Evaluating Methods of Correcting for Multiple Comparisons Implemented in SPM12 in Social Neuroscience fMRI Studies: An Example from Moral Psychology

Hyemin Han Andrea L. Glenn

University of Alabama

Author Note

Hyemin Han, Educational Psychology Program, University of Alabama.

Andrea L. Glenn, Center for the Prevention of Youth Behavior Problems, University of Alabama.

Correspondence concerning this manuscript should be addressed to Hyemin Han, Educational Psychology Program, University of Alabama, Box 870231, Tuscaloosa, AL 35487-0231, USA.

Abstract

In fMRI research, the goal of correcting for multiple comparisons is to identify areas of activity that reflect true effects, and thus would be expected to replicate in future studies. Finding an appropriate balance between trying to minimize false positives (Type I error) while not being too stringent and omitting true effects (Type II error) can be challenging. Furthermore, the advantages and disadvantages of these types of errors may differ for different areas of study. In many areas of social neuroscience that involve complex processes and considerable individual differences, such as the study of moral judgment, effects are typically smaller and statistical power weaker, leading to the suggestion that less stringent corrections that allow for more sensitivity may be beneficial, but also result in more false positives. Using moral judgment fMRI data, we evaluated four commonly used methods for multiple comparison correction implemented in SPM12 by examining which method produced the most precise overlap with results from a meta-analysis of relevant studies and with results from nonparametric permutation analyses. We found that voxel-wise thresholding with family-wise error correction based on Random Field Theory provides a more precise overlap (i.e., without omitting too few regions or encompassing too many additional regions) than either clusterwise thresholding, Bonferroni correction, or false discovery rate correction methods.

Evaluating Methods of Correcting for Multiple Comparisons Implemented in SPM12 in Social Neuroscience fMRI Studies: An Example from Moral Psychology

Correcting for multiple comparisons has been one of the most significant challenges in the statistical analysis of fMRI data (Bennett, Miller, & Wolford, 2009). Because more than one hundred thousand voxels are compared simultaneously during analysis, the chances of Type I error are very high in the absence of any correction (Genovese, Lazar, & Nichols, 2002). In order to address this issue, researchers have developed various correction methods. For instance, Bonferroni's correction method, one of the traditional methods for multiple comparison correction, divides the nominal significance level (e.g., $p < .05$) by the number of tests being performed (Bland & Altman, 1995). Although Bonferroni correction produces good control of Type I error, it has the disadvantage of removing both false and true positives when applied to whole brain analyses. To address this issue, many researchers use a family-wise error (FWE) correction method based on Random Field Theory (RFT) (Nichols, 2012). Unlike the traditional Bonferroni method, which only accounts for the total number of comparisons, this method assumes that the error fields can be a lattice approximation to an underlying random field usually with a Gaussian distribution (Brett, Penny, & Kiebel, 2004; Eklund, Nichols, & Knutsson, 2016). Moreover, the false discovery rate (FDR) correction method was developed. This method is thought to be more sensitive and less likely to produce Type II error than FWE correction methods. Unlike the aforementioned methods that control for the possibility of any false positives, this method focuses on the expected proportion of false positives only among survived entities (Genovese et al., 2002; Nichols, 2013). In terms of the implementation of the FDR correction method, neuroimaging has relied on the standard FDR procedure, the linear step-up procedure, or so-called Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995;

Benjamini, Krieger, & Yekutieli, 2006), although more sophisticated procedures, such as adaptive linear step-up procedures, have been developed (Benjamini et al., 2006).

These correction methods can be performed at different levels of inference (i.e., voxel-wise and clusterwise inference; Flandin & Novak (2013) using fMRI analysis software (e.g., SPM) and customized MATLAB codes. In the case of voxel-wise inference, each individual voxel is treated as a unit for analysis, and any voxel exceeding a threshold after applying one of the aforementioned correction methods is considered statistically significant in the whole brain or specified regions of interest (Nichols, 2012). In the case of clusterwise inference, statistically significant clusters showing activation are detected based on the number of contiguous voxels; this type of inference does not control the estimated false positive probability of each individual voxel in each region, but controls such a probability of the region as a whole (Woo, Krishnan, & Wager, 2014). This clusterwise inference has been one of the most popular methods used for multiple comparison correction because it is considered to be more sensitive than voxel-wise inference (Woo et al., 2014).

Although the aforementioned correction methods, particularly RFT FWE correction and FDR correction, have been implemented in widely used fMRI analysis software (e.g., SPM) with parametric assumptions, a nonparametric analysis tool, Statistical non-Parametric Mapping (SnPM) (Nichols & Holmes, 2002), uses permutations in order to correct for multiple comparisons without several assumptions required for parametric analysis, such as normally distributed data and mean parameterization. Instead, SnPM requires several minimal assumptions pertaining to the empirical null hypothesis (Nichols, 2012). For instance, in the case of a two-sample t-test, the subjects are assumed to be exchangeable under the null hypothesis, which might be violated if the subjects are related; in the case of a one-sample t-test, sign flipping,

which is based on an assumption that the errors have a symmetric distribution, is used (Winkler, Ridgway, Webster, Smith, & Nichols, 2014). The application of SnPM is considered less stringent than the Bonferroni and RFT-FWE correction methods applied by software supporting parametric analysis when it is applied using voxel-wise inference (Eklund et al., 2016; Nichols & Hayasaka, 2003); however, clusterwise inference with SnPM is considered to be more stringent (Eklund et al., 2016). Furthermore, the randomise function in FSL and the corresponding function in the BROCCOLI software (Eklund, Dufort, Villani, & LaConte, 2014), which are based on the same statistical principles used for SnPM, have been utilized to evaluate false-positive rates and sensitivity of traditional correction methods (Eklund et al., 2016).

Although clusterwise inference has become a popular method for multiple comparison correction, a recent study evaluating different correction methods has raised concerns that the RFT-applied FWE correction method for clusterwise inference implemented in widely-used fMRI analysis software, such as SPM and FSL, inflates false-positive rates and produces erroneous outcomes (Eklund et al., 2016). RFT clusterwise inference relies on two strong assumptions that might cause such erroneous outcomes. It assumes that "the spatial smoothness of the fMRI signal is constant over the brain" (Eklund et al., 2016, p. 7902), and that there is a specific shape in the spatial autocorrelation function (Eklund et al., 2016). Using resting-state fMRI and cognitive experimental fMRI data analyzed with putative task designs, Eklund et al. (2016) report that rather than a false-positive rate of 5%, the most common software packages (SPM, FSL, AFNI) resulted in false-positive rates of up to 70% when RFT clusterwise inference was performed. However, the RFT-applied voxel-wise inference produced conservative outcomes with a false-positive rate of 5% or less (Eklund et al, 2016). Furthermore, there have been concerns regarding the FDR method as well. Although the FDR correction method reduces

the probability of Type II error with enhanced sensitivity, this method is perhaps more likely to produce Type I error compared to other more conservative correction methods (Bennett, Wolford, & Miller, 2009).

The problem of appropriately correcting for multiple comparisons may be especially concerning in the field of social neuroscience in which balancing the risk of Type I and Type II error is sometimes an especially difficult task. Social and affective fMRI experiments often, though not always, produce smaller effect sizes and have weaker statistical power compared to, for example, sensory-motor experiments; there are several reasons for this (Lieberman & Cunningham, 2009). In many areas of social neuroscience, the psychological processes measured by these experiments are often poorly defined and are not directly observable (Poldrack, 2011). For example, in the field of moral psychology, the process of moral judgment – determining whether something is right or wrong –involves a number of different lower order processes, and is a task that can be accomplished in different ways by different individuals (Blasi & Hoeffel, 1974; Narvaez, Getz, Rest, & Thoma, 1999). Furthermore, the mental states of the individual are often not as certain. We cannot, for example, know with certainty that a person is "experiencing empathy" at a particular moment in time, nor expect the neural correlates of such a reported experience to be the same for all individuals. In contrast, with sensory and motor phenomena, there is a closer mapping between experimental inputs (e.g., a visual cue) and behavioral outputs (e.g., the person taps his fingers) and much less variability from trial-to-trial or person-to-person (Lieberman & Cunningham, 2009).

Although there are some examples of larger effects and more precise study designs within social neuroscience, such as the response of the fusiform face area to images of faces, many social neuroscience studies involve complex paradigms that may allow for multiple mental

processes to occur, and in which the timing of processing is less precise. Thus, it has been argued that attempts to diminish Type I errors may be problematic in many social neuroscience studies, as they increase the probability of missing true effects (Lieberman & Cunningham, 2009). As a result, more lenient correction methods have become common in the field, in order to improve sensitivity. However, given the recent report demonstrating the possibility of inflated false positive rates possibly produced by the application of methods with better sensitivity (e.g., clusterwise inference), researchers should be cautious when applying such methods for correction in social neuroscientific studies.

Because of the unique goals and challenges of social neuroscience research, it is important to evaluate the various methods for multiple correction reviewed above in the context of a social neuroscientific experiment. Previous studies have evaluated different correction methods by using data from various clinical, visual, cognitive, and behavioral experiments (Eklund et al., 2016; Nichols & Hayasaka, 2003; Nichols & Holmes, 2002) and using simulation data (Nichols & Hayasaka, 2003), but not using data collected in a social neuroscience experiment. Because of the issues raised above regarding small effects and reduced statistical power, one of the challenges with many types of social neuroscience research is that it is more difficult to determine which activations are indeed false positives. However, in the context of the more complex and noisier data of some social neuroscience experiments, one way to generate an activation map that may more closely reflect the "true" pattern of activity for a particular process – in order to evaluate these methods – is to use meta-analysis. Meta-analysis has been suggested as a feasible method to enhance the statistical power of fMRI analysis and better examine the overall brain activity pattern associated with a certain functionality of interest by analyzing a

large amount of data while addressing the issue of reverse inference (Eklund et al., 2016; Lieberman & Cunningham, 2009; Poldrack, 2011).

Thus, in the present study we compare the quality of various correction methods in the analysis of data collected in one area of social neuroscience that is faced with the challenges mentioned above – an fMRI study of moral judgment. We used datasets and results from a quantitative meta-analysis of previously published fMRI studies on moral cognition and emotion (Han, 2017) in order to identify which correction method results in activations that most closely resemble results from the meta-analysis. Furthermore, as recommended by Nichols & Holmes (2002), we also compare results from each correction method with results produced by SnPM analyses, which does not require any parametric assumptions. We test four thresholding methods which have been widely used in the field – FDR, thresholding with RFT clusterwise inference, RFT-based FWE voxel-wise thresholding implemented in SPM, and Bonferroni correction.

Methods and Materials

Subjects and Materials

In the present study we reanalyzed previously collected fMRI data (Han, 2016; Han, Chen, Jeong, & Glover, 2016; Han, Glover, & Jeong, 2014). This fMRI data was obtained from 16 participants (8 male) who attended a Northern Californian university. They were undergraduate and graduate students who ranged in age from 21 to 34 years (M = 28.59, SD = 3.18). Functional brain images were acquired while they were making judgments about 60 socio-moral dilemmas that were previously developed for fMRI experiments (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). The dilemma set consisted of three different types of dilemmas: moral-personal (22 dilemmas), moral-impersonal (18 dilemmas), and non-moral dilemmas (20 dilemmas). Moral-personal

dilemmas tend to provoke negative intuitive emotional responses among participants and often involve salient harm to human lives. Moral-impersonal dilemmas involved moral content but did not intend to provoke strong gut-level responses (e.g., should you return a lost wallet). Non-moral dilemmas included simple value-neutral mathematical problem sets. Participants were asked to choose one of two options to address each presented dilemma. Functional brain images were acquired using a spiral in-and-out sequence (TR = 2000ms, TE = 30ms, flip angle = 90) (Glover & Law, 2001). For the functional images, a total of 31 oblique axial slices were scanned parallel to the AC-PC with 4-mm slice thickness, 1-mm inter-slice skip. The resolution was 3.75 x 3.75 mm (FOV = 240mm, 64 x 64 matrix). Similar to Greene et al.'s (2001, 2004) experiments, we modeled brain activity four scans before, one during and three scans after the moment of response.

Procedures

Reanalysis of fMRI data. We reanalyzed the brain images using four different approaches for multiple comparison correction. These approaches were FDR correction (Genovese et al., 2002), clusterwise inference-applied thresholding (Flandin & Novak, 2013), voxel-wise thresholding with FWE correction based on the RFT implemented in SPM12 (Flandin & Novak, 2013; Nichols, 2013), and FWE correction using Bonferroni's method (Nichols & Holmes, 2002). Analyses were performed by using SPM12. The FDR inference was performed in MATLAB software and was based on example code developed by Nichols (2013). Also, a customized MATLAB code was composed to implement voxel-wise thresholding instead of peak-wise analysis that is the default in SPM. For the first-level estimation, a separate general linear model (GLM) was set for each participant that examined neural activity during each of three conditions. Each regressor was convolved with a standard hemodynamic response function

(HRF). For comparisons between conditions, second-level estimation was conducted. The performed comparisons were conducted for these pairs: all moral (moral-personal + moral-impersonal) versus non-moral, moral-personal versus non-moral, moral-impersonal versus non-moral, moral-personal versus moral-impersonal, and moral-impersonal versus moral-personal. We applied four different thresholds for comparison. First, a voxel-wise threshold of $p < .05$ was used after applying one of the three multiple comparison correction methods (i.e., FDR, RFT FWE, Bonferroni's FWE). Second, we also applied an uncorrected voxel-wise threshold of $p < .001$ with a clusterwise threshold of $p < .05$ after FWE correction, which was provided by SPM12. For cross-check, we also conducted the reanalysis with SnPM. Similarly, a voxel-wise threshold of $p < .05$ after applying FWE correction was used and 5,000 permutations were performed.

Meta-analysis of previous fMRI experiment for the basis for evaluation. We evaluated the four correction methods by comparing findings from our analyses of fMRI data to those from a meta-analysis of fMRI studies on moral cognition and emotion (Han, 2017). GingerALE software (version 2.3.6), which implements the activation likelihood estimation (ALE) method (Eickhoff et al., 2009; Eickhoff, Bzdok, Laird, Kurth, & Fox, 2012; Laird, Lancaster, & Fox, 2005), was employed in the meta-analysis. The meta-analysis examined a previously collected set of activation foci that were found by previous neuroimaging studies that compared neural correlates between moral and non-moral task conditions (for details see Han, 2017). This dataset included 45 experiments with 959 participants and 463 activation foci reported by 43 articles (see Table S1 for the list of included articles). In particular, comparisons between overall moral versus non-moral task conditions and moral versus non-moral judgment were performed. For the former comparison the whole dataset was meta-analyzed; for the

comparison between moral versus non-moral judgment, we meta-analyzed a subset of the whole dataset (18 experiments with 373 participants and 142 activation foci reported in 17 articles) that included previous studies focusing on moral judgment among various moral functions. For both meta-analyses, we used FDR of .01 as a cluster-forming threshold and .05 for clusterwise inference as suggested (Fox et al., 2013). The calculated ALE map was compared with results produced by the aforementioned four correction methods for quality evaluation.

Overlap index calculation and quality evaluation. The present study aimed to quantitatively examine the overlap between survived activation foci after the application of each correction method and those found by the meta-analysis, foci in the ALE map created by GingerALE, and SnPM. We may simply represent the degree of overlap with the ratio of the number of overlapped voxels to the total number of voxels of a reference area. In case of the present study, two different types of ratios can be calculated: the ratio of the number of overlapped voxels ($|V_{Ovl}|$) to the number of survived activation foci after applying correction method ($|V_{Cor}|$; $|V_{FDR}|$ in case of FDR correction, $|V_{CLU}|$ in case of clusterwise inference thresholding, $|V_{RFT}|$ in case of RFT-applied FWE voxel-wise thresholding, $|V_{Bon}|$ in case of Bonferroni's method-applied FWE correction) and that of the number of overlapped voxels to the number of activation foci found by meta-analysis ($|V_{Meta}|$) and SnPM ($|V_{SnPM}|$). However, these ratios would not provide us with enough information regarding the overall fit; instead, it may demonstrate a biased result. For instance, if survived voxels are completely contained by activation foci found by the meta-analysis, $|V_{Ovl}|/|V_{Cor}|$ becomes 1.00, but $|V_{Ovl}|/|V_{Meta}|$ can be smaller than 1.00. On the other hand, if resultant voxels from the meta-analysis are subsets of survived voxels, $|V_{Ovl}|/|V_{Meta}|$ is 1.00 while $|V_{Ovl}|/|V_{Cor}|$ can become smaller than 1.00. In these

situations, we cannot make an accurate decision about which case shows a greater overlap solely based on the two independent ratios.

Instead, we may consider employing a unified overlap index, which takes into account both ratios simultaneously. The harmonic mean, instead of the arithmetic mean, would be a feasible and reliable way to calculate the overall overlap index with two different ratios, given its definition (Marchiori & Latora, 2000). We can then calculate the overall overlap index (I_{Ovl}) as follows:

$$I_{Ovl} = \frac{n\prod_{j=1}^{n} x_j}{\sum_{i=1}^{n}\left\{\frac{1}{x_i}\prod_{j=1}^{n} x_j\right\}} = \frac{2\frac{|V_{Ovl}|\ |V_{Ovl}|}{|V_{Cor}||V_{Meta}|}}{\frac{|V_{Ovl}|}{|V_{Cor}|}+\frac{|V_{Ovl}|}{|V_{Meta}|}}$$ (for comparison with meta-analysis)

$$I_{Ovl} = \frac{n\prod_{j=1}^{n} x_j}{\sum_{i=1}^{n}\left\{\frac{1}{x_i}\prod_{j=1}^{n} x_j\right\}} = \frac{2\frac{|V_{Ovl}|\ |V_{Ovl}|}{|V_{Cor}||V_{SnPM}|}}{\frac{|V_{Ovl}|}{|V_{Cor}|}+\frac{|V_{Ovl}|}{|V_{SnPM}|}}$$ (for comparison with SnPM)

Five hypothetical cases are introduced as examples (see Figure 1). Red areas represent survived voxels after the application of correction method, yellow areas represent activation foci found by meta-analysis, and white areas represent overlapped voxels. Table 1 demonstrates two different ratios calculated for each case.

<Place Figure 1 about here>

Table 1. Sole Ratios Compared to Overlap Index for the Cases Depicted in Figure 1

	Case 1	Case 2	Case 3	Case 4	Case 5
$\|V_{Ovl}\| / \|V_{Cor}\|$	.15	.33	1.00	.00	1.00

$\lvert V_{Ovl} \rvert / \lvert V_{Meta} \rvert$	1.00	.48	.15	.00	1.00
I_{Ovl}	.26	.39	.26	.00	1.00

Both case 1 and 3 show why the sole ratio, instead of the overall overlap index, may be misleading when we attempt to evaluate the overlap quantitatively. Although one of two different types of overlap ratios is 1.00, another ratio is significantly smaller and vice versa (.15); in this case, we are not able to decide which one would better represent the overall trend of overlap. Instead, the overall overlap index would be a non-biased solution to address this issue. It can provide us with a unified value for overlap index by taking into account two different types of ratios at the same time. For instance, case 1 and 3 in fact show an identical degree of overlap as visualized in Figure 1 and their overall overlap indices are identical to each other. Furthermore, the overall overlap index value of case 2 is greater than the calculated index values of case 1 and 3; this result is consistent with what we would expect from the apparent ratio of white areas to other areas as presented in the diagrams.

For the evaluation of correction methods in the present study, we calculated the overall overlap index for one contrast, i.e., all moral versus non-moral, because included task conditions in the meta-analysis that intended to be performed are moral cognition and emotion in general, and moral judgment that did not distinguish task conditions according to the nature of moral dilemmas, i.e., moral-personal and moral-impersonal dilemmas. The overlap index was also calculated for case of the overlap between survived voxels after the application of each thresholding method and SnPM. As a result, four overlap indices were calculated ($I_{Ovl(FDR)}$, $I_{Ovl(CLU)}$, $I_{Ovl(RFT)}$, and $I_{Ovl(Bon)}$) for each type of meta-analysis (either moral cognition and emotion

in general, or moral judgment) and SnPM. We examined which correction method produced the highest overlap index value.

Results

Reanalysis of fMRI data

We reanalyzed the fMRI to examine which voxels were significantly activated by the different contrasts in the moral judgment task after the application of each correction method. Although we used the spiral in-and-out method that is more robust against the signal dropout in ventral areas in the prefrontal cortex compared to the EPI methods (Glover, 2012; Glover & Law, 2001), some prefrontal regions below z = -12 demonstrated signal loss and were excluded for our reanalysis (see Figure S1). Figure 2 demonstrates survived voxels for each contrast (all moral versus non-moral, moral-personal versus non-moral, moral-impersonal versus non-moral, moral-personal versus moral-impersonal, and moral-impersonal versus moral-personal). Table 2 summarizes the number of survived voxels for each correction method. In addition, Table S2 provides information regarding survived voxels for each contrast. As previous methodological studies have shown, the Bonferroni method-applied FWE voxel-wise thresholding was most conservative, and the FDR-applied thresholding was most lenient (Bennett, Wolford, et al., 2009; Nichols, 2013) among four different thresholding methods in terms of the number of survived voxels ($V_{FDR} \supset V_{CLU} \supset V_{RFT} \supset V_{Bon}$) for all contrasts, except for the contrast of moral-personal vs. moral-impersonal.

<Place Figure 2a-e about here>

Table 2. Number of Voxels Surviving with Four Different Thresholds and SnPM Method as a Reference

	FDR ($\|V_{FDR}\|$)	Clusterwise inference ($\|V_{CLU}\|$)	RFT FWE (voxel-wise) ($\|V_{RFT}\|$)	Bonferroni's FWE ($\|V_{Bon}\|$)	Voxel-wise SnPM ($\|V_{SnPM}\|$)
All moral vs. non-moral	24,524	14,424	264	66	416
Moral-personal vs. non-moral	20,883	13,540	817	227	2,084
Moral-impersonal vs. non-moral	13,787	7,458	0	0	20
Moral-personal vs. moral-impersonal	2,816	2,941	46	0	370
Moral-impersonal vs. moral-personal	11,474	6,826	22	2	235

Examination of Overlap with Meta-analysis and SnPM Results

By using the equation for the calculation of the overall overlap index, we examined the quality of each correction method. Survived voxels after the application of each correction method were compared with activation foci identified in the ALE maps created by GingerALE and SnPM. Although some ventral parts of the prefrontal cortex below z = -12 were excluded from our fMRI reanalysis due to signal dropout, no excluded voxels overlapped with any significant voxels in

the ALE maps as significant voxels were located above z = -8 in the ventromedial prefrontal cortex. Figure 3 demonstrates comparisons with common activation foci of moral cognition and emotion in general, and moral judgment, respectively. Table 3 shows the calculated overall overlap index for each case. As presented, the best overall overlap was achieved when the RFT FWE corrected voxel-wise thresholding was applied. In all cases, the RFT FWE showed the best performance. In the comparisons with the meta-analysis of moral cognition and emotion in general, FDR resulted in 21.4% less overlap with meta-analysis results than RFT FWE, thresholding with clusterwise inference resulted in 2.9% less overlap, and Bonferroni FWE resulted in 66% less overlap. In the comparisons with the meta-analysis of moral judgment, FDR resulted in 72.8% less overlap with meta-analysis results than RFT FWE, thresholding with clusterwise inference resulted in 66.2% less overlap, and Bonferroni FWE resulted in 54.3% less overlap. Finally, when the results were compared with SnPM, FDR resulted in 87.3% less overlap with SnPM results than RFT FWE, thresholding with clusterwise inference resulted in 78.4% less overlap, and Bonferroni FWE resulted in 32.8% less overlap.

<Place Figure 3a-c about here>

Table 3. Overlap Index Representing the Degree of Overlap between Activated Voxels Using Each Correction Method and Results from the Meta-Analyses and SnPM

	FDR ($I_{Ovl(FDR)}$) (24,524 voxels)	Clusterwise inference ($I_{Ovl(CLU)}$) (14,424 voxels)	RFT FWE ($I_{Ovl(RFT)}$) (264 voxels)	Bonferroni's FWE ($I_{Ovl(Bon)}$) (66 voxels)

Meta-analysis: moral cognition and emotion in general (3,258 voxels)	.081	.100	.103	.035
Meta-analysis: moral judgment (1,287 voxels)	.041	.051	.151	.069
SnPM: moral-personal + moral-impersonal (416 voxels)	.033	.056	.259	.174

Note. The number of survived voxels for each case is presented in parentheses.

Discussion

There has been much discussion in the literature about the importance of finding a balance between Type I and Type II errors in fMRI studies. Although some have advocated for less stringent thresholds in order to reduce the risk of missing true effects in social and affective neuroscience studies (Lieberman & Cunningham, 2009), recent studies by Eklund et al. (2016) and Bennett et al. (2009) suggest that some of these methods may be far more lenient than expected or desired. Determining an appropriate method for balancing Type I and Type II error is particularly difficult in social neuroscience studies, which often (though not always) involve complex paradigms that measure less precise and more variable mental processes, resulting in smaller effect sizes and weaker power.

In the present study, we evaluated four different methods for correcting for multiple comparisons within the social neuroscience domain of moral judgment by examining which method produced results that most closely matched the effects identified in a meta-analysis (i.e., an estimate of the "real" effects) and SnPM (i.e., an estimate without any parametric assumptions). As in previous studies, we found that the RFT-applied FWE correction implemented in SPM was similar to or slightly less conservative than the Bonferroni FWE correction method, but more conservative than the FDR correction (Bennett, Wolford, et al., 2009; Eklund et al., 2016; Nichols, 2013; Nichols & Hayasaka, 2003; Nichols & Holmes, 2002) in the case of voxel-wise inference. Also, as recently reported, the clusterwise thresholding provided in SPM12 by default was also more lenient than the voxel-wise FWE-applied thresholding (Eklund et al., 2016). When compared to results from meta-analyses of studies on moral cognition and emotion in general, and moral judgment specifically, as well as those from the application of SnPM, activation maps from the RFT FWE-applied voxel-wise thresholding demonstrated the most overlap, as calculated by the overall overlap index.

These findings suggest that the RFT FWE-applied voxel-wise thresholding may be an acceptable correction method for studies of moral psychology despite the limitations that have been discussed in the field of social neuroscience (e.g., conservativeness and a lack of statistical power). The RFT FWE-applied voxel-wise thresholding may provide a more appropriate balance between false positives and false negatives than other correction methods. Although this method implemented in SPM is susceptible to reduced sensitivity, in the context of some areas of social neuroscience the RFT FWE-applied voxel-wise thresholding may still be considered a viable correction method given the calculated overall overlap indices despite the statistical power issue. However, researchers in social neuroscience should seriously consider cross-checking the

thresholding method with meta-analyses and/or SnPM in order to evaluate whether it is appropriate to apply the method in the context of their study.

In the present study, we compared different correction methods in the context of a moral judgment task. However, this unique approach of comparing the results from different correction methods to those of a meta-analysis could be applied to many different domains of social and affective neuroscience. Future studies using this approach can provide more information about whether voxel-wise RFT FWE results in the most overlap with meta-analysis results in other domains within social and affective neuroscience.

The primary limitation of the present study is the imperfect nature of the meta-analysis that was used to examine the quality of the multiple correction methods. The meta-analysis was based on studies that implement a variety of tasks related to moral cognition and emotion, and that use a variety of correction methods and thresholds, thus introducing the possibility of biased results. However, by aggregating data from many studies, the hope is that the meta-analysis technique can provide a relatively unbiased indicator of the areas that appear to be commonly activated across many studies – particularly in social neuroscience studies that involve rather complex processes. In addition, as Eklund et al. (2016) warned that the employment of meta-analysis does not necessarily mitigate the need for the application of valid inferential methodologies for individual studies.

Another limitation is that, as is common in meta-analyses, our meta-analysis was not limited to previous studies that utilized the same paradigm (Greene et al.'s (2001, 2004) experimental design). Thus, the activation foci identified in the meta-analysis might include brain voxels not directly associated with moral judgment, the process of interest, and might be

inappropriate to be used for a comparison and evaluation. Of course, the best way to address this issue is to only meta-analyze published studies that used Greene et al.'s (2001, 2004) dilemmas; however, due to the limited number of studies, this was not feasible. To address this issue, we tried to apply a more stringent inclusion rule (i.e., meta-analyzing studies related to moral judgment instead of moral cognition and emotion in general) in order to address this issue; the results demonstrated that there were greater overlaps between activation foci found by the aforementioned fMRI experiment and activation foci found by a meta-analysis of studies related moral judgment compared to those found by a meta-analysis of studies related moral cognition and emotion in general.

Given these limitations associated with meta-analysis, researchers may have to employ additional cross-checking methods, such as SnPM, which is free from any error originating from parametric assumptions, to provide further evaluation of findings (Nichols & Hayasaka, 2003). Researchers may also consider utilizing alternative correction methods not assessed here, such as the threshold-free cluster enhancement (TFCE), which is considered to have greater sensitivity than the traditional methods and allows for the false positive rate to be set at a predetermined level by the permutation test (Smith & Nichols, 2009). Although, this function has not been implemented in SPM, which was examined in the present study, it is available in FSL as an option in the randomise permutation-based inference tool.

Conclusion

The goal of correcting for multiple comparisons in fMRI studies is to generate clusters of activity that reflect true effects, and thus would be expected to replicate in future studies. Here we show that using the RFT FWE-applied voxel-wise thresholding method in a study of moral judgment produced the most overlap with results from a meta-analysis on moral cognition and

moral emotion and the most overlap with SnPM analyses, suggesting that this method may be the best for achieving the goal of identifying true effects. Although this method suffers from potentially insufficient statistical power, which has been a significant issue in social neuroscience, it may be an acceptable option in the context of experiments focusing on morality and possibly other domains of social neuroscience, as long as its application is cross-checked with other methods.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*(1), 289–300. doi:10.2307/2346101

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, *93*(3), 491–507. doi:10.1093/biomet/93.3.491

Bennett, C. M., Miller, M. B., & Wolford, G. L. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. *NeuroImage*, *47*(Suppl 1), S125.

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience*, *4*(4), 417–422. doi:10.1093/scan/nsp053

Bland, J. M., & Altman, D. G. (1995). Statistics notes: Multiple significance tests: the Bonferroni method. *BMJ*, *310*(6973), 170–170. doi:10.1136/bmj.310.6973.170

Blasi, A., & Hoeffel, E. C. (1974). Adolescence and Formal Operations. *Human Development*, *17*(5), 344–363. doi:10.1159/000271357

Brett, M., Penny, W., & Kiebel, S. J. (2004). Introduction to random field theory. In R. S. J. Frackowiak, K. J. Friston, C. Fritch, R. J. Dolan, C. Price, & W. Penny (Eds.), *Human Brain Function, 2nd ed.* (pp. 867–880). New York, NY: Academic Press.

Cui, X., Li, J., & Song, X. (2015). xjview. Retrieved June 28, 2015, from http://www.alivelearn.net/xjview

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, *59*(3), 2349–2361. doi:10.1016/j.neuroimage.2011.09.017

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*, 2907–2926.

Eklund, A., Dufort, P., Villani, M., & LaConte, S. (2014). BROCCOLI: Software for fast fMRI analysis on many-core CPUs and GPUs. *Frontiers in Neuroinformatics*, *8*. doi:10.3389/fninf.2014.00024

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905. doi:10.1073/pnas.1602413113

Flandin, G., & Novak, M. J. U. (2013). fMRI data analysis using SPM. In S. Ulmer & O. Jansen (Eds.), *fMRI* (pp. 51–76). Heidelberg, Germany: Springer Berlin Heidelberg. doi:10.1007/978-3-642-34342-1_6

Fox, P. T., Laird, A. R., Eickhoff, S. B., Lancaster, J. L., Fox, M., Uecker, A. M., … Ray, K. L. (2013). User manual for GingerALE 2.3. Retrieved from https://www.brainmap.org/ale/manual.pdf

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*(4), 870–878.

doi:10.1006/nimg.2001.1037

Glover, G. H. (2012). Spiral imaging in fMRI. *Neuroimage*, *62*(2), 706–712. doi:DOI 10.1016/j.neuroimage.2011.10.039

Glover, G. H., & Law, C. S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magnetic Resonance in Medicine*, *46*(3), 515–522. doi:10.1002/Mrm.1222

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. doi:10.1016/j.neuron.2004.09.027

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), 2105–2108. doi:10.1126/science.1062872

Han, H. (2016). *Neuroscientific and social psychological investigation on psychological effects of stories of moral exemplars*. Stanford University.

Han, H. (2017). Neural correlates of moral sensitivity and moral judgment associated with brain circuitries of selfhood: A meta-analysis. *Journal of Moral Education*. doi:10.1080/03057240.2016.1262834

Han, H., Chen, J., Jeong, C., & Glover, G. H. (2016). Influence of the cortical midline structures on moral emotion and motivation in moral decision-making. *Behavioural Brain Research*, *302*, 237–251. doi:10.1016/j.bbr.2016.01.001

Han, H., Glover, G. H., & Jeong, C. (2014). Cultural influences on the neural correlate of moral

decision making processes. *Behavioural Brain Research*, *259*, 215–228. doi:10.1016/j.bbr.2013.11.012

Laird, A. R., Lancaster, J. L., & Fox, P. T. (2005). BrainMap: the social evolution of a human brain mapping database. *Neuroinformatics*, *3*, 65–78.

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: Re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*, 423–428. doi:10.1093/scan/nsp052

Marchiori, M., & Latora, V. (2000). Harmony in the small-world. *Physica A: Statistical Mechanics and Its Applications*, *285*(3–4), 539–546. doi:10.1016/S0378-4371(00)00311-3

Narvaez, D., Getz, I., Rest, J. R., & Thoma, S. J. (1999). Individual moral judgment and cultural ideologies. *Developmental Psychology*, *35*, 478–488. doi:10.1037/0012-1649.35.2.478

Nichols, T. (2012). Multiple testing corrections, nonparametric methods, and random field theory. *NeuroImage*, *62*(2), 811–815. doi:10.1016/j.neuroimage.2012.04.014

Nichols, T. (2013). False Discovery Rate. Retrieved June 20, 2015, from http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/software/fdr

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*(5), 419–446. doi:10.1191/0962280203sm341ra

Nichols, T., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. doi:10.1002/hbm.1058

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. doi:10.1016/j.neuron.2011.11.001

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. doi:10.1016/j.neuroimage.2008.03.061

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, *92*, 381–397. doi:10.1016/j.neuroimage.2014.01.060

Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, *91*, 412–419. doi:10.1016/j.neuroimage.2013.12.058

Figure Legends

Figure 1. Sample cases for the overall overlap index calculation

Figure 2. Survived voxels after the application of correction methods. Blue: voxel-wise FDR; green: clusterwise inference; yellow: voxel-wise RFT FWE; red: voxel-wise Bonferroni FWE. (a) All moral (moral-personal + moral-impersonal) vs. non-moral. (b) Moral-personal vs. non-moral. (c) Moral-impersonal vs. non-moral. (d) Moral-personal vs. moral-impersonal. (e) Moral-impersonal vs. moral-personal. Figures created with XjView (Cui, Li, & Song, 2015).

Figure 3. Comparisons between voxels identified using the four correction methods (sky blue: voxel-wise FDR; crimson red: clusterwise inference; yellow: voxel-wise RFT FWE; bright red: voxel-wise Bonferroni FWE) and voxels identified in the meta-analyses and SnPM (areas surrounded by blue lines). (a) Comparisons with meta-analysis of general moral cognition and emotion in general. (b) Comparisons with meta-analysis of moral judgment. (c) Comparisons with voxels identified by SnPM. Figures created with XjView (Cui et al., 2015).

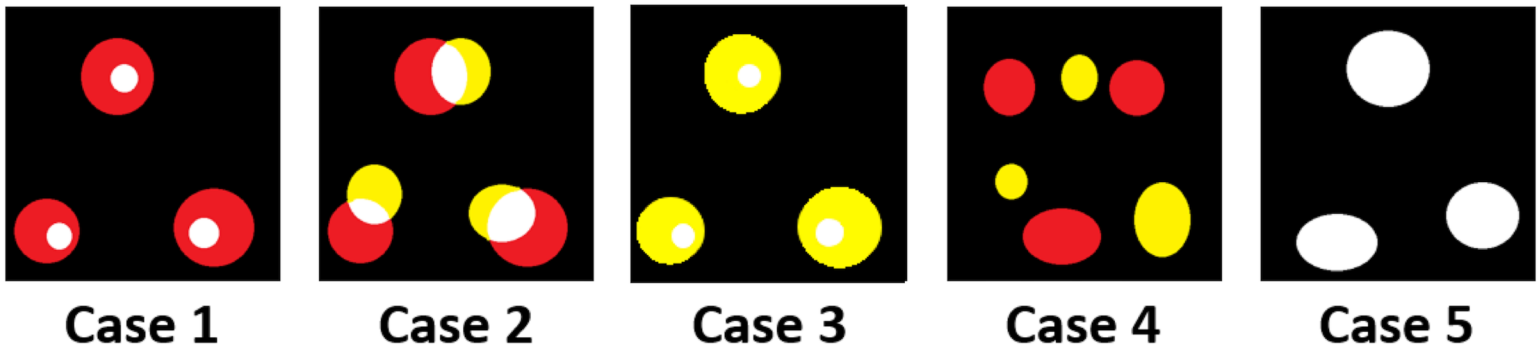

Figure 1

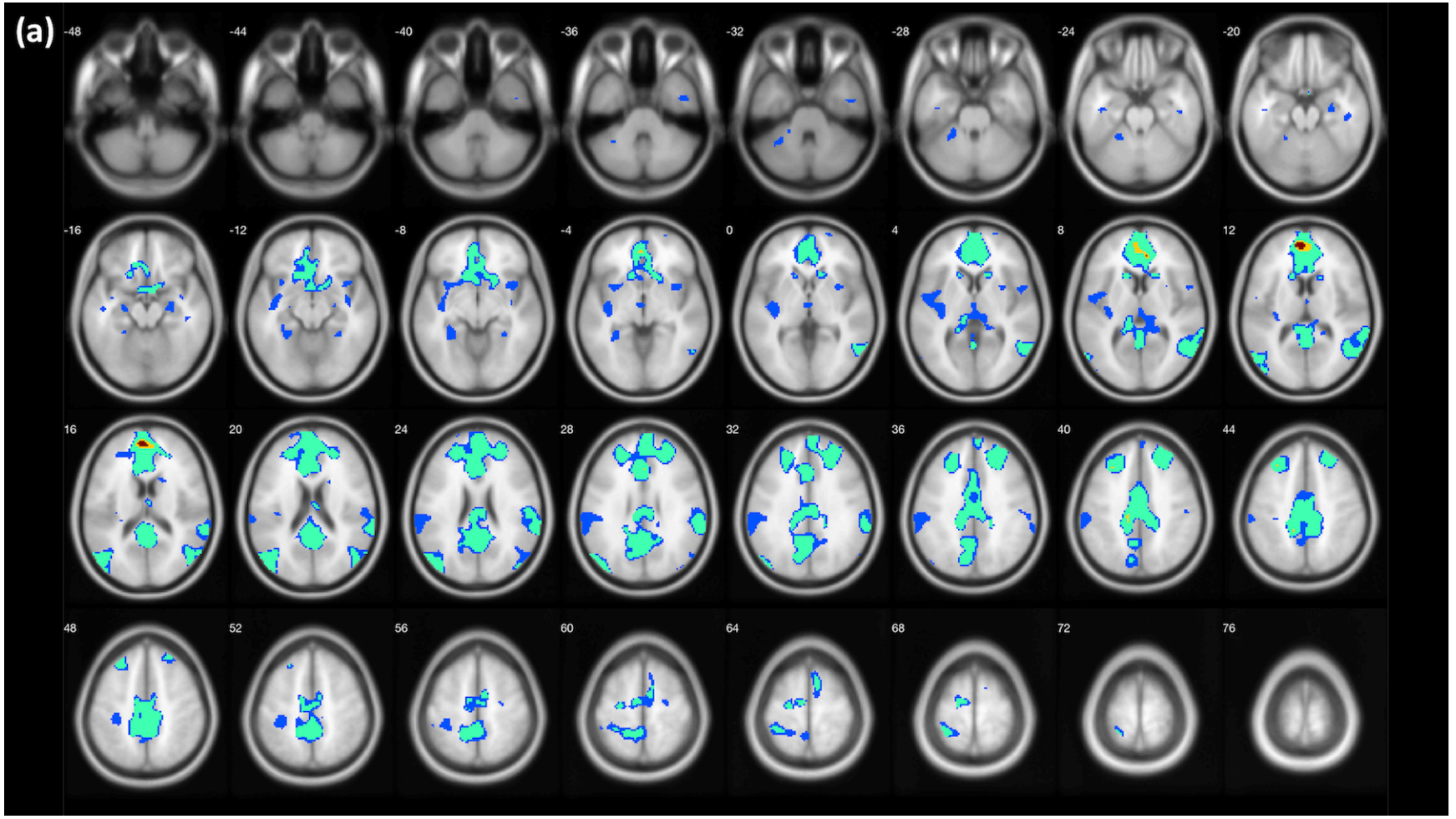
(a)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 2(a)

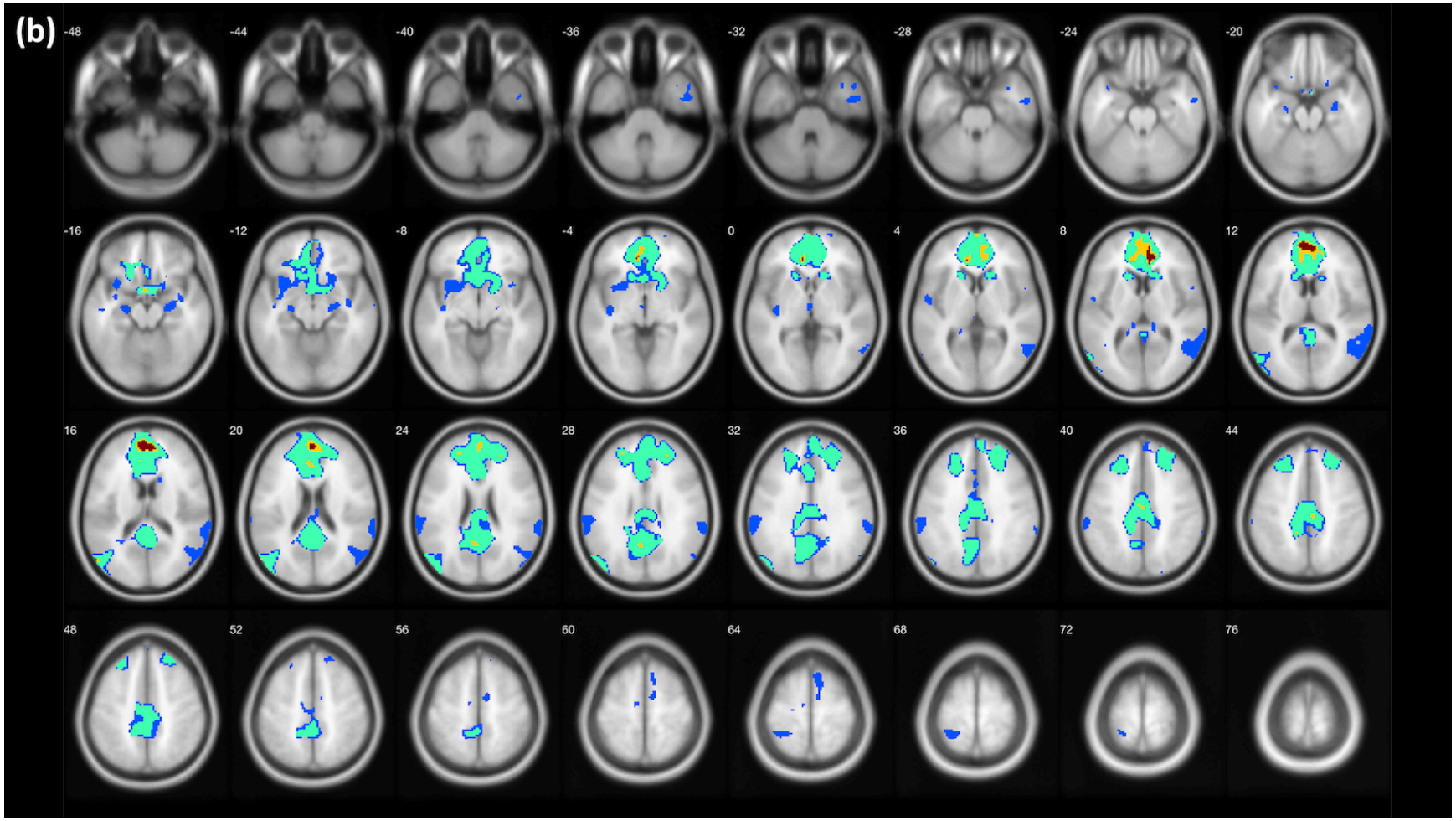
(b)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 2(b)

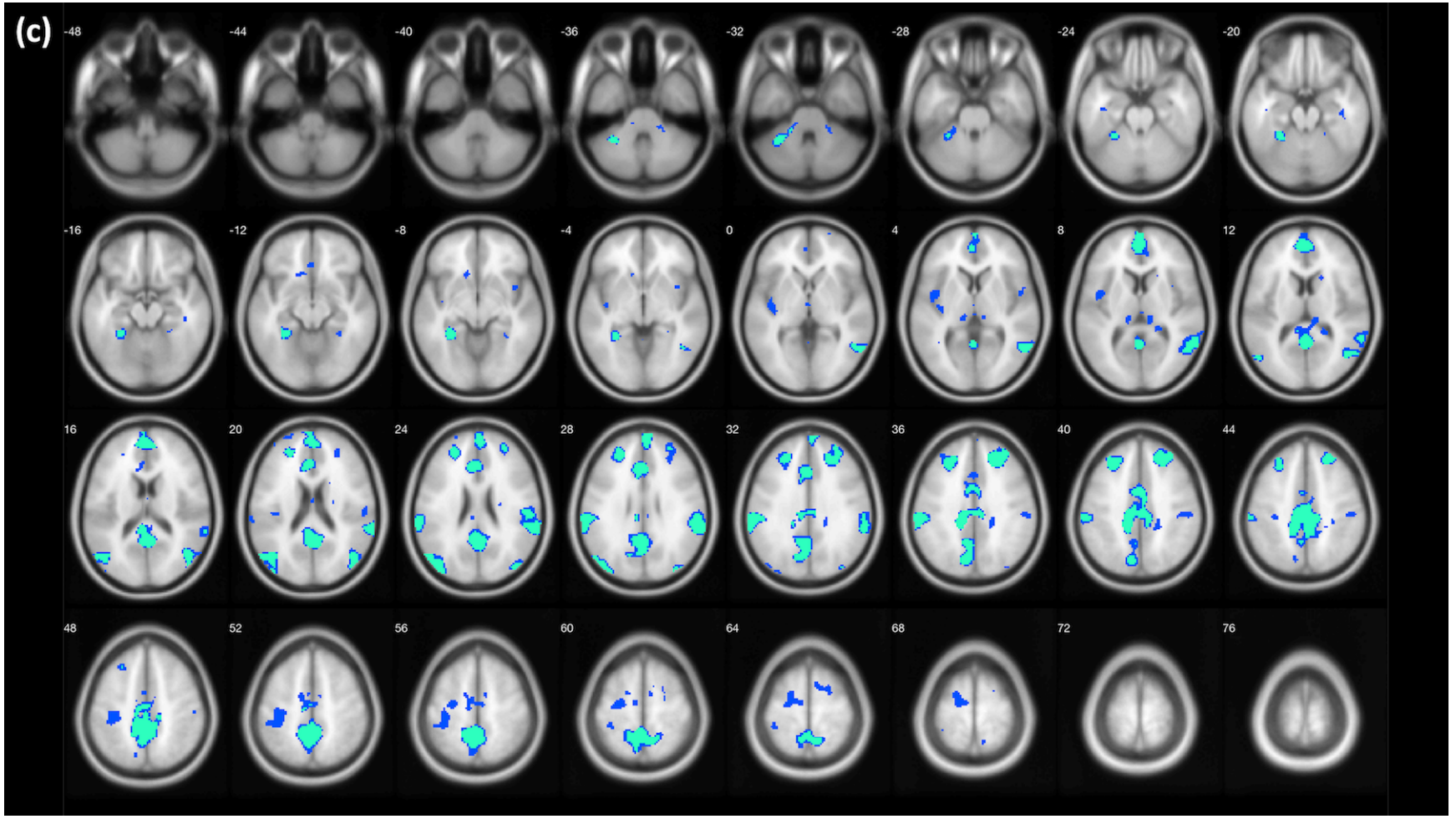
(c)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 2(c)

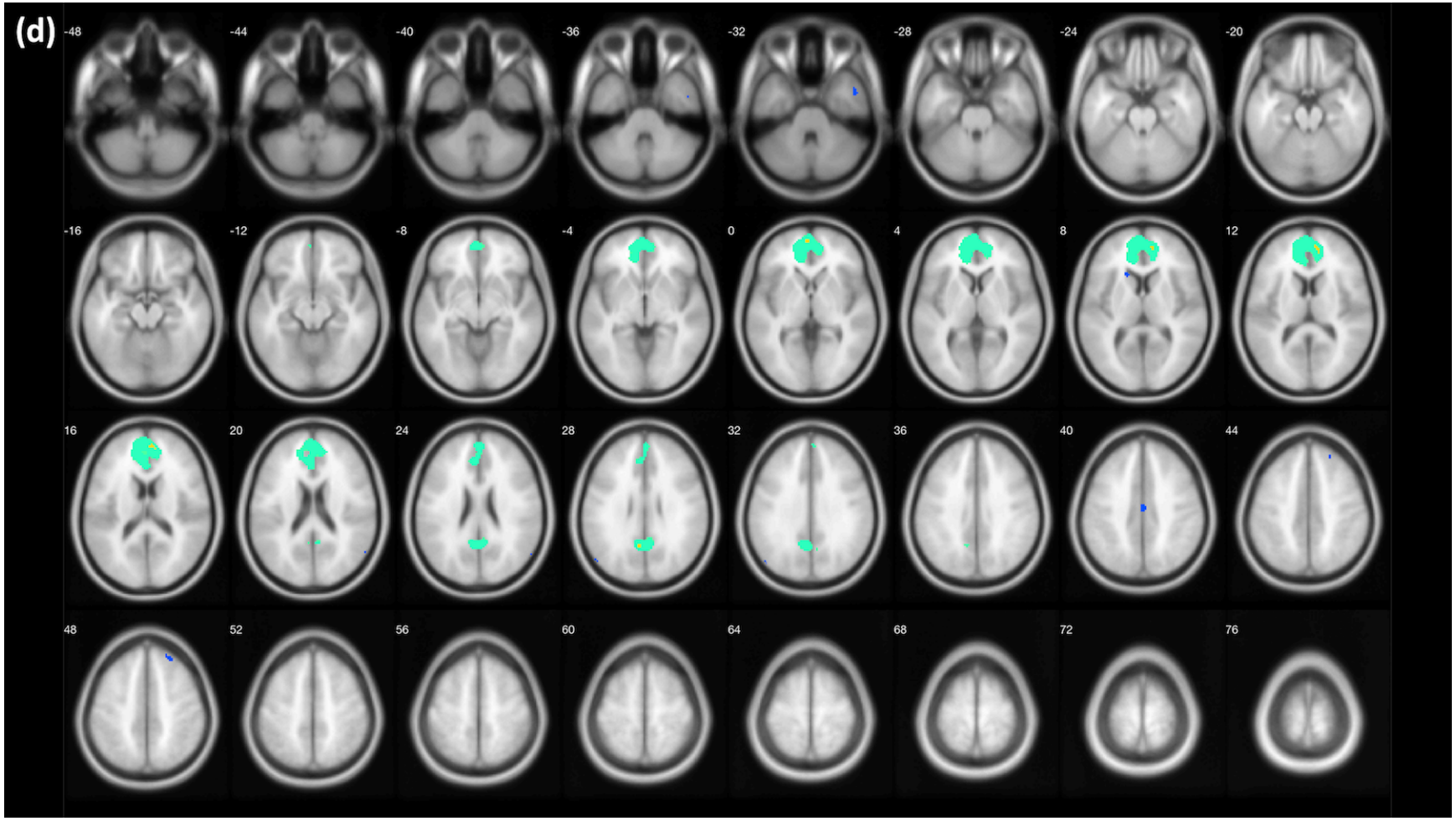
(d)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 2(d)

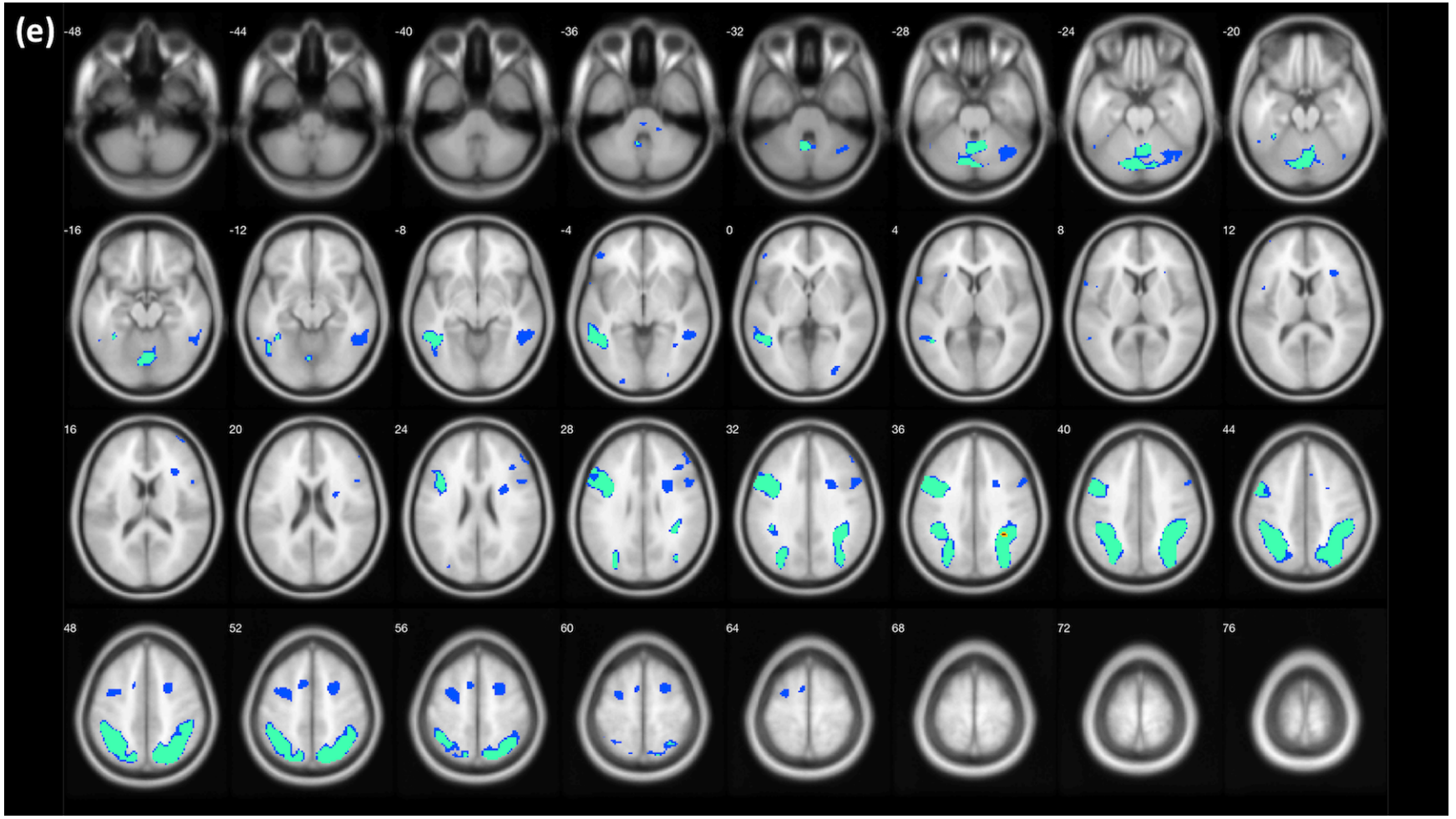
(e)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 2(e)

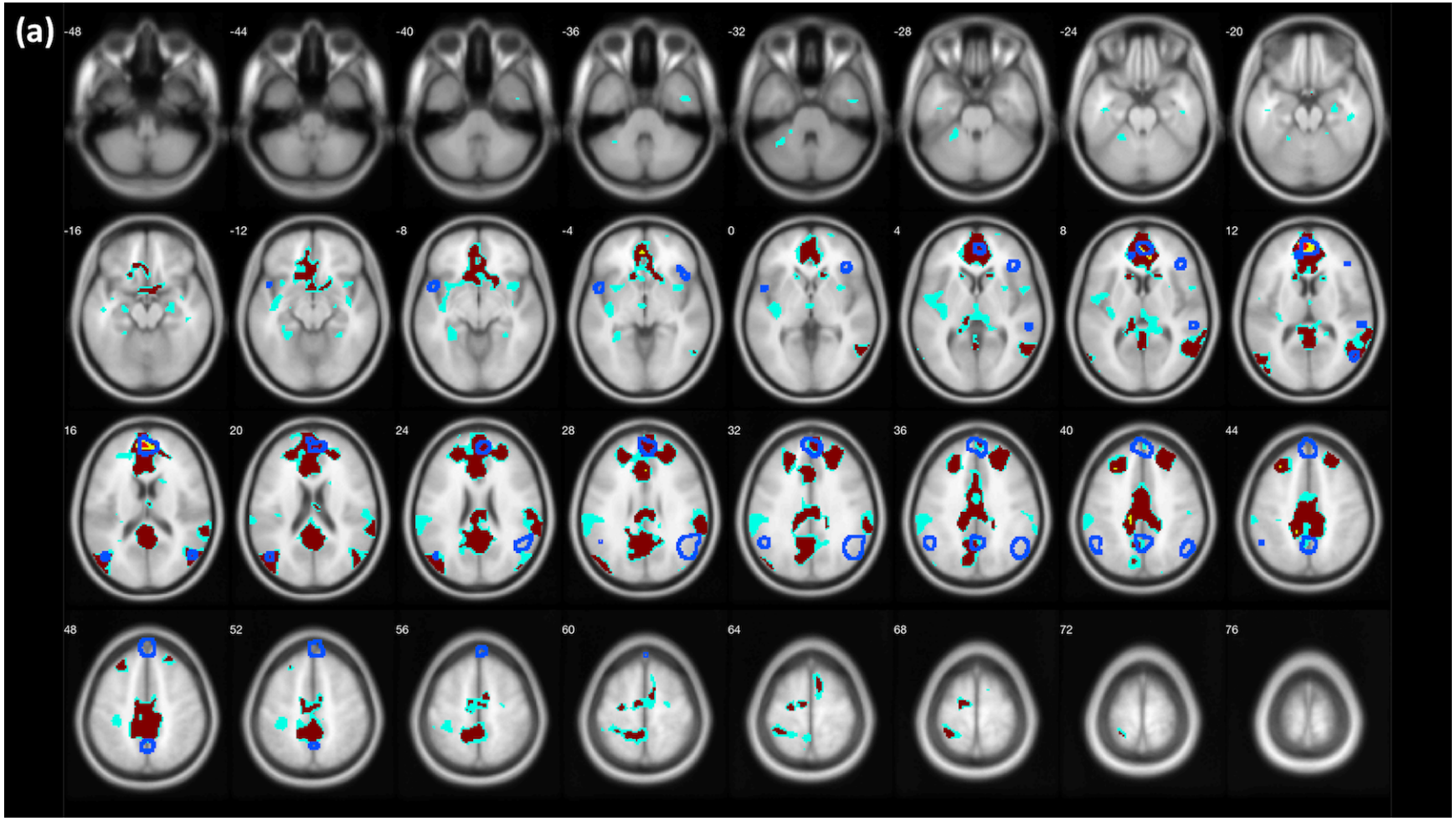
(a)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 3(a)

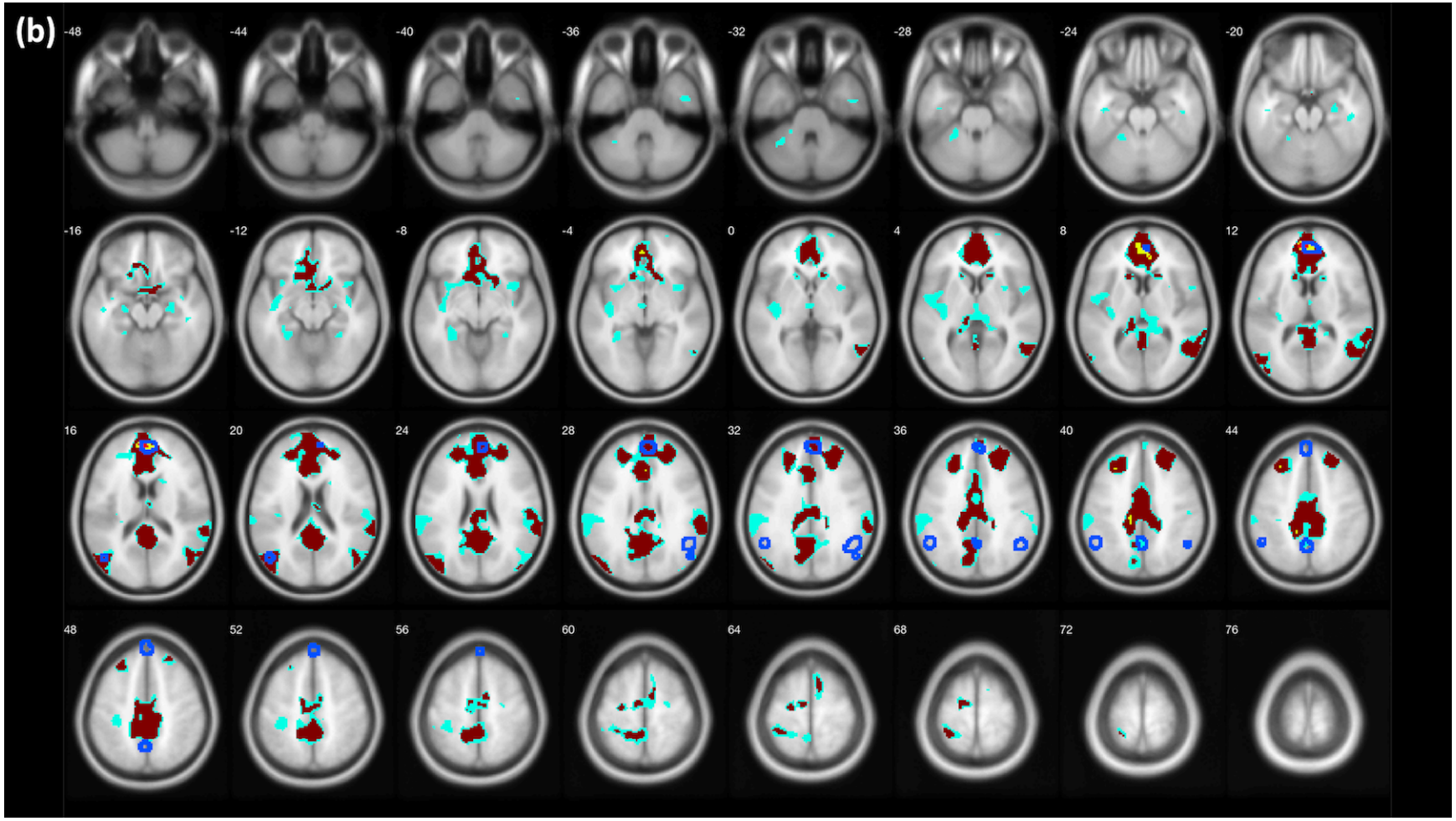
(b)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 3(b)

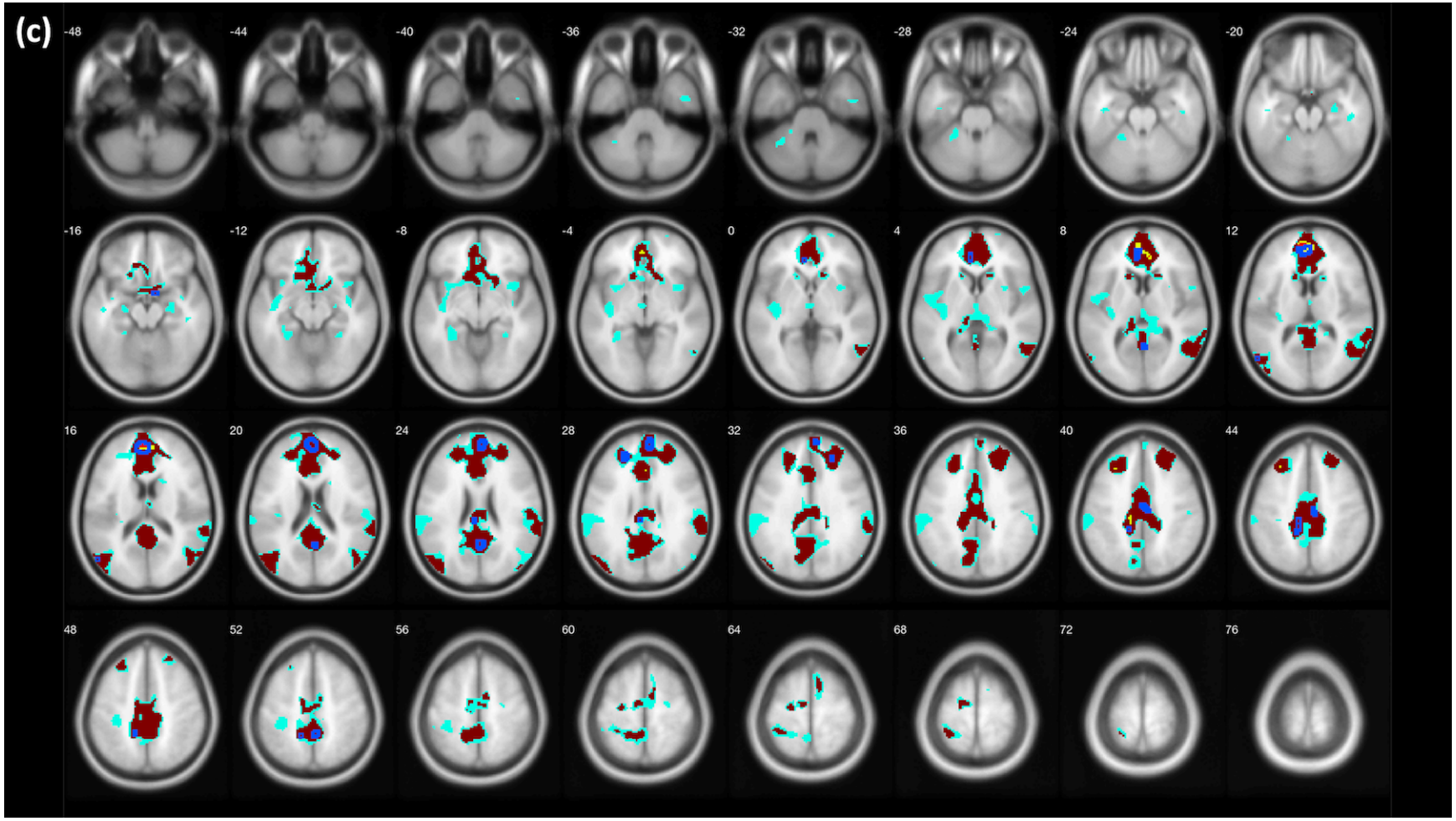
(c)
-48
-44
-40
-36
-32
-28
-24
-20
-16
-12
-8
-4
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60
64
68
72
76

Figure 3(c)

Supplementary Materials

Table S1. Experiments Included in the Meta-analysis

References		Subject	Foci	Type		Contrast	Threshold
				Judgment	Non-judgment		
Judgment Experiments							
Moll, de Oliveira-Souza, Bramati, et al. (2002)		7	3	x		Moral vs. Non-moral emotions	$p < .001$, uncorr.
Heekeren et al. (2003)		8	9	x		Moral vs. Semantic judgment	$p < .005$, uncorr.
Heekeren et al. (2005)		12	8	x		Moral vs. Semantic judgment	$p < .001$, uncorr.
Berthoz et al. (2006)		12	8	x		Moral-related intentional vs. Accidental violation judgment	$p < .001$, uncorr.
Borg et al. (2006)		24	7	x		Moral vs. Non-moral harm	$p < .05$, corr
Young et al. (2007)	1	10	6	x		Moral belief task vs. Photo presentation	$p < .001$, uncorr.
	2	17	6				
Harrison et al., (2008)		22	8	x		Moral dilemma vs. Stroop task	$p < .05$, corr.
Harenski et al. (2008)		33	16	x		Moral vs. Non-moral violation	$p < .001$, uncorr.
Young and Saxe (2008)		17	6	x		Moral vs. value-neutral judgment	$p < .001$, uncorr.
Sommer et al. (2010)		12	6	x		Moral vs. Neutral conflicts	$p < .05$, corr.
Schleim et al. (2011)		40	6	x		Moral vs. value-neutral judgment	$p < .005$, corr.
FeldmanHall et al. (2012)		14	14	x		Moral (real + hypothetical) vs. Non-moral	$p < .05$, corr.
Reniers et al. (2012)		24	6	x		Moral vs. Non-moral decision-making	$p < .05$, corr.
FeldmanHall et al. (2014)		38	1	x		Moral vs. Non-moral dilemmas	$p < .05$, corr.

Han et al. (2014)	16	11	x		Moral (personal + impersonal) vs. Non-moral (arithmetic)	$p < .001$, uncorr.
Shenhav and Greene (2014)	35	5	x		Integrative moral vs. utilitarian & emotional judgment	$p < .05$, corr.
Sommer et al. (2014)	32	16	x		Moral vs. Neutral conflicts	$p < .05$, corr.
Subtotal	**373**	**142**				
Non-judgment experiment						
Moll et al. (2001)	10	10		x	Moral vs. factual evaluation	$p < .0001$, uncorr.
Berthoz et al. (2002)	12	20		x	Socio-moral vs. Non-socio-moral violation stories	$p < .0001$, uncorr.
Moll, de Oliveira-Souza, Eslinger, et al. (2002)	7	17		x	Moral vs. Non-moral pictures	$p < .005$, uncorr.
Singer et al. (2004)	11	13		x	Moral vs. Non-moral status face watching	$p < .001$, uncorr.
Takahashi et al. (2004)	19	15		x	Moral guilt and embarrassment vs. Neutral feeling	$p < .001$, uncorr.
Moll, de Oliveira-Souza, et al. (2005)	13	22		x	Moral indignation vs. Basic disgust, neutral	$p < .005$, uncorr.
Finger et al. (2006)	16	5		x	Moral vs. Conventional transgression evaluation	$p < .05$, corr.
Harenski and Hamann (2006)	10	2		x	Moral vs. Non-moral violation watching	$p < .001$, uncorr.
Moll et al. (2007)	12	31		x	Moral (Guilt, Embarrassment, Compassion, Indignation) vs. Neutral agency	$p < .005$, uncorr.
Robertson et al. (2007)	16	10		x	Justice/care vs. Non-moral strategy evaluation	$p < .001$, uncorr.
Borg et al. (2008)	50	24		x	Socio-moral vs. Pathogen disgust feeling	$p < .05$, corr.
Prehn et al. (2008)	23	6		x	Socio-moral vs. Grammatical errors	$p < .05$, corr.
Takahashi et al. (2008)	15	7		x	Moral beauty + depravity vs. Emotion-neutral	$p < .001$, uncorr.
Immordino-Yang et al. (2009)	13	8		x	Moral vs. Physical admiration	$p < .05$, corr.
Young and Saxe (2009) 1	14	6		x	Moral vs. Non-moral intention evaluation	$p < .001$, uncorr.
Young and Saxe (2009) 2	14	6				
Cope et al. (2010)	100	17		x	Moral vs. Nonmoral wrongdoing evaluation	$p < .05$, corr.
Harenski et al. (2010)	30	10		x	Moral vs. Non-moral picture viewing	$p < .05$, corr.

Young et al. (2010)	17	5	x	Moral vs. Non-moral story reading	$p < .001$, uncorr.
Parkinson et al. (2011)	38	2	x	Moral vs. Neutral transgression stories	$p < .05$, corr.
Young et al. (2011)	17	6	x	Moral vs. Non-moral intuitive verdict evaluation	$p < .001$, uncorr.
Englander et al. (2012)	10	8	x	Moral vs. Physical admiration	$p < .05$, corr.
Harenski et al. (2012)	51	17	x	Moral vs. Non-moral violation	$p < .05$, corr.
Avram et al. (2013)	16	8	x	Moral vs. Esthetic judgment	$p < .05$, corr.
Avram et al. (2014)	16	10	x	First- and third-person moral vs. Non-moral evaluation	$p < .005$, uncorr.
Fourie et al. (2014)	22	8	x	Moral-related prejudice vs. neutral feeling	$p < .001$, uncorr.
Michl et al. (2014)	14	28	x	Shame and guilt vs. neutral emotion	$p < .0002$, uncorr.
Subtotal	**586**	**321**			
Grand Total	**959**	**463**			

Included Articles

Avram, M., Gutyrchik, E., Bao, Y., Pöppel, E., Reiser, M., & Blautzik, J. (2013). Neurofunctional correlates of esthetic and moral judgments. *Neuroscience Letters*, *534*, 128–32. http://doi.org/10.1016/j.neulet.2012.11.053

Avram, M., Hennig-Fast, K., Bao, Y., Pöppel, E., Reiser, M., Blautzik, J., … Gutyrchik, E. (2014). Neural correlates of moral judgments in first- and third-person perspectives: implications for neuroethics and beyond. *BMC Neuroscience*, *15*, 39. http://doi.org/10.1186/1471-2202-15-39

Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, *125*, 1696–1708. http://doi.org/10.1093/brain/awf190

Berthoz, S., Grèzes, J., Armony, J. L., Passingham, R. E., & Dolan, R. J. (2006). Affective response to one's own moral violations. *NeuroImage*, *31*(2), 945–50. http://doi.org/10.1016/j.neuroimage.2005.12.039

Borg, J. S., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience*, *20*, 1529–1546. http://doi.org/10.1162/jocn.2008.20109

Cope, L. M., Borg, J. S., Harenski, C. L., Sinnott-Armstrong, W., Lieberman, D., Nyalakanti, P. K., … Kiehl, K. A. (2010). Hemispheric Asymmetries during Processing of Immoral Stimuli. *Frontiers in Evolutionary Neuroscience*, *2*, 110. http://doi.org/10.3389/fnevo.2010.00110

Englander, Z. A., Haidt, J., & Morris, J. P. (2012). Neural Basis of Moral Elevation Demonstrated through Inter-Subject Synchronization of Cortical Activity during Free-Viewing. *Plos One*, *7*(6), e39384. Journal Article. http://doi.org/10.1371/journal.pone.0039384

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, *7*, 743–51. http://doi.org/10.1093/scan/nss069

FeldmanHall, O., Mobbs, D., & Dalgleish, T. (2014). Deconstructing the brain's moral network: Dissociable functionality between the temporoparietal junction and ventro-medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *9*, 297–306. http://doi.org/10.1093/scan/nss139

Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G. V, & Blair, J. R. (2006). Caught in the act: the impact of audience on the neural response to morally and socially inappropriate behavior. *NeuroImage*, *33*(1), 414–21. http://doi.org/10.1016/j.neuroimage.2006.06.011

Fourie, M. M., Thomas, K. G. F., Amodio, D. M., Warton, C. M. R., & Meintjes, E. M. (2014). Neural correlates of experienced moral emotion: an fMRI investigation of emotion in response to prejudice feedback. *Social Neuroscience*, *9*(2), 203–18. http://doi.org/10.1080/17470919.2013.878750

Han, H., Glover, G. H., & Jeong, C. (2014). Cultural influences on the neural correlate of moral decision making processes. *Behavioural Brain Research*, *259*, 215–28. http://doi.org/10.1016/j.bbr.2013.11.012

Harenski, C. L., Antonenko, O., Shane, M. S., & Kiehl, K. A. (2010). A functional imaging investigation of moral deliberation and moral intuition. *NeuroImage*, *49*(3), 2707–16. http://doi.org/10.1016/j.neuroimage.2009.10.062

Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, *30*(1), 313–24. http://doi.org/10.1016/j.neuroimage.2005.09.034

Harrison, B. J., Pujol, J., López-Solà, M., Hernández-Ribas, R., Deus, J., Ortiz, H., … Cardoner, N. (2008). Consistency and functional specialization in the default mode brain network. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(28), 9781–6. http://doi.org/10.1073/pnas.0711791105

Heekeren, H. R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H.-P., & Villringer, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage*, *24*(3), 887–97. http://doi.org/10.1016/j.neuroimage.2004.09.026

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, *14*, 1215–1219. http://doi.org/10.1097/00001756-200307010-00005

Immordino-Yang, M. H., McColl, A., Damasio, H., & Damasio, A. (2009). Neural correlates of admiration and compassion. *Proceedings of the National Academy of Sciences*, *106*(19), 8021–8026. Journal Article. http://doi.org/10.1073/pnas.0810363106

Michl, P., Meindl, T., Meister, F., Born, C., Engel, R. R., Reiser, M., & Hennig-Fast, K. (2014). Neurobiological underpinnings of shame and guilt: a pilot fMRI study. *Social Cognitive and Affective Neuroscience*, *9*(2), 150–7. http://doi.org/10.1093/scan/nss114

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002). Functional networks in emotional moral and nonmoral social

judgments. *NeuroImage*, *16*, 696–703. http://doi.org/10.1006/nimg.2002.1118

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, *22*(7), 2730–2736. Journal Article.

Moll, J., De Oliveira-Souza, R., Garrido, G. J., Bramati, I. E., Caparelli-Daquer, E. M. A., Paiva, M. L. M. F., … Grafman, J. (2007). The self as a moral agent: linking the neural bases of social agency and moral sensitivity. *Social Neuroscience*, *2*, 336–352. http://doi.org/10.1080/17470910701392024

Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignácio, F. A., Bramati, I. E., Caparelli-Dáquer, E. M., & Eslinger, P. J. (2005). The moral affiliations of disgust: a functional MRI study. *Cognitive and Behavioral Neurology*, *18*, 68–78. http://doi.org/00146965-200503000-00008 [pii]

Moll, J., Eslinger, P. J., & Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. *Arquivos de Neuro-Psiquiatria*, *59*(3–B), 657–64.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is Morality Unified? Evidence that Distinct Neural Systems Underlie Moral Judgments of Harm, Dishonesty, and Disgust. *Journal of Cognitive Neuroscience*, *23*, 3162–3180. http://doi.org/10.1162/jocn_a_00017

Prehn, K., Wartenburger, I., Meriau, K., Scheibe, C., Goodenough, O. R., Villringer, A., … Heekeren, H. R. (2008). Individual differences in moral judgment competence influence neural correlates of socio-normative judgments. *Social Cognitive and Affective Neuroscience*, *3*(1), 33–46. Journal Article. http://doi.org/10.1093/Scan/Nsm037

Reniers, R. L. E. P., Corcoran, R., Völlm, B. A., Mashru, A., Howard, R., & Liddle, P. F. (2012). Moral decision-making, ToM, empathy and the default mode network. *Biological Psychology*, *90*(3), 202–10. http://doi.org/10.1016/j.biopsycho.2012.03.009

Robertson, D., Snarey, J., Ousley, O., Harenski, K., Bowman, E. D., Gilkey, R., & Kilts, C. (2007). The neural processing of moral sensitivity to issues of justice and care. *Neuropsychologia*, *45*(4), 755–766. Journal Article. http://doi.org/10.1016/j.neuropsychologia.2006.08.014

Schleim, S., Spranger, T. M., Erk, S., & Walter, H. (2011). From moral to legal judgment: the influence of normative context in lawyers and other academics. *Social Cognitive and Affective Neuroscience*, *6*(1), 48–57. http://doi.org/10.1093/scan/nsq010

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(13), 4741–9. http://doi.org/10.1523/JNEUROSCI.3390-13.2014

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain Responses to the Acquired Moral Status of Faces. *Neuron*, *41*, 653–662. http://doi.org/10.1016/S0896-6273(04)00014-5

Takahashi, H., Kato, M., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., & Okubo, Y. (2008). Neural correlates of human virtue

judgment. *Cerebral Cortex*, *18*, 1886–1891. http://doi.org/10.1093/cercor/bhm214
Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *NeuroImage*, *23*(3), 967–74. http://doi.org/10.1016/j.neuroimage.2004.07.054
Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 8235–8240. http://doi.org/10.1073/pnas.0701408104
Young, L., Nichols, S., & Saxe, R. (2010). Investigating the Neural and Cognitive Basis of Moral Luck: It's Not What You Do but What You Know. *Review of Philosophy and Psychology*, *1*(3), 333–349. http://doi.org/10.1007/s13164-010-0027-y
Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*(4), 1912–20. http://doi.org/10.1016/j.neuroimage.2008.01.057
Young, L., & Saxe, R. (2009). An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience*, *21*, 1396–1405. http://doi.org/10.1162/jocn.2009.21137
Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for "intuitive prosecution": the use of mental state information for negative moral verdicts. *Social Neuroscience*, *6*(3), 302–15. http://doi.org/10.1080/17470919.2010.529712

Table S2. Activation foci for each contrast

Region	BA	MNI coordinates			*t*	*k*
		x	y	z		
Moral Personal + Impersonal vs. Non-moral						
FDR applied						
Medial Frontal Gyrus	10, 24, 31, 32	-6	54	12	11.37	19742
Middle Temporal Gyrus	19, 22, 39, 40	54	-60	4	7.25	2250
Middle Temporal Gyrus	19, 39	-56	-72	14	7.52	888
Inferior Parietal Lobule	2, 40	-66	-32	30	6.42	729
Parahippocampa Gyrus	19, 36, 37	-28	-44	4	4.35	339
Insula	13	34	8	-6	4.78	178
Parahippocampa Gyrus		28	-14	-18	5.22	75
Superior Temporal Gyrus	22	54	6	4	4.06	60
Inferior Temporal Gyrus	20, 21	50	-6	-34	3.71	53
Insula	21	40	-8	-12	4.38	48
Fusiform Gyrus		46	-26	-18	4.19	39
Middle Frontal Gyrus	6	26	-12	60	3.63	35
Hippocampus		-24	-16	-16	3.76	28
Parahippocampa Gyrus	19, 37	30	-44	-10	3.95	26
Superior Frontal Gyrus	10	22	66	0	5.36	22
Culmen		22	-42	-18	3.13	4
Middle Temporal Gyrus	21	-64	-60	4	3.13	2
Temporal Lobe		-30	-52	16	3.08	2
Middle Temporal Gyrus		-40	-54	6	3.08	1
Temporal Lobe		48	-16	-26	3.07	1
Precuneus	19	22	-86	40	3.05	1
Precuneus		-6	-62	64	3.05	1
Cluster-wise correction applied						
Medial Frontal Gyrus	10, 24, 31, 32	-6	54	12	11.37	12185
Middle Temporal Gyrus	19, 22, 39, 40	54	-60	4	7.25	1165
Middle Temporal Gyrus	19, 39	-56	-72	14	7.52	632
Inferior Parietal Lobule	2, 40	-66	-32	30	6.42	442
RFT-FWE applied						
Medial Frontal Gyrus	9, 10, 32	-6	54	12	11.37	223
Medial Cingulate Cortex	31	-12	-28	38	8.13	14
Middle Frontal Gyrus		-30	28	42	8.93	8
Medial Frontal Gyrus	10, 32	-2	46	-6	8.46	8
Precuneus		-12	-42	44	7.90	4
Medial Cingulate Cortex		6	-24	42	7.90	3
Superior Frontal Gyrus	10, 32	-24	44	26	7.93	2

Middle Temporal Gyrus		-56	-72	14	7.52	1
Anterior Cingulate Cortex		2	26	28	7.51	1
Bonferroni's FWE						
Medial Frontal Gyrus	10	-6	54	12	11.37	63
Middle Frontal Gyrus		-30	28	42	8.93	2
Anterior Cingulate Cortex	32	8	42	8	8.88	1
SnPM						
Medial Frontal Gyrus	9, 10, 32	-6	52	16	8.69	264
Posterior Cingulate Cortex	23, 31	0	-54	24	7.19	41
Medial Cingulate Cortex	7, 31	-14	-30	42	7.81	26
Medial Cingulate Cortex	24, 31	4	-16	42	6.92	21
Superior Frontal Gyrus	10	-24	44	26	8.71	18
Paracentral Lobule	5	2	-42	52	6.67	14
Corpus Callosum		-6	-26	26	7.13	6
Posterior Cingulate Cortex		0	-56	6	6.48	6
Superior Frontal Gyrus		22	42	32	6.91	4
Subcallosal Gyrus	25, 34	8	4	-16	6.61	3
Medial Cingulate Cortex	31	8	-30	42	6.29	3
Corpus Callosum		-8	-42	18	6.75	2
Middle Temporal Gyrus	39	-56	-70	16	6.53	2
Superior Occipital Gyrus	19	-42	-82	26	6.47	2
Anterior Cingulate Cortex		10	38	10	7.01	1
Middle Temporal Gyrus	19	-58	-68	12	6.73	1
Medial Cingulate Cortex	24	-6	-8	38	6.37	1
Medial Cingulate Cortex		-6	-4	38	6.28	1
Moral Personal vs. Non-moral						
FDR applied						
Medial Frontal Gyrus	9, 10, 31, 32	8	40	8	12.06	17317
Superior Temporal Gyrus	19, 22, 39, 40	68	-42	10	6.40	1278
Middle Temporal Gyrus	19, 39	-56	-72	16	9.10	776
Inferior Parietal Lobule	40	-68	-30	28	5.59	426
Hippocampus	21, 35	24	-14	-14	5.47	194
Inferior Temporal Gyrus	20, 21	50	-6	-34	5.11	186
Superior Frontal Gyrus	6	8	16	64	4.35	179
Postcentral Gyrus	2, 5	-32	-46	68	6.20	153
Insula	13, 21	-36	-16	-2	4.86	103
Hippocampus	28, 35	-24	-14	-14	4.90	79
Superior Temporal Gyrus	6, 22	-54	-2	4	3.71	43
Thalamus		0	-12	0	4.02	35
Superior Temporal Gyrus	38	34	10	-32	4.51	24

Corpus Callosum		-16	-40	6	3.99	21
Insula	13	38	10	-8	3.44	16
Thalamus		12	-36	8	3.44	15
Middle Temporal Gyrus		-48	-16	-16	3.62	10
Superior Frontal Gyrus	10	22	66	0	3.53	10
Precentral Gyrus		54	6	6	3.41	9
Middle Frontal Gyrus		-20	-18	64	3.38	2
Middle Temporal Gyrus	21	68	-12	-12	3.52	1
Lentiform Nucleus		-22	-8	-6	3.52	1
Postcentral Gyrus		-44	-40	62	3.16	1
Parahippocampa Gyrus	34	16	-2	-18	3.15	1
Temporal Lobe		-40	-56	8	3.15	1
Extra-Nuclear		-2	-24	4	3.11	1
Precuneus	19	22	-86	40	3.11	1
Cluster-wise correction applied						
Medial Frontal Gyrus	9, 10, 24, 32	8	40	8	12.06	8573
Cingulate Gyrus	5, 7, 24, 31	6	-24	42	8.26	3904
Middle Temporal Gyrus	19, 39	-56	-72	16	9.10	529
Middle Temporal Gyrus	19, 22, 39	54	-62	8	6.22	296
Superior Temporal Gyrus	22, 40, 42	68	-42	10	6.40	238
RFT-FWE applied						
Medial Frontal Gyrus	9, 10, 32	8	40	8	12.06	715
Anterior Cingulate Cortex	24, 32	-4	34	20	8.18	28
Precuneus	23, 31	-4	-56	26	8.06	24
Middle Temporal Gyrus	39	-56	-72	16	9.10	10
Superior Frontal Gyrus	10	-22	44	26	8.57	10
Medial Cingulate Cortex	24	6	-24	42	8.26	10
Suballosal Gyrus	25	2	4	-16	8.24	9
Superior Frontal Gyrus		24	42	26	7.90	5
Medial Cingulate Cortex	24, 32	-2	-12	40	7.75	5
Precuneus		-12	-42	44	7.56	1
Bonferroni's FWE						
Medial Frontal Gyrus	9, 10, 32	8	40	8	12.06	217
Anterior Cingulate Cortex		-8	40	-2	9.26	6
Anterior Cingulate Cortex		-10	38	4	8.99	3
Middle Temporal Gyrus		-56	-72	16	9.10	1
SnPM						
Anterior Cingulate Cortex	9, 10, 24, 32	8	40	8	11.69	1502
Posterior Cingulate Cortex	23, 30, 31	8	-54	16	9.07	274
Medial Cingulate Cortex	24, 31	6	-24	42	8.55	106

Medial Cingulate Cortex	31	-14	-34	40	8.56	92
Superior Frontal Gyrus	10	24	42	26	8.68	57
Superior Frontal Gyrus	10	-24	46	26	8.73	43
Middle Frontal Gyrus		-30	30	42	7.13	6
Corpus Callosum		-8	-42	18	6.88	2
Suballosal Gyrus	25	-4	6	-14	6.72	1
Middle Frontal Gyrus		26	34	42	6.61	1
Moral Impersonal vs. Non-moral						
FDR applied						
Posterior Cingulate Cortex	5, 7, 24, 31	-12	-32	38	6.54	6905
Middle Temporal Gyrus	19, 22, 39, 40	66	-38	24	5.75	1774
Medial Frontal Gyrus	9, 10, 32	-4	52	10	6.54	965
Superior Frontal Gyrus	8, 9, 10	24	32	38	6.84	801
Superior Frontal Gyrus	8, 9, 10	-24	44	26	5.42	743
Inferior Parietal Lobule	1, 2, 3, 40	-62	-28	30	7.38	637
Culmen		-32	-50	-34	5.80	606
Middle Temporal Gyrus	19, 39	-56	-70	22	5.31	599
Insula	6, 13, 44	-42	0	6	5.26	209
Middle Frontal Gyrus	6	22	2	64	3.99	95
Anterior Cingulate Cortex	11, 25	-10	22	-8	4.27	45
Hippocampus		20	-40	10	4.43	44
Pons		18	-34	-36	4.10	37
Thalamus		-12	-26	6	4.02	37
Parahippocampa Gyrus	19, 36, 37	32	-46	-10	3.61	31
Superior Frontal Gyrus	10, 46	-24	62	22	3.94	30
Insula	13	40	6	-6	3.52	30
Medial Frontal Gyrus	11, 32	0	32	-12	3.88	28
Parahippocampa Gyrus		44	-28	-18	3.56	27
Superior Temporal Gyrus	22	52	0	4	3.65	18
Temporal Lobe	20	-40	-14	-24	4.15	17
Insula	13	-38	-22	30	3.72	16
Caudate		20	20	10	3.85	14
Corpus Callosum		2	-6	18	3.71	12
Parietal Lobe		30	-32	44	3.68	12
Thalamus		2	-22	6	3.57	12
Extra-Nuclear		26	-14	24	3.66	11
Thalamus		0	-12	2	4.17	9
Superior Frontal Gyrus	10, 46	-40	54	20	3.81	5
Caudate		-14	20	10	3.69	3
Temporal Lobe		-40	-52	2	3.49	3

Hippocampus		-18	-40	6	3.37	2
Superior Frontal Gyrus	10	24	66	0	3.36	2
Insula		34	-18	2	3.31	2
Parahippocampa Gyrus		28	-42	-4	3.37	1
Extra-Nuclear		24	12	14	3.35	1
Caudate		20	10	20	3.33	1
Culmen		0	-50	-4	3.32	1
Postcentral Gyrus	5	-28	-44	70	3.32	1
Insula		36	8	8	3.30	1
Cluster-wise correction applied						
Cingulate Gyrus	5, 7, 24, 31	-12	-32	28	6.54	2297
Posterior Cingulate Cortex	7, 23, 30, 31	6	-54	18	6.37	1205
Medial Frontal Gyrus	9, 10, 32	-4	52	10	6.54	633
Superior Frontal Gyrus	8, 9, 10	24	32	38	6.84	495
Middle Temporal Gyrus	19, 23, 22, 39	46	-72	18	5.63	459
Inferior Parietal Lobule	1, 2, 40	-62	-28	30	7.38	439
Middle Frontal Gyrus	8, 9, 10	-24	44	26	5.42	437
Inferior Parietal Lobule	22, 40	66	-38	24	5.75	432
Middle Temporal Gyrus	19, 39	-56	-70	22	5.31	397
Culmen	19, 37	-32	-50	-34	5.80	340
Anterior Cingulate Cortex	24, 32	0	30	28	5.83	324
RFT-FWE applied						
N/A						
Bonferroni's FWE						
N/A						
SnPM						
Inferior Parietal Lobule	40	-62	-28	30	7.38	13
Superior Frontal Gyrus		24	32	38	6.84	6
Moral Personal vs. Moral Impersonal						
FDR applied						
Medial Frontal Gyrus	9, 10, 24, 32	8	52	16	8.34	2394
Precuneus	7, 23, 31	-4	-56	26	8.17	305
Middle Temporal Gyrus	20, 21	48	0	-34	4.44	27
Superior Frontal Gyrus	8	26	40	48	5.10	23
Medial Cingulate Cortex	24	2	-16	40	4.12	21
Caudate		-16	22	8	4.81	16
Angular Gyrus	39	-50	-76	39	4.59	10
Superior Temporal Gyrus	39	58	-64	22	4.09	6
Extra-Nuclear		14	-4	-10	4.03	4
Superior Frontal Gyrus	6	12	22	64	3.94	4

Insula		-26	12	-14	3.96	2
Middle Frontal Gyrus		-22	34	-14	3.93	2
Medial Frontal Gyrus		-12	68	8	4.01	1
Middle Temporal Gyrus	9, 10, 24, 32	-56	-70	26	3.92	1
Cluster-wise correction applied						
Medial Frontal Gyrus	9, 10, 32	8	52	16	8.34	2596
Precuneus	7, 23, 31	-4	-56	26	8.17	345
RFT-FWE applied						
Medial Frontal Gyrus	9, 10	8	52	16	8.34	28
Medial Frontal Gyrus	10	-2	62	2	8.06	9
Posterior Cingulate Cortex	31	-4	-56	26	8.17	8
Medial Frontal Gyrus	10	-2	54	-2	7.36	1
Bonferroni's FWE						
N/A						
SnPM						
Medial Frontal Gyrus	9, 10, 32	10	50	16	9.66	141
Medial Frontal Gyrus	10	-2	60	0	8.48	83
Anterior Cingulate Cortex	10, 32	-10	40	4	8.23	77
Anterior Cingulate Cortex	24, 32	-2	32	20	7.60	33
Precuneus	31	-4	-56	26	8.32	25
Insula	47	-28	10	-16	7.05	6
Corpus Callosum		-10	20	18	7.16	3
Middle Temporal Gyrus	21	50	2	-30	7.23	2
Moral Impersonal vs. Moral Personal						
FDR applied						
Inferior Parietal Lobule	7, 19, 39, 40	32	-46	36	9.03	3301
Inferior Parietal Lobule	7, 19, 39, 40	-28	-60	46	6.55	2661
Middle Frontal Gyrus	6, 8, 9, 46	-38	8	32	6.35	1706
Declive		-4	-76	-28	7.01	1399
Middle Temporal Gyrus	20, 21, 22, 37	-48	-50	-4	6.90	794
Middle Temporal Gyrus	20, 37	52	44	-6	6.78	344
Middle Frontal Gyrus	6, 24, 32	24	6	54	4.87	289
Extra-Nuclear		24	12	34	4.78	226
Inferior Frontal Gyrus	8, 9	52	10	30	4.94	214
Medial Frontal Gyrus	6, 24, 32	-12	8	52	4.52	123
Insula	13	30	24	13	4.52	66
Middle Frontal Gyrus	46	46	42	30	4.17	62
Middle Frontal Gyrus	10, 47	-46	42	-4	4.51	47
Middle Occipital Gyrus	18	28	-84	0	3.66	45
Middle Frontal Gyrus		40	28	26	3.92	42

Inferior Frontal Gyrus	22, 44, 45	-60	12	8	3.61	36
Declive		-48	-54	-30	4.38	15
Pons		18	-38	-38	4.31	13
Cuneus		-22	-96	4	3.56	13
Middle Frontal Gyrus	10, 47	42	56	16	3.54	11
Precentral Gyrus		-44	6	12	4.06	10
Pons		0	-30	-36	3.80	9
Insula	13	-30	20	4	3.68	9
Cingulate Gyrus	32	6	20	46	3.59	6
Middle Frontal Gyrus	10, 47	-38	58	14	3.66	5
Temporal Lobe		34	-58	-4	3.45	5
Inferior Frontal Gyrus	45, 47	-58	32	0	3.60	4
Superior Temporal Gyrus		-56	-50	14	3.39	4
Cuneus	8	16	-82	16	3.66	3
Cuneus		22	-98	-4	3.39	3
Fusiform Gyrus	20	36	-38	-20	3.38	3
Temporal Lobe		30	-44	0	3.66	2
Orbitofrontal Cortex		46	46	-16	3.66	1
Middle Frontal Gyrus	10, 47	-30	64	14	3.51	1
Middle Frontal Gyrus	10, 47	-32	62	16	3.42	1
Extra-Nuclear		28	28	4	3.34	1
Cluster-wise correction applied						
Inferior Parietal Lobule	7, 19, 39, 40	32	-46	36	9.03	2621
Inferior Parietal Lobule	7, 19, 40	-28	-60	46	6.55	1979
Inferior Frontal Gyrus	6, 8, 9, 46	-38	8	32	6.35	991
Declive		-4	-76	-28	7.01	768
Middle Temporal Gyrus	19, 21, 22, 37	-48	-50	-4	6.90	467
RFT-FWE applied						
Angular Gyrus		32	-46	36	9.03	22
Bonferroni's FWE						
Parietal Lobe		32	-46	36	9.03	2
SnPM						
Angular Gyrus	7, 19, 39, 40	32	-46	36	9.03	203
Declive		-4	-76	-28	7.01	10
Temporal Lobe		-48	-50	-4	6.90	6
Temporal Lobe		52	-44	-6	6.78	6
Fastigium		8	-58	-26	6.70	5
Middle Temporal Gyrus	20	56	-48	-14	6.56	3
Superior Parietal Lobule		-28	-60	46	6.55	2

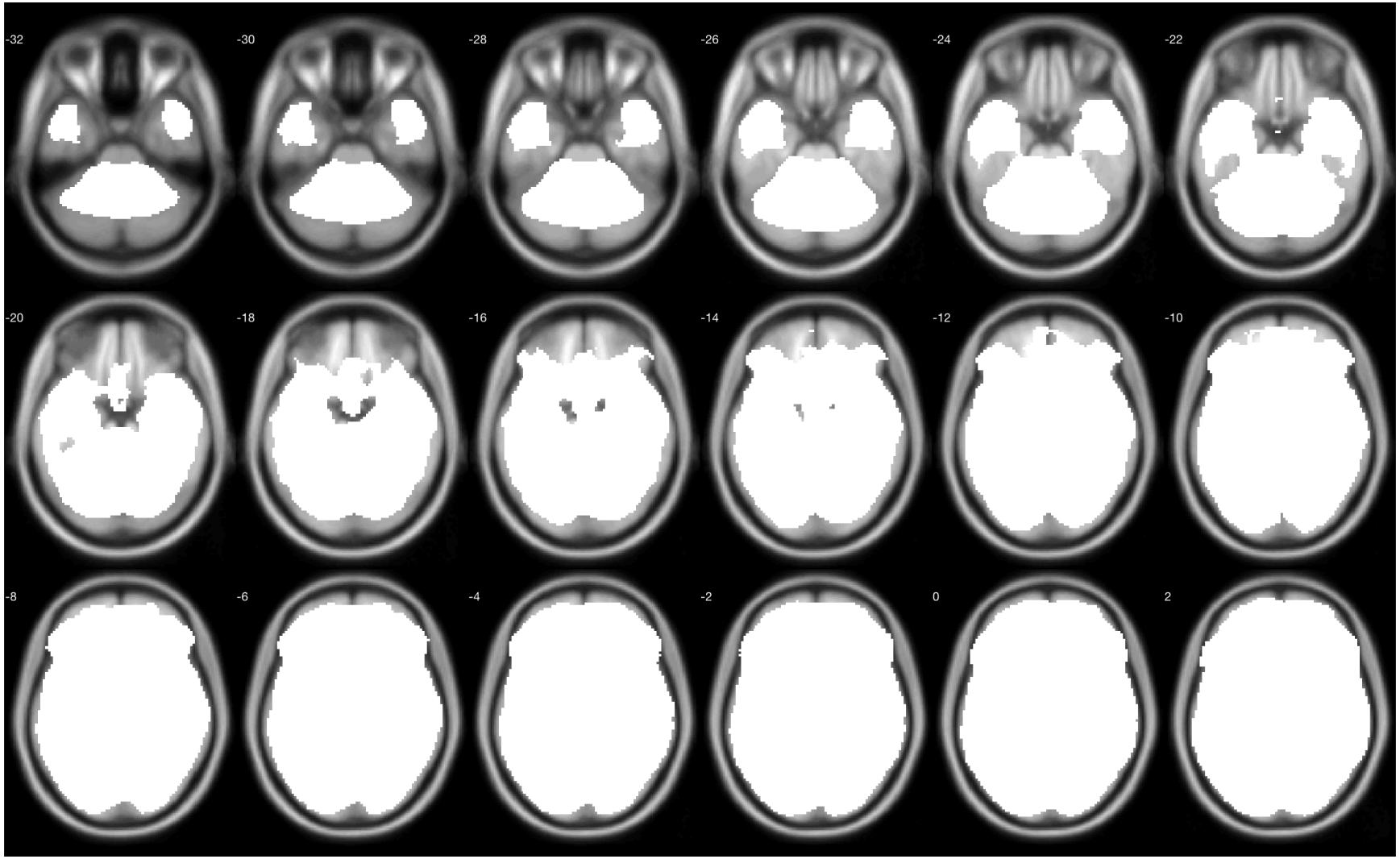

Figure S1. The implicit mask in SPM near the VMPFC.