non-controversial act to be wrong than when judging a non-controversial act to be not wrong, the higher level reflected a smaller decrease from baseline rather than a larger increase. This was true for more dorsal regions of dorsomedial BA 10 and medial BA 9 (which overlapped with regions previously published in the functional imaging literature; i.e., Greene et al., 2001, 2004) as well.

Group-level *t*-tests: comparison of event-locked and stimulus-locked models

To test whether the dissociation between moral deliberation-related activity and moral verdict-related activity was specific to the activity immediately preceding the report of a verdict, group-level *t*-tests were repeated, using (1) stimulus-locked events (modeled with no specified durations) rather than response-locked events, and (2) stimulus-locked events with specified durations modeled as variable length, boxcar functions (encompassing all the activity beginning with the stimulus onset and ending with the response). Insula activity remained correlated with judging an act to be wrong vs. not wrong in these analyses, but to a lesser extent than in the response-locked analysis, while activity in the amygdala and basal ganglia decreased when events were time-locked to stimuli (modeled with no specified durations) and disappeared altogether when all the averaged stimulus-related activity preceding and leading up to a decision was included in the model as a variable length boxcar. Again, no activity was identified in the vmPFC, posterior cingulate, and TPJ in the *wrong* > *not-wrong* contrast (Figure 4 and Supplementary Information 6). This indicates that the lack of verdict-correlated activity in these areas is not likely to be explained by the type of event or how much time around an event is included in the analysis. It also indicates that verdict-correlated activity is strongest immediately preceding the commitment to a moral

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

verdict, rather than earlier when deliberation is still underway and a verdict is not yet likely to have been reached.

Unlike verdict-correlated activity, activity in the vmPFC, posterior cingulate, and TPJ increased and became more significant in the *controversial* > *non-controversial* contrasts when events were stimulus-locked, as opposed to response-locked, to an even greater degree when durations were specified by boxcar functions. These changes illustrate that task-related activity in these brain regions was stronger during moral deliberation than during commitment to a moral verdict, and was relatively homogeneous over the course of the deliberation such that averaging over longer periods of time in the box-car designs increased statistical power rather than decreased it (in contrast to verdict-related activity, which was less significant when a box-car design was used).

DISCUSSION

This study was designed to identify the neural processes that underlie the specific conclusion that something is morally wrong. It is often intuited that high-level cortical brain regions should be responsible for making moral judgments, given the complexity of morality and its uniqueness to the human species. Yet, by applying a model of moral judgment that distinguishes moral verdicts from moral deliberation, the present study demonstrates that the bilateral anterior insula and subcortical regions, including the basal ganglia, are particularly strong candidates for the regions responsible for verdicts about what is morally wrong. Activity in these brain regions correlates with judging acts to be wrong across controversial and non-controversial scenarios, and correlates most strongly immediately preceding a negative moral judgment rather than during the deliberation leading up to the judgment.

The anterior insula, in particular, is not often discussed in the context of moral judgment, but results presented here are consistent with our hypotheses and consistent with evidence that the anterior insula and basal ganglia are active during subjective feelings of aversion in other contexts (as reviewed in the opening paragraphs of this paper). Consistent with their structure deep in the brain and given their participation in negative judgments in many contexts, the role of the anterior insula and basal ganglia in judging an act to be morally wrong likely represents a general role for these regions in encoding negative valence and avoidance of aversive stimuli rather than a unique role in contributing to negative moral verdicts. It is possible that one of the functions of the anterior insula in negative moral verdicts is to generate or regulate arousal. Similar to a previously published result from a study that asked people to rate concepts as “bad” or “good” (Cunningham, Raye, & Johnson, 2004), a separate set of 40 people rated the acts judged to be morally “wrong” in the present study as eliciting significantly greater emotional intensity than the acts judged to be morally “not wrong” ($p < .001$ for non-controversial acts, 5.33 for *non-controversial wrong* on a 7-point scale from “extremely calm/not at all emotional” to “extremely “worked up”/emotional,” 4.30 for *non-controversial not wrong; p* < .001 for controversial acts, 4.75 for *controversial wrong*, 3.09 for *controversial not wrong*). However, a couple of things call into question whether arousal is the only role of the insula in moral verdicts. First, activity in the anterior insula correlated with valence in the previous study asking people to rate concepts as “bad” or “good,” not arousal (Cunningham et al., 2004). Activity in the amygdala, on the other hand, correlated with arousal. Furthermore, even if activity in the anterior insula did correlate with arousal, it may have correlated specifically with negative arousal rather than positive arousal (Lewis, Critchley, Rotshtein, & Dolan, 2007), suggesting that its function might still be specific to negative moral verdicts as opposed to other verdicts. Given these issues, additional experiments will be needed to explore the functional specificity of the anterior insula and the basal ganglia in moral judgments, especially given that our analyses suggest that these brain regions are involved with moral deliberation as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

well as with committing to moral verdicts. In the meantime, understanding the role that these regions play in making moral verdicts might give critical, previously unavailable insight into what systems to tap into to modify and improve moral decision-making and behavior.

We focused on the role of the anterior insula and the basal ganglia in this study because, as reviewed in the beginning of this paper, previous research showing the role of these regions in rejection of unfair offers or allocations (Sanfey et al., 2003) and decisions not to donate to charity or purchase items in a shopping task (Knutson et al., 2007; Moll et al., 2006) made them particularly strong a priori candidates for brain regions likely to be involved in negative moral verdicts. This emphasis was supported further by the established role of these regions in empathy (Decety & Lamm, 2006). However, in addition to the anterior insula and basal ganglia, our results indicate that the left lingual gyrus and cuneus of the visual cortex may play an unappreciated role in making moral judgments. These regions were reliably more active when acts were judged to be wrong than when acts were judged to be not wrong for both controversial and non-controversial stimuli, and, intriguingly, more variance was accounted for by the *wrong* > *not-wrong* main effect in the left lingual gyrus and cuneus than any other regions tested, including the anterior insula and basal ganglia. Since we did not track eye movement, it cannot be ruled out that this activity represents task-unrelated visual activity (i.e., making random eye movements while judging an act to be morally wrong), but this possibility is unlikely given the regions we identified were so strongly lateralized. Furthermore, the left visual cortex is often identified in fMRI studies of social emotion (Critchley, Elliott, Mathias, & Dolan, 2000; Graff-Guerrero et al., 2008; Stoeter et al., 2007; Taylor et al., 1998; Völlm et al., 2006). Thus, our results suggest that the role of the left lingual gyrus and left cuneus in social rejection judgments should be replicated and studied more thoroughly in the future.

On the other hand, activity in the vmPFC, posterior cingulate, and cortex around the TPJ did *not* correlate with judging an act to be morally wrong in the present study, and activity in these regions increased above baseline only when acts were controversial. Accordingly, activity in these regions correlated with moral deliberation. The vmPFC, posterior cingulate, and cortex around the TPJ have reliably been implicated in studies of morality (Finger et al., 2006; Greene et al., 2001, 2004; Harenski & Hamann, 2006;Hauke R. Heekeren et al., 2003;H. R. Heekeren et al., 2005; Moll et al., 2002, 2005; Robertson et al., 2007; Schaich Borg et al., 2006, 2008). It has been hypothesized that this network might be involved in morality because of its role in emotional self-referential processing (Harrison et al., 2008) or representations of others’ beliefs (Van Overwalle, 2009), both of which might be preferentially engaged by morally controversial items if they inspire references to one’s own personal moral beliefs or comparisons to what others are likely to think about the controversial item. However, with one exception (Greene et al., 2004), all previous studies have examined moral vs. non-moral deliberation or deliberation in different kinds of moral conflicts. Since these studies did not examine different moral judgments of “wrong” as opposed to “not wrong,” they did not account for moral verdicts in their analyses. Furthermore, most previous studies of morality time-locked their statistical analyses to the presentation of a moral stimulus rather than the report of a moral judgment or verdict. Our comparison of statistical models suggests that this may have made it difficult to detect signal correlated with moral verdicts. Therefore, the results presented here offer a significant refinement in how previous results might be interpreted. Strong evidence is provided that the activity detected in the vmPFC, posterior cingulate, and cortex around the TPJ in past studies likely represented their role in the accumulation of morally relevant information (by mediating the detection, filtering, or weighing of relevant moral principles, heuristics, or concepts, for example, which might include many of the cognitive processes previously shown to contribute to moral decision-making such as references to self, cognitive control,

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

or representations of theory of mind) more than a role in committing to a specific moral verdict. In other words, the vmPFC, posterior cingulate, and cortex around the TPJ might represent or calculate the factors that contribute to moral deliberation, but they do not appear to be the final arbiters of the commitment to the specific verdict that the act is morally wrong.

Like the results from the moral verdict analyses, it is possible that the results from the moral deliberation analyses are also mediated by arousal. However, this explanation is unlikely because, although 40 independent participants rated the controversial and non-controversial items as eliciting significantly different arousal, controversial items were actually rated as *less* arousing than non-controversial items ($p < .001$; 3.60 on a 7-point scale for *controversial* items, 4.82 for *non-controversial* items). Especially given that activity in no brain regions passed multiple-comparisons corrections in the non-controversial > controversial contrast, it is not likely that arousal explains the differences observed in the moral deliberation analyses.

Also of note, despite their differences in moral valence, we did not find activity that correlated more with the verdict “this is morally not wrong” than the verdict “this is morally wrong.” The *non-controversial not wrong* stimuli we used were rated about as pleasant in valence as the *non-controversial wrong* stimuli were rated unpleasant, but there were still no voxels that withstood multiple-comparisons corrections in the simple or main effects of the *not wrong* > *wrong* contrasts. It is possible that the *not wrong* stimuli evoked moral neutrality or ambivalence, which might be unlikely to recruit additional or unique brain regions compared to stimuli that elicited *wrong* moral verdicts. However, previous evidence calls that explanation in question. Like us, Cunningham et al. (2004) found positive correlations between brain activity in the anterior insula and ratings of “badness” of social concepts, but did not find any positive correlations between brain activity and “goodness” of social concepts. Thus, separate neural systems might be used to encode negative moral verdicts compared to positive moral verdicts. More research will be needed to uncover the neural basis of positive moral verdicts, especially those representing moral obligation.

The results reported here and the distinction between moral deliberation and moral verdict are useful for many reasons. First, the distinction between moral deliberation and moral verdict provides vocabulary to start breaking down moral judgment into its component cognitive parts, and this, in turn, helps make predictions for future studies. Any processes that ostensibly push final moral judgments in a particular direction, regardless of when or how, should be of interest. However, it will be useful to know when and at what part of the decision these processes are most likely to have their influence. As the field of neuroeconomics has shown, understanding the details of how the component neural decision-making processes interact can dramatically improve our ability to understand and predict human behavior. Given that many of the robust, confusing behavioral phenomena studied in economic decision-making also manifest themselves in moral judgment and decisions (Kahneman, 1994; Shenhav & Greene, 2010), there is good reason to be optimistic that refining our understanding of the neural mechanisms underlying moral judgment will also improve our ability to understand, predict, and modify moral behavior.

Second, one area that can greatly benefit from a refined understanding of the distinct mechanisms underlying different parts of moral judgment is clinical psychiatry. Our results support a model of moral decision-making that can account for how previous brain-imaging results can be reconciled with two clinical observations. (1) Patients with prefrontal cortex damage can and do make moral judgments, and *most* of the moral judgments they report are normal (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007;Michael Koenigs et al., 2007; Mendez & Anderson, 2005; Saver & Damasio, 1991), even if their judgments are sometimes

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

inconsistent with their behavior (Saver & Damasio, 1991). (2) Adult psychopaths, who are hypothesized to have impaired vmPFC function (M. Koenigs, Kruepke, & Newman, 2010), also report normal judgments about what is morally wrong or not wrong (Blair & Cipolotti, 2000; Cima, Tonnaer, & Hauser, 2010; Damasio, 1994; Saver & Damasio, 1991). By distinguishing moral processing from moral verdict, our data clarify that the vmPFC, posterior cingulate, and TPJ do not likely mediate the final commitment to the moral verdict “This is wrong.” Instead, they likely contribute the outcome of the moral verdict in a more distal, indirect fashion, by identifying and interpreting factors that need to be integrated before the processes that read or translate the verdict are applied. This suggests that vmPFC patients and psychopaths should have mostly normal moral processing and should be able to integrate available information to commit to moral alternatives as long as normal, successful accumulation of morally relevant information is not vmPFC-, TPJ-, or posterior cingulated-dependent and their anterior insula and basal ganglia functionality is intact. Indeed, this is what has been shown in the case of lesion patients. Thus, differentiating moral processing from moral judgment provides a framework that can account for the counterintuitive observation that antisocial patients often have intact moral judgment.

In addition, the data described here support new hypotheses about why psychopaths and vmPFC patients have abnormal moral behavior despite their ability to integrate many types of moral information. One hypothesis highlighted by the results of the present study is that perhaps psychopaths and vmPFC patients have deficits in how they commit to moral verdicts at the end of their otherwise fairly normal moral deliberation. For example, they might have deficits in the motivation that should be associated with their moral verdict, and that motivation might be mediated by some of the brain regions shown to be involved in moral verdicts in the present study. If true, this hypothesis clarifies the possibility that psychopathic symptoms might be alleviated through cognitive therapy using prefrontal-cortex-independent methods to attach significance and motivation to things judged to be wrong (Fellows, 2006) rather than through therapy that targets the process patients use to arrive at that judgment.” Another previously proposed hypothesis is that vmPFC and/or anterior cingulate dysfunction in psychopaths and vmPFC patients (vmPFC lesions often extend into the anterior cingulate) lead to a more general deficit in adapting all kinds of behavior—including social behavior— to the expected value of its outcome (Rangel & Hare, 2010), consistent with the propensity of psychopaths and vmPFC patients to behave suboptimally in many arenas in life (Damasio, 1994; Kiehl, 2006). If this second hypothesis is true, the data presented here suggest the roles of the vmPFC in adapting behavior would be in addition to (and perhaps later in time than) the roles of this region in moral processing. Future research will be needed to determine whether either of these hypotheses is accurate, but differentiating moral processing from moral judgment provides a framework that both adds to and refines these hypotheses about moral behavior in populations that have thus far proven very difficult to treat.

Third, the distinction between moral deliberation and moral verdict provides interesting twists in the theoretical debates over whether reason or emotion causes moral judgment and whether lesion patients should be held legally culpable for their actions. These questions were purposely not addressed here, but activity in the vmPFC is usually cited as evidence that “emotion” is critical for judging something to be morally wrong (Greene et al., 2001;Michael Koenigs et al., 2007; Young & Koenigs, 2007) and has already been described in court as “the seat” of moral judgments and moral culpability (*State* vs. *Stanko*, South Carolina Supreme Court, 2008). Our results suggest that this specific type of activity may be important for creating the emotional (or cognitive) context for moral deliberation but is likely not the most proximate cause of the final moral verdict that something is “wrong” or “right.” Other types of emotion mediated by the insula and basal ganglia, such as disgust, might be more likely to contribute more proximately to moral verdicts that something is

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

morally wrong, consistent with reports that manipulations of these emotions also manipulate moral judgments (Jones & Fitness, 2008; Wheatley & Haidt, 2005).

The framework presented here is just one preliminary step in breaking down moral judgment. It is not yet clear which factors or computations comprise deliberation in moral judgments, for example, although some promising headway has been made (Shenhav & Greene, 2010). Importantly, though, the present results indicating that moral deliberation is distinguishable from moral verdict raise significant questions for future research. In particular, given the evidence that moral emotions affect moral judgment and contribute to moral behavior (Tangney, Stuewig, & Mashek, 2007), do moral emotions like guilt, anger, or disgust contribute more to moral deliberation or to moral verdicts? And do different emotions interact with moral deliberation and moral verdicts in different ways, sometimes acting as a cause and other times acting as a response? Questions like these provide a useful framework to help integrate the psychology of morality and the neuroscience of moral judgments in directions that generate new hypotheses and provide new insights into how and why humans are able to judge acts to be morally right or wrong.

References

Aharoni E, Funk C, Sinnott-Armstrong W, Gazzaniga M. Can neurological evidence help courts assess criminal responsibility? Lessons from law and neuroscience. Annals of the New York Academy of Sciences. 2008; 1124:145–160. [PubMed: 18400929]

Blair RJR, Cipolotti L. Impaired social response reversal: A case of ‘acquired sociopathy’. Brain. 2000; 123(6):1122–1141. [PubMed: 10825352]

Calhoun VD, Stevens MS, Pearlson GD, Kiehl KA. fMRI analysis with the general linear model: Removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. NeuroImage. 2004; 22:252–257. [PubMed: 15110015]

Caria A, Sitaram R, Veit R, Begliomini C, Birbaumer N. Volitional control of anterior insula activity modulates the response to aversive stimuli. A real-time functional magnetic resonance imaging study [doi: 10.1016/j.biopsych.2010.04.020]. Biological Psychiatry. 2010; 68(5):425–432. [PubMed: 20570245]

Ciaramelli E, Muccioli M, Ladavas E, di Pellegrino G. Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. Social Cognitive and Affective Neuroscience. 2007; 2(2):84–92. [PubMed: 18985127]

Cima M, Tonnaer F, Hauser MD. Psychopaths know right from wrong but don’t care. Social Cognitive and Affective Neuroscience. 2010

Critchley HD, Elliott R, Mathias CJ, Dolan RJ. Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. Journal of Neuroscience. 2000; 20(8):3033–3040. [PubMed: 10751455]

Cunningham WA, Raye CL, Johnson MK. Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. Journal of Cognitive Neuroscience. 2004; 16(10):1717–1729. [PubMed: 15701224]

Damasio, A. Descartes’ error. New York, NY: Grosset/Putnam; 1994.

de Quervain DJF, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, et al. The neural basis of altruistic punishment. Science. 2004; 305(5688):1254–1258. [PubMed: 15333831]

Decety J, Lamm C. Human empathy through the lens of social neuroscience. Scientific World Journal. 2006; 6:1146–1163. [PubMed: 16998603]

Eslinger PJ, Robinson-Long M, Realmuto J, Moll J, deOliveira-Souza R, Tovar-Moll F, et al. Developmental frontal lobe imaging in moral judgment: Arthur Benton’s enduring influence 60 years later. Journal of Clinical and Experimental Neuropsychology. 2009; 31(2):158–169. [PubMed: 19048446]

Fellows LK. Deciding how to decide: Ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. Brain. 2006; 129(4):944–952. [PubMed: 16455794]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Finger EC, Marsh AA, Kamel N, Mitchell DG, Blair JR. Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. NeuroImage. 2006; 33(1):414–421. [PubMed: 16891125]

Freire L, Roche A, Mangin JF. What is the best similarity measure for motion correction in fMRI time series? IEEE Transactions on Medical Imaging. 2002; 21(5):470–484. [PubMed: 12071618]

Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. Human Brain Mapping. 1994; 1(2):153–171.

Gold JI, Shadlen MN. The neural basis of decision making. Annual Review of Neuroscience. 2007; 30(1):535–574.

Graff-Guerrero A, Pellicer F, Mendoza-Espinosa Y, Martínez-Medina P, Romero-Romo J, de la Fuente-Sandoval C. Cerebral blood flow changes associated with experimental pain stimulation in patients with major depression. Journal of Affective Disorders. 2008; 107(1–3):161–168. [PubMed: 17904643]

Greene JD, Nystrom LE, Engell AD, Darley JM. The neural bases of cognitive conflict and control in moral judgment. Neuron. 2004; 44:389–400. [PubMed: 15473975]

Greene JD, Sommerville R, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. Science. 2001; 293(5537):2105–2108. [PubMed: 11557895]

Harenski CL, Hamann S. Neural correlates of regulating negative emotions related to moral violations. NeuroImage. 2006; 30(1):313–324. [PubMed: 16249098]

Harris S, Sheth SA, Cohen MS. Functional neuroimaging of belief, disbelief, and uncertainty. Annals of Neurology. 2008; 63(2):141–147. [PubMed: 18072236]

Harrison BJ, Pujol J, Lopez-Sola M, Hernandez-Ribas R, Deus J, Ortiz H, et al. Consistency and functional specialization in the default mode brain network. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(28):9781–9786. [PubMed: 18621692]

Heekeren HR, Wartenburger I, Schmidt H, Prehn K, Schwintowski HP, Villringer A. Influence of bodily harm on neural correlates of semantic and moral decision-making. NeuroImage. 2005; 24(3):887–897. [PubMed: 15652323]

Heekeren HR, Wartenburger I, Schmidt H, Schwintowski HP, Villringer A. An fMRI study of simple ethical decision-making. NeuroReport. 2003; 14(9):1215–1219. [PubMed: 12824762]

Hsu M, Anen C, Quartz SR. The right and the good: Distributive justice and neural encoding of equity and efficiency. Science. 2008; 320(5879):1092–1095. [PubMed: 18467558]

Jones A, Fitness J. Moral hypervigilance: The influence of disgust sensitivity in the moral domain. Emotion. 2008; 8(5):613–627. [PubMed: 18837611]

Kahneman, D. The cognitive psychology of consequences and moral intuition. Paper presented at the the Tanner Lecture in Human Values; Ann Arbor: University of Michigan; 1994.

Kiehl KA. A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. Psychiatry Research. 2006; 142(2–3):107–128. [PubMed: 16712954]

Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G. Neural predictors of purchases. Neuron. 2007; 53(1):147–156. [PubMed: 17196537]

Koenigs M, Kruepke M, Newman JP. Economic decision-making in psychopathy: A comparison with ventromedial prefrontal lesion patients. Neuropsychologia. 2010; 48(7):2198–2204. [PubMed: 20403367]

Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, et al. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature. 2007; 446:908–911. [PubMed: 17377536]

Kriegeskorte N, Lindquist MA, Nichols TE, Poldrack RA, Vul E. Everything you never wanted to know about circular analysis, but were afraid to ask. Journal of Cerebral Blood Flow and Metabolism. 2010; 30(9):1551–1557. [PubMed: 20571517]

Lewis PA, Critchley HD, Rotshtein P, Dolan RJ. Neural correlates of processing valence and arousal in affective words. Cerebral Cortex. 2007; 17(3):742–748. [PubMed: 16699082]

Lieberman MD, Berkman ET, Wager TD. Correlations in social neuroscience aren't voodoo: Commentary on Vul et al. (2009). Perspectives on Psychological Science. 2009; 4(3):299–307.

Mendez M, Anderson E. An investigation of moral judgement in frontotemporal dementia. Cognitive and Behavioral Neurology. 2005; 18(4):193–197. [PubMed: 16340391]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Moll J, de Oliviera-Souza R, Eslinger PJ, Bramati IE, Mourao-Miranda J, Andreiuolo PA, et al. The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. Journal of Neuroscience. 2002; 22(7):2730–2736. [PubMed: 11923438]

Moll J, de Oliveira-Souza R, Moll FT, Ignacio FA, Bramati IE, Caparelli-Daquer EM, et al. The moral affiliations of disgust: A functional MRI study. Cognitive and Behavioral Neurology. 2005; 18(1): 68–78. [PubMed: 15761278]

Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J. Human fronto-mesolimbic networks guide decisions about charitable donation. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(42):15623–15628. [PubMed: 17030808]

Olejnik S, Algina J. Generalized eta and omega squared statistics: Measures of effect size for some common research designs. Psychological Methods. 2003; 8(4):434–447. [PubMed: 14664681]

Rangel A, Hare T. Neural computations associated with goal-directed choice. Current Opinion in Neurobiology. 2010; 20(2):262–270. [PubMed: 20338744]

Robertson D, Snarey J, Ousley O, Harenski K, Bowman F, Gilkey R, et al. The neural processing of moral sensitivity to issues of justice and care. Neuropsychologia. 2007; 45(4):755–766. [PubMed: 17174987]

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. The neural basis of economic decision-making in the ultimatum game. Science. 2003; 300(5626):1755–1758. [PubMed: 12805551]

Saver JL, Damasio AR. Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. Neuropsychologia. 1991; 29(12):1241–1249. [PubMed: 1791934]

Schaich Borg J, Hynes C, Van Horn J, Grafton S, Sinnott-Armstrong W. Consequences, action, and intention as factors in moral judgments: An fMRI investigation. Journal of Cognitive Neuroscience. 2006; 18(5):803–817. [PubMed: 16768379]

Schaich Borg J, Lieberman D, Kiehl KA. Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. Journal of Cognitive Neuroscience. 2008; 20(9):1–19. [PubMed: 17919082]

Shenhav A, Greene JD. Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. [doi: 10.1016/j.neuron.2010.07.020]. Neuron. 2010; 67(4):667–677. [PubMed: 20797542]

Stoeter P, Bauermann T, Nickel R, Corluka L, Gawehn J, Vucurevic G, et al. Cerebral activation in patients with somatoform pain disorder exposed to pain and stress: An fMRI study. NeuroImage. 2007; 36(2):418–430. [PubMed: 17428684]

Tangney JP, Stuewig J, Mashek DJ. Moral emotions and moral behavior. Annual Review of Psychology. 2007; 58(1):345–372.

Taylor SF, Liberzon I, Fig LM, Decker LR, Minoshima S, Koeppe RA. The effect of emotional content on visual recognition memory: A PET activation study. NeuroImage. 1998; 8(2):188–197. [PubMed: 9740761]

Van Overwalle F. Social cognition and the brain: A meta-analysis. Human Brain Mapping. 2009; 30(3):829–858. [PubMed: 18381770]

Völlm BA, Taylor ANW, Richardson P, Corcoran R, Stirling J, McKie S, et al. Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. NeuroImage. 2006; 29(1):90–98. [PubMed: 16122944]

Vul E, Harris C, Winkielman P, Pashler H. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspectives on Psychological Science. 2009; 4(3):274–290.

Wheatley T, Haidt J. Hypnotic disgust makes moral judgments more severe. Psychological Science. 2005; 16(10):780–784. [PubMed: 16181440]

Young L, Koenigs M. Investigating emotion in moral cognition: A review of evidence from functional neuroimaging and neuropsychology. British Medical Bulletin. 2007:69–79. [PubMed: 18029385]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

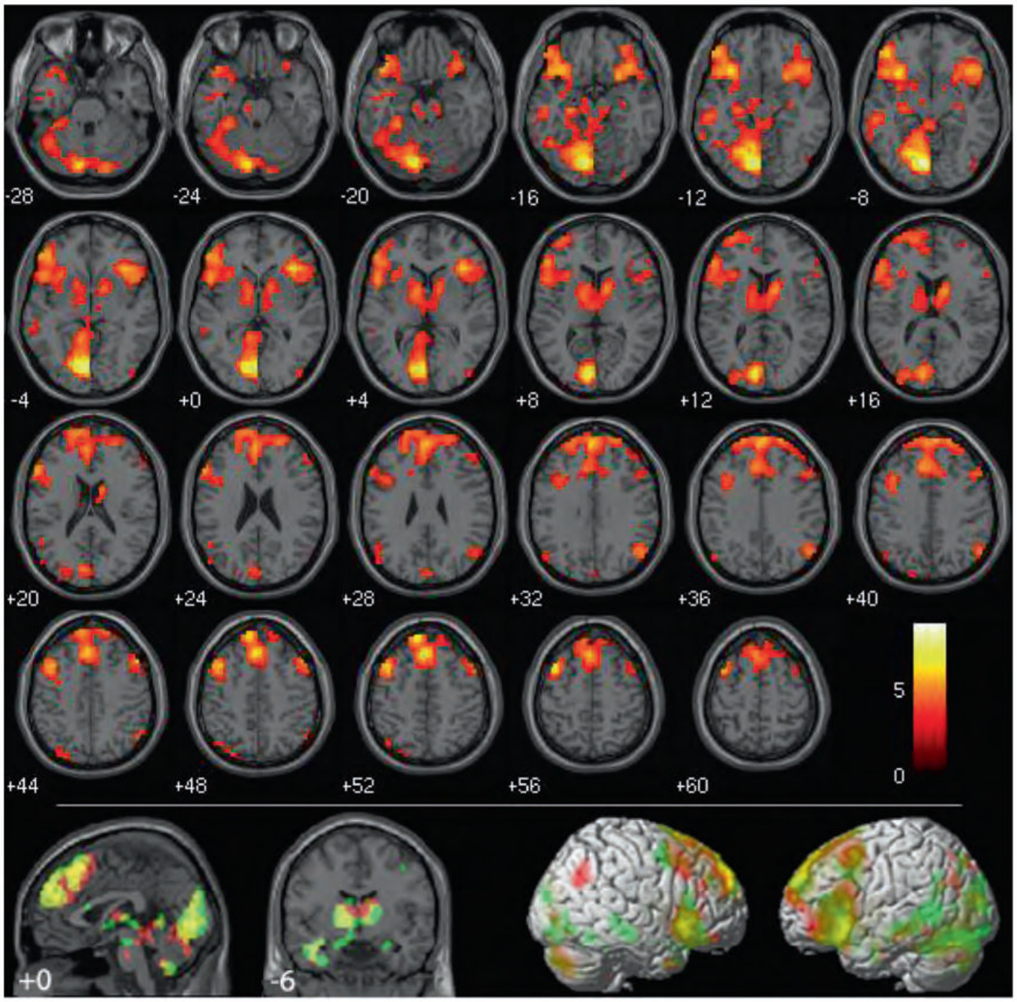

Figure 1.

Wrong (controversial + non-controversial) > not-wrong (controversial + non-controversial). Results are overlaid on SPM2 canonical T1 image, FDR corrected, p < .01, clusters ≥ 10 contiguous voxels. Bottom: wrong (controversial + non-controversial) > not-wrong (controversial + non-controversial) in red overlapped with non-controversial wrong > non-controversial not wrong in green.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

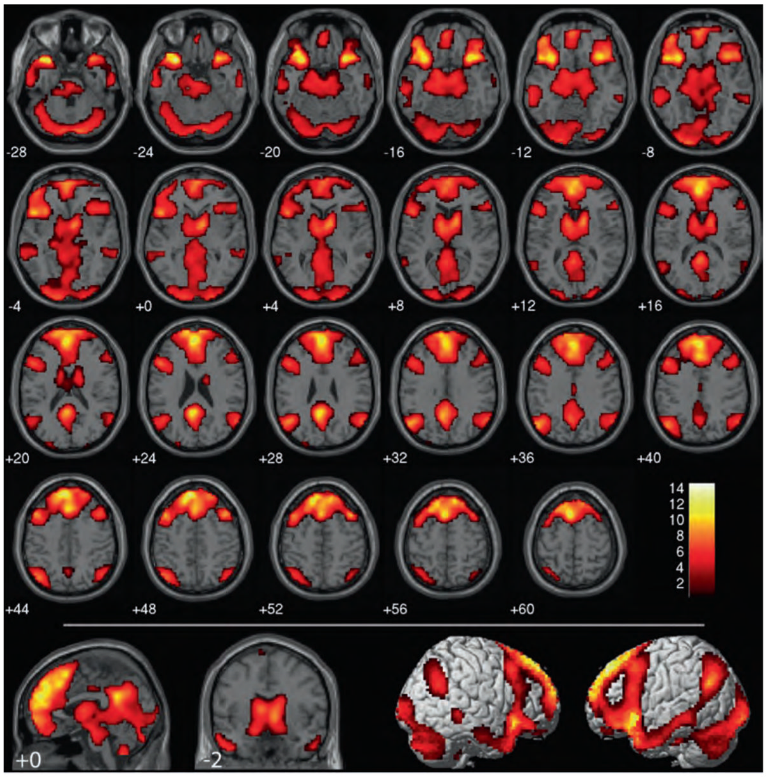

Figure 2. Controversial (wrong + not wrong) > non-controversial (wrong + not wrong)
Results are overlaid on SPM2 canonical T1 image, FDR corrected, $p < .01$, clusters ≥ 10 contiguous voxels.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

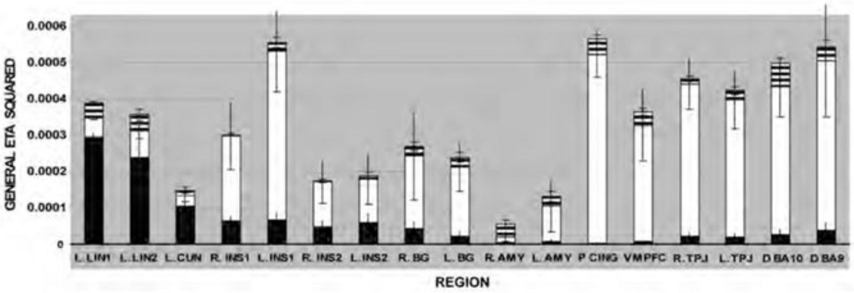

Figure 3. Dissociation in moral deliberation and moral verdict as measured by the generalized eta squared statistic

Error bars represent *SDs*. Black bar: variance associated with wrong vs. not wrong. White bar: variance associated with controversial vs. non-controversial. Striped bar: variance associated with interaction between wrong vs. not wrong and controversial vs. non-controversial.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

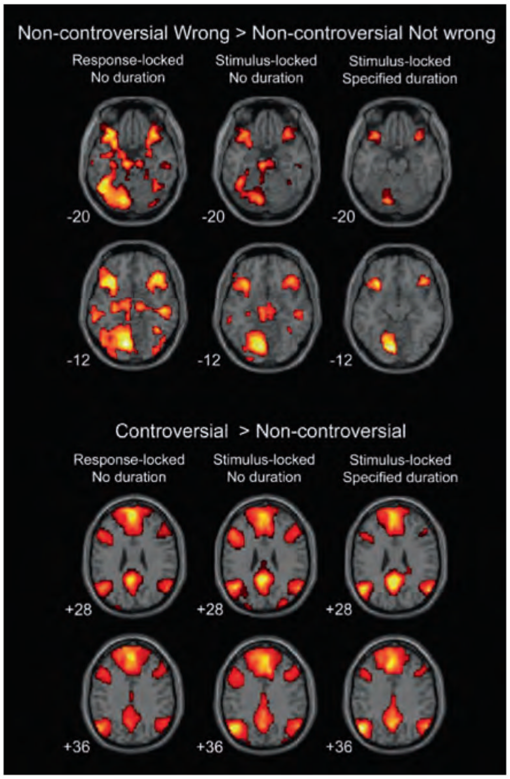
Non-controversial Wrong > Non-controversial Not wrong
Response-locked No duration
Stimulus-locked No duration
Stimulus-locked Specified duration
-20
-20
-20
-12
-12
-12
Controversial > Non-controversial
Response-locked No duration
Stimulus-locked No duration
Stimulus-locked Specified duration
+28
+28
+28
+36
+36
+36

Figure 4.

Effects of different statistical designs on non-controversial wrong > non-controversial not wrong and controversial > non-controversial contrasts. Activity in the amygdala (top row), insula, and basal ganglia (second to top row) decreased when stimulus-locked models were used. Activity in the vmPFC, posterior cingulate (second to bottom row), and areas around the TPJ (bottom row) increased when stimulus-locked models were used.