



Published in final edited form as:

AJOB Neurosci. 2020 ; 11(4): 269–284. doi:10.1080/21507740.2020.1811798.

The dark side of morality:

Neural mechanisms underpinning moral convictions and judgments about violence

Clifford I. Workman^{1,2}, Keith J. Yoder¹, Jean Decety^{1,†}

¹Department of Psychology, University of Chicago, Chicago, IL, 60637, USA

²Department of Neurology, University of Pennsylvania, Pennsylvania, PA, 19104, USA

Abstract

People are motivated by shared social values that, when held with moral conviction, can serve as compelling mandates capable of facilitating support for ideological violence. The current study examined this dark side of morality by identifying specific cognitive and neural mechanisms associated with beliefs about the appropriateness of sociopolitical violence, and determining the extent to which the engagement of these mechanisms was predicted by moral convictions. Participants reported their moral convictions about a variety of sociopolitical issues prior to undergoing functional MRI scanning. During scanning, they were asked to evaluate the appropriateness of violent protests that were ostensibly congruent or incongruent with their views about sociopolitical issues. Complementary univariate and multivariate analytical strategies comparing neural responses to congruent and incongruent violence identified neural mechanisms implicated in processing salience and in the encoding of subjective value. As predicted, neuro-hemodynamic response was modulated parametrically by individuals' beliefs about the appropriateness of congruent relative to incongruent sociopolitical violence in ventromedial prefrontal cortex, and by moral conviction in ventral striatum. Overall moral conviction was predicted by neural response to congruent relative to incongruent violence in amygdala. Together, these findings indicate that moral conviction about sociopolitical issues serves to increase their subjective value, overriding natural aversion to interpersonal harm.

Keywords

Morality; moral conviction; violence; political neuroscience; computational neuroscience; social decision-making; ventromedial prefrontal cortex

Introduction

The view that morality is a set of biological and cultural solutions for solving cooperative problems in social interactions has recently gained traction (Curry, Mullins, and Whitehouse 2019). Yet, moral values do not always bear directly on cooperation and are, in fact,

[†]Corresponding author: decety@uchicago.edu.

Disclosure of interest: The authors report no conflict of interest.

Data availability statement: The data supporting the findings of this study, including the code for statistical models and analyses as well as the statistical maps and data used in the figures, are available from <https://github.com/cliffworkman/tdsom>.

sometimes *detrimental* to social cohesion (Smith and Kurzban 2019). Even when moral values appear *prima facie* motivated towards cooperation, they may, in fact, be motivated by self-interest (Kurzban, Dukes, and Weeden 2010; Smith and Kurzban 2019). Morality is concerned with more than just cooperation, sometimes also motivating people to behave in ways that are selfish and destructive. A complete theory of morality requires understanding both sides of the moral coin, so-to-speak. At the proximate level, morality involves multiple dimensions including harm aversion, intention understanding, empathic concern, values, and conformity to social norms, and relies on both implicit and deliberative information processing (Decety and Cowell 2018; Miller and Cushman 2018; Skitka and Conway 2019). While significant theoretical progress has been made in elucidating the mechanisms underlying these dimensions, much of this knowledge relates to the bright side of morality, like its role in promoting cooperation and minimizing conflict. Far less is known about morality's dark side – in particular, its role in justifying violence. If preventing violent ideological conflict is a priority at the societal level, a critical first step is understanding how individuals come to accept violence as appropriate in pursuing shared sociopolitical objectives.

Importantly, while violence is often described as antithetical to sociality, it can be motivated by moral values with the ultimate goal of regulating social relationships (Fiske and Rai 2014). In fact, most violence in the world appears to be rooted in conflict between moral values (Ginges 2019). Across cultures and history, violence has been used with the intention to sustain order and can be expressed in war, torture, genocide, and homicide (Rai and Fiske 2011). What distance separates accepting “deserved” vigilantism from others and justifying any behavior—rioting, warfare—as means to morally desirable ends? People who bomb family planning clinics and those who violently oppose war (e.g., the Weathermen's protests of the Vietnam War) may have different sociopolitical ideologies, but both are motivated by deep moral convictions.

Moral beliefs and values differ from other attitudes in important ways. Our most highly valued moral beliefs are those held with extreme *moral conviction*. Moral values exert a powerful motivational force that vary both in direction and intensity, guide the differentiation of just from unjust courses of action, and direct behavior towards desirable outcomes (Halevy et al. 2015), and serve as compelling mandates regarding what one “ought” to do in certain circumstances. This can be seen, for instance, in expressions of support for war, which depend primarily on perceptions of moral righteousness rather than strategic efficacy. Beyond the level of the self, moral convictions also dictate beliefs about what *others* ought to do. Ethical vegans, for instance, are rarely satisfied with the status quo, instead wishing to bring societal attitudes towards animal welfare in line with their own (Thomas et al. 2019). When the morally convicted are confronted with societal attitudes out of sync with their moral values, some may find this sufficiently intolerable to justify violence against those who challenge their beliefs (Skitka 2010). The sorts of beliefs capable of being tinged with moral conviction are highly variable, which underscores the urgent need to characterize how people distinguish courses of action as being morally just or unjust. This view aligns well with theoretical accounts emphasizing the importance of morality in regulating the behavior of individuals in group settings (Rai and Fiske 2011; Tooby and Cosmides 2010).

Moral conviction provides the psychological foundation and justification for any number of extreme acts intended to achieve morally desirable ends. In contrast to social rules (e.g., appropriate greetings for different situations), which are matters of convention that vary by locale, moral rules are often viewed as universal and compulsory. People are often objectivists about their moral principles (i.e., believing a moral rule is entirely true or false), and are intolerant of behavior that deviates from them, regardless of other people's beliefs and desires and with little room for compromise (Skitka 2010). The morally convicted think that others *ought* to universally agree with their views on right and wrong, or at least that they *would* agree if only they knew the facts (Skitka, Bauman, and Sargis 2005). Moral convictions motivate judgments of right and wrong as well as the willingness to support and even to participate in violent collective action (Ginges and Atran 2009).

Moral conviction may function by altering the decision-making calculus through the subordination of social prohibitions against violence, thereby requiring less top-down inhibition. This hypothesis holds that moral conviction facilitates support for ideological violence by increasing Utilitarian commitments to a "greater good" even at the expense of others. An alternative hypothesis is that moral conviction increases the *subjective value* of certain actions, where violence in service of those convictions is underpinned by deontological judgments about one's moral responsibilities to sociopolitical causes.

Functional neuroimaging is well suited to examine these hypotheses with forward inference (Henson 2006) since the mechanisms they purport are associated with partially dissociable neural circuits and computations (Buckholz and Marois 2012; Decety and Yoder 2017). The central executive network, anchored by the dorsolateral prefrontal cortex (dlPFC), is essential for using social norms to exert top-down control over moral judgment (e.g., whether to harm others; Buckholz et al. 2008; Ruff, Ugazio, and Fehr 2013). The first hypothesis would predict that moral conviction is associated with decreased dlPFC responding when individuals believe it is appropriate to eschew social prohibitions against violence committed in service of a sociopolitical position held with moral conviction. The second hypothesis would predict that moral conviction is associated with increased activity in ventral striatum (VS) and ventromedial prefrontal cortex (vmPFC) during appropriateness evaluations of violence in service of one's sociopolitical goals, reflecting greater subjective value placed either on one's goals or on the use of violence in achieving them (Krastev et al. 2016; Kable and Glimcher 2007; Delgado et al. 2016). The vmPFC, VS, and amygdala comprise a reward circuit implicated in representing subjective value (Ruff and Fehr 2014). The vmPFC is critical for the representation of reward and value-based decision-making, through interactions with ventral striatum and amygdala, and as such is a key node of subcortical and cortical networks that underpin moral decision-making (Hiser and Koenigs 2018). Finally, since moral conviction is expected to make social issues more important to oneself, it is also predicted to demonstrate associations with areas critical for self and other-related information processing (e.g., dorsomedial prefrontal cortex [dmPFC] and temporoparietal junction [TPJ]; Carter and Huettel 2013; Decety and Lamm 2007).

Despite the suitability of functional neuroimaging to address important questions about how people make third-party moral evaluations, much of the cognitive neuroscience literature over the past 20 years has focused on sacrificial moral scenarios that are lacking in

ecological validity (Bauman et al. 2014; Moll et al. 2005; Casebeer 2003). Judgments made in response to such scenarios are poor predictors of real-world behavior (Bostyn, Sevenhant, and Roets 2018; FeldmanHall et al. 2012), perhaps owing to the implausibility of the scenarios themselves. Most people have limited relevant experience to draw on when judging, for instance, whether they would push someone in front of a trolley if they thought this would save the lives of several others. In contrast, moral convictions about sociopolitical issues are deeply personal and are known to increase the risk of violent behavior in support of those convictions (Ginges and Atran 2009; Decety, Pape, and Workman 2018).

In the current study, participants underwent functional MRI (fMRI) scanning while evaluating the appropriateness of violent protests about real-world issues that were either congruent or incongruent with their own sociopolitical views. After seeing photographs of real protests, and learning the ostensible motivations for those protests, participants rated how appropriate they found the protestors' violence (Figure 1). Complementary fMRI analysis strategies characterized the neural response associated with beliefs about the appropriateness of sociopolitical violence and identified patterns of activation that distinguished violence that was either congruent or incongruent with participants' own values and beliefs. Since believing oneself to be a victim of injustice has been linked to supporting violence (Decety, Pape, and Workman 2018), relations between dispositional justice sensitivity and neural responses that distinguished the congruency of sociopolitical violence were also assessed.

Materials and Methods

Participants

A sample of 41 healthy adults (22 female) aged between 18 and 38 years (mean \pm SD: 22.8 \pm 5.3) was enrolled in the study, which was approved by the University of Chicago Institutional Review Board. Participants were recruited from the Chicago metropolitan area using print and online advertisements and were compensated for their time. A target sample size of 40 was chosen in view of recent evidence that samples of this size generally provide sufficient power to detect large group-level effects with fMRI (Geuter et al. 2018; of note, however, the authors of this study examined samples of $N = 20$ and $N = 40$ —not $N = 30$ —with the jump in power between these sample sizes likely to benefit intermediate sample sizes as well; e.g., Yarkoni 2009). Of the 41 people enrolled into the study, 34 identified themselves as having liberal sociopolitical views. Since most participants were liberal, the remaining seven non-liberal participants were excluded from further analysis ($N = 5$ moderates, $N = 2$ conservatives). Additional information about the criteria used to identify and exclude these participants is given in the next section. Two additional participants were excluded for excessive head motion during scanning (over 3mm translational or 3° rotational motion). The final sample consisted of 32 participants (18 female) aged between 18 and 38 years (23.0 \pm 5.0).

Experimental Design and Statistical Analysis

Online Survey—Participants completed an online survey in advance of their scanning visit, which included a demographics questionnaire with questions about age, sex, ethnicity,

level of education, income, and religiosity. The survey also included questions about political party affiliation (e.g., Democrat, Republican), frequency of political engagement, and beliefs about the appropriate and justified use of violence. Overall political orientation was calculated as the mean response to the following questions: “How would you describe your political orientation with regard to social issues?” and “How would you describe your political orientation with regard to economic issues?” (1 = “Very liberal”, 7 = “Very conservative”). Participants with overall political orientation scores below 4 were categorized as liberals, scores equal to 4 were categorized as moderates, and scores greater than 4 were categorized as conservatives.

Opinions on specific social and political issues (e.g., free speech, social justice) were assessed with an updated 30-item version of the Wilson-Patterson Inventory (adapted from Fessler et al., 2017; Appendix S1). Half of these issues are generally supported by liberals (e.g., abortion rights, foreign aid) and half by conservatives (e.g., tax cuts, small government). After reporting their agreement with each issue, participants indicated whether they had ever supported the issue passively (e.g., with donations) and/or actively (e.g., by engaging in protests) with “yes” or “no” responses. Moral convictions were measured with the following item (Skitka and Morgan 2014): “To what extent is your position on this a reflection of your core moral beliefs and convictions?” (1 = “Not at all”, 7 = “Very much”; Cronbach’s $\alpha = 0.887$, $M = 3.320$, $SD = 0.555$). Finally, the survey also included the 40-item Justice Sensitivity Inventory (JSI) to measure trait sensitivity to injustices against oneself or others (Schmitt et al. 2010).

Functional MRI Scanning Task—Participants completed a novel task during fMRI scanning in which they viewed a series of photographs depicting violent protests from throughout the United States and judged the appropriateness of the protestors’ violence (see Figure 1 for a graphical overview of the task). First, after jittered fixation (average 500 ms), participants saw a short text description of the sociopolitical issue that purportedly motivated the protest (4 s). A subset of 20 social and political issues from the updated Wilson-Patterson Inventory was used in this task, with 10 supported by liberals and 10 by conservatives. See Table S1 for the complete list of issues. Each issue was paired with an image of either a “thumbs up” or “thumbs down” to indicate the protestors’ support or opposition to the issue. Each issue appeared once per run in each of two separate functional runs, once with a “thumbs up” and once with a “thumbs down”, such that participants saw a total of 40 issues. In an additional 20 trials, the issue and thumbs up or down were replaced with the text “RIOT,” resulting in a total of 60 trials that comprised the task (30 per run with two functional runs per participant).

Next, after jittered fixation (average 500 ms), participants saw one of 60 photographs depicting violent protests (4 s). The photographs were identified through web searches and edited to obscure any features revealing the true causes of the protests (e.g., banners and signs, clothing featuring political slogans). The photographs were distributed in a random and non-repeating fashion across the 60 trials of the task: 20 trials where the violence was purportedly congruent with the participant’s own liberal views (i.e., in support of a liberal issue or in opposition to a conservative issue), 20 trials where it was purportedly incongruent (i.e., in support of a conservative issue or in opposition to a liberal issue), and 20 “riot” trials

(prior to completing the task, participants were instructed that the violence in photographs preceded by the text “RIOT” was unrelated to specific social or political issues; these data were not included in the analyses reported below and are therefore excluded from further discussion).

Importantly, by virtue of their purported connection to divisive sociopolitical issues as described above, these images depicted the sorts of acts of ideologically motivated violence this study sought to examine. The order of the trials was pseudo-randomized (no more than two contiguous trials of the same type within each run).

After each photograph, participants had 3 s to rate the appropriateness of the violence in the photograph along a 7-point scale (1 = “Extremely inappropriate, 7 = “Extremely appropriate”) using three buttons on a five-button box (pointer finger on the right hand to move the selection box left, ring finger to move it right, and middle finger to submit). The average participant failed to respond only once across all 60 trials (misses: 1.13 ± 1.66), suggesting participants were engaged with the task.

Each of the 60 trials was separated by periods of fixation (12 s). Taken together, the timing of the fMRI task was designed to balance overall exposure to each stage of stimulus presentation against the disambiguation of individual trials. Prior to starting the task, participants completed at least 6 practice rounds with the option to complete more if the instructions were unclear.

Image Acquisition and Processing—MRI scanning was conducted with a Philips Achieva 3T scanner (Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel SENSE head coil located at the University of Chicago’s MRI Research Center. Functional runs were collected in interleaved 2.5 mm thick transverse slices using a single-shot EPI sequence (TR = 2500 ms; TE = 22 ms; flip angle = 80° ; SENSE factor = 2; matrix = 80×78 ; field-of-view = 200 mm x 200 mm; voxel size = 2.5 mm^3 , slice gap: 0.5 mm). A Z-Shim sequence was used to recover signal in areas adversely affected by magnetic field inhomogeneities owing to their adjacency to the sinuses. Two functional runs lasting 8 mins and 32 s each were acquired during completion of the task. Alertness was monitored throughout each run using a long-range mounted EyeLink eye-tracking camera (SR Research, Ontario, Canada) trained on the left eye. In addition to these two functional runs, a 3-dimensional T1-weighted magnetization-prepared rapid-acquisition gradient-echo (MPRAGE) structural scan lasting 5 mins 39 s was also acquired for each participant (TR = 8 ms; TE = 3.5 ms; matrix = 256×244 ; voxel size = $.9 \text{ mm}^3$). Although not reported here, an additional functional run was acquired during completion of the Emotional Stroop Switching Task lasting 10 mins 23 s, as was a resting-state fMRI scan lasting 8 mins 14 s. In total, each MRI scan took approximately 45 minutes per participant.

Pre-processing of the MRI data was performed with SPM12 (Wellcome Centre for Human Neuroimaging, London) in MATLAB (MathWorks). First, the functional runs underwent frame-by-frame realignment to adjust for head motion. Next, the MPRAGE images were co-registered to the mean EPIs generated during frame-by-frame realignment and then segmented. The nonlinear deformation parameters resulting from segmentation were applied

to the EPIs to warp them into standard Montreal Neurological Institute (MNI) space. After normalization, the EPIs were smoothed with a 6 mm full width at half-maximum (FWHM) Gaussian kernel. High-motion volumes identified with the ArtRepair toolbox (greater than 0.5 mm between TRs) were replaced with interpolated volumes and partially de-weighted in first-level models (Mazaika et al. 2009). Low frequency artefacts were removed during first-level model estimation with a high-pass filter (128 s cut-off).

Univariate fMRI Analysis—At the first-level, general linear models were constructed for each participant. Regressors defined the 4 s epochs in which the two categories of violent protest photographs (Congruent, Incongruent) were presented to participants. Regressors were convolved with the canonical hemodynamic response function and resulting models were corrected for temporal autocorrelation using first-order autoregressive models. Images reflecting the contrast between the Congruent and Incongruent conditions (Congruent > Incongruent, Incongruent > Congruent) were then used for region of interest (ROI) analysis (see below) and were also separately submitted to a one-sample t-test for voxelwise analysis at the second-level. This t-test was used to identify brain areas preferentially activated when viewing congruent compared to incongruent acts of sociopolitical violence and vice versa. A cluster-level FWE-corrected threshold of $p < 0.01$ was achieved using an uncorrected voxel-level threshold of $p < 0.005$ in conjunction with an extent threshold of 98 voxels (estimated with 10,000 Monte Carlo simulations in the 3dClustSim routine in AFNI; Forman et al., 1995).

Parametric Modulation fMRI Analysis—The first-level models were then re-estimated with mean-centered appropriateness ratings for each category included as parametric modulators. Contrast images reflecting positive associations between hemodynamic response amplitude and appropriateness ratings for the congruent and incongruent conditions were submitted to a flexible factorial model at the second-level. This analysis identified brain areas in which this parametric relationship maximally distinguished between the congruent and incongruent conditions. The above procedure was repeated with moral conviction scores for each issue included as parametric modulators into the first-level models. A cluster-level FWE-corrected threshold of $p < 0.01$ was achieved using an uncorrected voxel-level threshold of $p < 0.005$ in conjunction with an extent threshold of 91 voxels (estimated with 3dClustSim).

Multivariate fMRI Analysis—In contrast to the univariate approach described above, which is used to localize and measure activity in clusters of voxels, MVPA is concerned with identifying relations between voxels that optimally distinguish experimental conditions. These patterns arguably approximate neural “representations”, or the neural population codes that store information in the brain (Haxby et al., 2014; cf. Ritchie et al., 2017). In order to provide a thorough characterization of the features distinguishing the Congruent and Incongruent conditions, complimentary univariate and multivariate analyses are reported below.

For MVPA, each of the two runs was split in half by removing the central 12 s period of fixation in order to create a total of four “chunks” (minimum two Congruent and Incongruent trials in each chunk). Each chunk was separately realigned, co-registered, and

re-sliced, but not smoothed or normalized. The same general linear model as in the univariate analysis was applied separately for each chunk, producing eight beta maps (Congruent and Incongruent from four blocks). A 3-voxel radius moving searchlight was applied, such that every brain voxel served as the center of a sphere (or spherical segment for voxels near the boundary of the brain). Beta weights within each sphere were used to train a Support Vector Machine (SVM) classifier to distinguish between Congruent and Incongruent targets. Leave-one-chunk-out cross-validation was applied to determine classifier accuracy. Participant's native-space accuracy maps were then normalized to MNI template space and smoothed (4mm FWHM). In order to identify reliable clusters across participants with above chance classification accuracy, normalized accuracy maps were subjected to threshold-free cluster enhancement. An FWE-corrected threshold of $p < 0.01$ was determined using 10,000 Monte Carlo simulations.

Region of Interest Analysis—Spheres of 6 mm in diameter were placed in the following regions, discussed in the introduction for their theoretical relevance to moral conviction: 1. dlPFC (MNI coordinates: $\pm 39, 37, 26$; Yoder and Decety, 2014), 2. dmPFC / dorsal anterior cingulate cortex (dACC; $\pm 10, 34, 24$; Seeley et al., 2007), 3. vmPFC ($\pm 10, 44, -8$; Lebreton et al., 2009), 4. VS ($\pm 8, 8, -4$; Lebreton et al., 2009), and 5. TPJ ($\pm 62, -54, 16$; Yoder and Decety, 2014). In addition, 6. an amygdala mask was extracted from the Automatic Anatomical Labelling atlas (Tzourio-Mazoyer et al. 2002). Parameter estimates (i.e., percent signal change) were extracted from ROIs with the MarsBaR toolbox (Brett et al. 2002).

Relations between the extracted parameter estimates and moral conviction were analyzed with non-parametric Spearman rank correlations in RStudio version 1.2.5001 and considered significant below a corrected $\alpha = 0.05 / 6 \text{ ROIs} = 0.008$ (two-tailed). Overall moral conviction was calculated as the mean difference in the extent to which people viewed congruent relative to incongruent issues as central to their moral beliefs ($M = 0.357, SD = 0.504$). Additional exploratory analyses examined relations between the parameter estimates and scores on the victimization subscale of the JSI. Larger victimization scores ($M = 38.438, SD = 10.797$) reflect heightened sensitivity to feeling like the victim of injustices.

Results

Increased hemodynamic response to photographs depicting violent protests that were purportedly congruent relative to incongruent with participants' sociopolitical views was detected in left dmPFC extending into right dmPFC and bilateral pre-supplementary motor area, right dlPFC (BA 46), left dlPFC (BA 9), right aINS extending into lateral orbitofrontal cortex (BA 47), right PCC (BA 23) extending bilaterally, and left precuneus (BA 7) (Table 1). The reverse contrast, which sought to identify larger responses to incongruent compared to congruent violence, was not significant.

Above chance classification using MVPA was possible with searchlights placed in several cortical areas (Table 2; Figure 2) including a large contiguous cluster of significant voxels spanning bilateral parietal and retrosplenial cortices, lateral prefrontal cortex, and midline structures including dACC, dmPFC, and vmPFC. High classification accuracy was also obtained in bilateral insula and inferior frontal gyri. Significant MVPA clusters overlapped

with the univariate analysis in bilateral dIPFC, TPJ, and ventrolateral prefrontal cortex (vlPFC).

The next analysis identified regions in which the parametric relationship between amplitude of hemodynamic response to sociopolitical violence and in-scanner appropriateness ratings maximally differentiated between the congruent and incongruent conditions. Stronger positive associations between appropriateness and neural response amplitudes for congruent compared to incongruent violence were found in right posterior medial prefrontal cortex (pmPFC) extending into left pmPFC and bilateral dACC, left parahippocampal gyrus extending into left hippocampus, left VS extending into right VS and bilateral sgACC as well as left vmPFC, left middle occipital and medial occipitotemporal gyri, left middle temporal gyrus, left dmPFC extending into right dmPFC, and bilateral PCC (Table 3; Figure 3). No regions showed stronger positive associations for the incongruent condition compared to the congruent condition.

A second parametric modulation analysis identified regions that maximally differentiated congruency on the basis of relations between amplitude of hemodynamic response to violent protests in support of sociopolitical issues and participant's moral conviction scores corresponding to these issues. Stronger positive associations between moral conviction and neural response amplitudes for congruent compared to incongruent violence were found in bilateral middle occipital gyri extending into bilateral fusiform gyri, and in left VS extending into left sgACC (Figure 4). The difference in moral conviction scores for congruent compared to incongruent violence did not correlate significantly with the same difference in in-scanner appropriateness ratings ($r_s(32) = -0.119$, $p = 0.518$).

Finally, overall moral conviction was negatively associated with signal change in amygdala ($r_s(32) = -0.488$, $p = 0.005$). Scores on the victimization subscale of the JSI correlated positively with signal change in dmPFC / dACC ($r_s(32) = 0.376$, $p = 0.034$), although this result would not survive correction for multiple comparisons.

Dispositional measures of justice sensitivity didn't yield any significant correlation with neural responses after controlling for multiple comparisons. This strengthens the importance of examining issue-level moral convictions, rather than person-level personality factors.

Discussion

People are motivated by beliefs and values that, when held with moral conviction, serve as compelling mandates capable of facilitating decisions to support ideological violence. With radicalization and support for sociopolitical violence seemingly on the rise around the world (Villasenor 2017), it is critical to increase understanding of how moral beliefs and values coalesce to motivate and normalize the acceptance of interpersonal harm as an appropriate means of furthering one's sociopolitical goals. This study examined alternative hypotheses about moral conviction's role in modulating beliefs about the appropriateness of sociopolitical violence; namely, either through reduced top-down control over expressions of support, or by increasing the subjective value of violence. Decision neuroscience has consistently identified a small number of areas associated with value-based choice (Levy and

Glimcher 2015; Rangel and Clithero 2018; Bartra, McGuire, and Kable 2013). This distributed network includes the vmPFC, cingulate cortex, insula, VS, and amygdala. There is broad consensus that specific subregions in prefrontal cortex, including vmPFC and dlPFC, together with the VS, are implicated in encoding aspects of subjective value (Bartra, McGuire, and Kable 2013).

Stronger beliefs about the appropriateness of congruent relative to incongruent sociopolitical violence showed a positive parametric relationship with the strength of neural response to photographs ostensibly depicting these acts of violence in the reward circuit (i.e., a cluster extending from VS into vmPFC). When in-scanner appropriateness ratings were not modelled, contrasting the congruency conditions failed to localize any significant activations to VS or vmPFC. This provides further evidence for the notion that beliefs about sociopolitical violence are specific to particular social and economic issues rather than monolithic political identity. The VS and vmPFC appear to modulate the online expression of pro-violence attitudes, which is in keeping with previous work linking these regions to political radicalism (Cristofori et al. 2015; Zamboni et al. 2009) and in tracking the magnitude of third-party punishment (Buckholz et al. 2008). The willingness to fight and die on behalf of one's sociopolitical in-group was found to be associated with activation of the vmPFC and its functional connectivity to the dlPFC in Pakistani participants who support the Kashmiri cause (Pretus et al. 2019). The VS and vmPFC are connected via corticostriatal circuits, with afferents from vmPFC providing input to VS (Berendse, Graaf, and Groenewegen 1992). Although both VS and vmPFC are sensitive to subjective value, the vmPFC is additionally implicated in integrating expected values and outcomes and regulating the deployment of effort in the pursuit of those outcomes (Gourley et al. 2016). This functional distinction provides insight into the finding that patterns of activation in a vmPFC cluster, which overlapped with the cluster that varied parametrically with appropriateness ratings, reliably distinguished the congruent from incongruent conditions whereas VS did not (Figures 2 and 3). These patterns of activation in the vmPFC may represent the expected value of outcomes from congruent compared to incongruent violence, also reflected in the in-scanner appropriateness ratings.

Moral conviction scores for each of the issues that purportedly motivated the congruent versus incongruent violent protests showed a positive parametric relationship with hemodynamic response in a cluster extending from VS into subgenual cingulate but not vmPFC. The VS and vmPFC both receive afferent projections from the amygdala (Ongür, Price, and Ongur 2000; Alheid 2006). Overall moral conviction was negatively associated with percent signal change in amygdala for congruent relative to incongruent sociopolitical violence. These findings fit with previous work implicating the amygdala as well as the VS in perceiving third-party interpersonal harm (Decety, Michalska, & Kinzler, 2012; Yoder, Porges, & Decety, 2015) or merely imagining intentionally hurting another person (Decety and Porges 2011). Blunted amygdala responding was detected while participants judged political violence in support of their favored causes, suggesting participants found the prospect of congruent violence less aversive. The amygdala is also implicated in affective learning (Delgado et al. 2008), possibly by mediating loss aversion, with amygdala lesions shown to impair reward-related behavior (Janak and Tye 2015). Together, the findings from the current study point to an affective learning-based neural mechanism that generates and

sustains extreme sociopolitical beliefs that, when held with moral conviction, may lead to finding violence in support of those views less objectionable.

The centrality of emotion to moral conviction was underscored in a recent study, which found that moral emotions (e.g., guilt, disgust, outrage), as well as nudging, correlated with moral convictions about eating meat (Feinberg et al. 2019). Moral emotions play a critical role in motivating behaviors linked to the welfare of others (Curry, Mullins, and Whitehouse 2019), with moral motivations emerging from the integration of motivational/emotional states and associated goals (Moll et al. 2005). Moral emotions are underpinned by fronto-temporo-mesolimbic circuits comparable to those identified in this study (Zahn et al. 2009). A similar fronto-temporo-mesolimbic circuit has been found to support facets of social cognition in non-human primates, including cooperation and competition (Wittmann, Lockwood, and Rushworth 2018), suggesting much of this neural circuitry has been preserved since the evolutionary divergence of humans from non-human primates. However, while violence is a natural part of life for our closest relatives—the chimpanzees—in which lethal aggression seems to be an adaptive strategy (Wilson et al. 2014), violence motivated by moral convictions, which is seen as both proper and necessary for regulating social relationships, is specific to humans.

Affiliative emotions have been shown to activate VS and basal forebrain structures such as the septo-hypothalamic area (Krueger et al. 2007; Moll et al. 2012), a region implicated in social conformity (Stallen, Smidts, and Sanfey 2013). Damage to the ventromedial prefrontal and lateral orbitofrontal cortices results in an insensitivity to feelings of guilt quantified with a battery of economic games (Krajbich et al. 2009). The processing of anticipated guilt and of moral transgressions engages overlapping regions in vmPFC and amygdala, respectively (Seara-Cardoso et al. 2016). The VS and subgenual frontal cortex have been linked to the excessive guilt and anhedonia characteristic of major depressive disorder (Workman et al. 2016) and to the lack of empathy characteristic of psychopathy (Decety, Skelly, and Kiehl 2013), with normal variation in psychopathic traits shown to track activity in the subgenual cingulate cortex (Wiech et al. 2013).

Beyond finding a link between moral conviction and increased hemodynamic response in regions associated with valuation, it is notable that activity in the dlPFC, a region involved in overriding prepotent responses and applying social norms (Buckholz and Marois 2012), was not significantly associated with moral conviction. This is perhaps surprising in view of previous work demonstrating greater dlPFC responding during evaluations of interpersonal assistance compared to interpersonal harm (Yoder and Decety 2014) and implicating the dlPFC in third-party punishment (Buckholz et al. 2008). Although failure to detect an effect may stem from limitations in sensitivity, rather than revealing the absence of an effect *per se*, the weight of the evidence presented above suggests that moral conviction does not facilitate the subordination of other norms (e.g., social prohibitions on violence). Interestingly, the multivariate analysis identified a cluster in the neighboring vlPFC that overlapped with the univariate analysis. Activation in the vlPFC has been shown to occur alongside dlPFC and other regions central to decision-making, including vmPFC, striatum, and posterior parietal cortex (Chung, Tymula, and Glimcher 2017). The vlPFC has been linked to emotion

regulation, with pathways through VS and amygdala, respectively, predicting better or worse chances for successful reappraisal (Wager et al. 2008).

The positive parametric relationship with appropriateness ratings that characterized the vmPFC in this study may appear to contradict findings reviewed above that indicate a central role of the vmPFC in giving rise to prosocial moral emotions, specifically (Moll and de Oliveira-Souza 2007). Decisions to act violently can be motivated, not only by a desire to act antisocially towards outgroup members, but also by a desire to behave prosocially towards in-group members (Molenberghs et al. 2015; Domínguez et al. 2018; Molenberghs et al. 2016). This view is consistent with the gene-culture coevolution theory, which accounts for human other-regarding preferences like fairness, the capacity to empathize, and salience of morality (Gintis 2011). Mathematical modeling of social evolution, together with evidence from anthropological studies, indicates that intragroup (tribal) motivations to help and cooperate with others co-evolved through intergroup competition over valuable resources (Boyd and Richerson 2009). During the late Pleistocene, groups that had higher numbers of prosocial people worked together more effectively and were thusly able to outcompete, passing genes relating to adaptive behaviors transmitted to the next generation and resulting in the spread of the hyperprosociality that characterizes our species (Bowles 2012).

Blame-related feelings may serve as proximate psychological mechanisms that act to regulate violent behavior. The tendency to engage in blame directed at oneself or others is shaped by relevant dispositions, such as the propensity to feel moral outrage in response to feeling victimized (Gollwitzer and Rothmund 2011). In this study, scores on the victimization subscale of the JSI correlated positively with signal change in dmPFC / dACC.

Inferences about the mental states of others are enabled by the dmPFC (Apps & Sallet, 2017). This area demonstrates sensitivity to information about normative, but not subjective, value (Apps & Ramnani, 2017), which is consistent with meta-analytic evidence suggesting the dmPFC mediates in-group favoritism (Volz, Kessler, and von Cramon 2009). Social norms exert powerful influences over moral feelings, judgments, and behaviors (Cikara and Van Bavel 2014). These effects persist even for norms that do not reflect the actual preferences of in-groups (Pryor, Perfors, and Howe 2019). Recent accounts suggest the dmPFC integrates information about rewards and beliefs (Rouault, Drugowitsch, and Koechlin 2019), with conformity potentially reflecting beliefs about the value of “fitting in.” Although social conformity may exert undesirable effects, leading to e.g. mob violence (Russell 2004), a compelling line of research suggests social information may be able to modulate the extremity of one’s views (Kelly et al. 2017). Together, these findings point towards a potential neurocognitive target for tempering moral conviction about sociopolitical issues likely to be accompanied by the acceptance of violence as a means to an end.

The efficacy of interventions designed to prevent ideological violence hinges on first understanding the risk factors that facilitate the transformation of violence from morally forbidden into morally appropriate. Although the findings described above point towards an affective learning-based mechanism, burgeoning behavioral research indicates that cognitive empathy, rather than emotional empathy (Simas, Clifford, and Kirkland 2019), motivates justice concerns (Decety and Yoder 2016) and may reduce polarization in highly politicized

contexts (Fido and Harper 2018). Many have argued (Haidt 2012), and at least some research suggests, that a deeper understanding of our political opponents reduces polarization. Furthermore, people tend to overestimate the differences that separate them from their political opponents (Levendusky and Malhotra 2016), and correcting these errors may reduce polarization (Ahler 2014). Interventions centered on reasoning – especially on understanding others – may be our best bet for bridging the ideological divide.

In recent months, reports of violence perpetrated by people from across the ideological spectrum provide a sobering reminder that violence has no political allegiance. For all the ways liberals and conservatives ostensibly differ (e.g., Waytz et al. 2019), small yet growing minorities on both sides are increasingly sympathetic to the prospect of leveraging violence to meet moral ends (Ditto et al. 2018; Crawford and Brandt 2020). In fact, only a fraction of U.S. citizens even subscribe to coherent and stable political orientations (Kalmoe 2020), which aligns with our finding that variation in moral beliefs about specific issues—which wash out at the level of monolithic orientation—tracked differentiated neural responses to violent protests. On this basis, one might expect liberals and conservatives to engage similar neurocognitive mechanisms when making moral judgments about sociopolitical violence. The opposite prediction—that liberals and conservatives exploit fundamentally different mechanisms to fuel their intolerance—is also plausible. This is because some differences in the moral psychologies of liberals and conservatives have been documented (Waytz et al. 2019)—that said, counter to an extensive literature that reported effects of political orientation on threat sensitivity, the recent failure of a pre-registered replication study to reproduce these earlier findings suggests differences in political views may not reflect fundamental differences in physiology (Bakker et al. 2020). Although these are interesting possibilities, they could not be tested here—it was not possible to recruit a sufficiently large sample of conservatives to justify group comparisons by political orientation. Additional research is therefore needed to establish the generalizability of these findings to people with conservative social and/or economic political leanings.

Conclusion

Sociopolitical beliefs, when held with moral conviction, can license violence in service of those beliefs despite ordinarily being morally prohibited. This study characterized the neurocognitive mechanisms underpinning moral convictions about sociopolitical issues and beliefs about the appropriateness of ideological violence. Believing violence to be appropriate was associated with parametric increases in vmPFC, while stronger moral conviction was associated with parametric increases in VS and, overall, correlated negatively with activation in amygdala. Together, these findings indicate that moral convictions about sociopolitical issues may raise their subjective value, overriding our natural aversion to interpersonal harm. Value-based decision-making in social contexts provides a powerful platform for understanding sociopolitical and moral decision-making. Disentangling the neurocognitive mechanisms fundamental to value and choice can reveal—at a granular level—physiological boundaries that ultimately constrain our moral psychologies. The theoretical advances afforded by such an undertaking enable more accurate predictions about moral conviction and its impact on social decision-making.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

The study was supported by the National Institute of Mental Health R01 MH109329 and seed grants from the University of Chicago MRI Research Center and University of Chicago Grossman Institute for Neuroscience, Quantitative Biology and Human Behavior.

References

- Ahler Douglas J. 2014. "Self-Fulfilling Misperceptions of Public Polarization." *The Journal of Politics* 76 (3): 607–20. 10.1017/S0022381614000085.
- Alheid George F. 2006. "Extended Amygdala and Basal Forebrain." *Annals of the New York Academy of Sciences* 985 (1): 185–205. 10.1111/j.1749-6632.2003.tb07082.x.
- Apps MAJ, and Ramnani N. 2017. "Contributions of the Medial Prefrontal Cortex to Social Influence in Economic Decision-Making." *Cerebral Cortex* 27 (9): 4635–48. 10.1093/cercor/bhx183. [PubMed: 28922858]
- Apps Matthew A.J., and Sallet Jérôme. 2017. "Social Learning in the Medial Prefrontal Cortex." *Trends in Cognitive Sciences* 21 (3): 151–52. 10.1016/j.tics.2017.01.008. [PubMed: 28173997]
- Bakker Bert N., Schumacher Gijs, Gothreau Claire, and Arceneaux Kevin. 2020. "Conservatives and Liberals Have Similar Physiological Responses to Threats." *Nature Human Behaviour*, 2. 10.1038/s41562-020-0823-z.
- Bartra Oscar, McGuire Joseph T., and Kable Joseph W.. 2013. "The Valuation System: A Coordinate-Based Meta-Analysis of BOLD fMRI Experiments Examining Neural Correlates of Subjective Value." *NeuroImage* 76 (August): 412–27. 10.1016/j.neuroimage.2013.02.063. [PubMed: 23507394]
- Bauman Christopher W., Peter McGraw A, Bartels Daniel M., and Warren Caleb. 2014. "Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology." *Social and Personality Psychology Compass* 8 (9): 536–54. 10.1111/spc3.12131.
- Berendse Henk W, Galis-de Graaf Yvonne, and Groenewegen Henk J. 1992. "Topographical Organization and Relationship with Ventral Striatal Compartments of Prefrontal Corticostriatal Projections in the Rat." *The Journal of Comparative Neurology* 316 (3): 314–47. 10.1002/cne.903160305. [PubMed: 1577988]
- Bostyn Dries H., Sevenhant Sybren, and Roets Arne. 2018. "Of Mice, Men, and Trolleys: Hypothetical Judgment Versus Real-Life Behavior in Trolley-Style Moral Dilemmas." *Psychological Science* 29 (7): 1084–93. 10.1177/0956797617752640. [PubMed: 29741993]
- Bowles Samuel. 2012. "Warriors, Levelers, and the Role of Conflict in Human Social Evolution." *Science* 336 (6083): 876–79. 10.1126/science.1217336. [PubMed: 22605768]
- Boyd Robert, and Richerson Peter J.. 2009. "Culture and the Evolution of Human Cooperation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1533): 3281–88. 10.1098/rstb.2009.0134.
- Brett M, J. Anton, Valabregue R, and Poline JB. 2002. "Region of Interest Analysis Using an SPM Toolbox." In 8th International Conference on Functional Mapping of the Human Brain. Sendai, Japan.
- Buckholz Joshua W., Asplund Christopher L., Dux Paul E., Zald David H., Gore John C., Jones Owen D., and Marois René. 2008. "The Neural Correlates of Third-Party Punishment." *Neuron* 60 (5): 930–40. 10.1016/j.neuron.2008.10.016. [PubMed: 19081385]
- Buckholz Joshua W, and Marois René. 2012. "The Roots of Modern Justice: Cognitive and Neural Foundations of Social Norms and Their Enforcement." *Nature Neuroscience* 15 (5): 655–61. 10.1038/nn.3087. [PubMed: 22534578]
- Carter R. McKell, and Huettel Scott A.. 2013. "A Nexus Model of the Temporal–Parietal Junction." *Trends in Cognitive Sciences* 17 (7): 328–36. 10.1016/j.tics.2013.05.007. [PubMed: 23790322]

- Casebeer William D. 2003. "Moral Cognition and Its Neural Constituents." *Nature Reviews Neuroscience* 4 (10): 840–46. 10.1038/nrn1223. [PubMed: 14523383]
- Chung Hui-Kuan, Tymula Agnieszka, and Glimcher Paul. 2017. "The Reduction of Ventrolateral Prefrontal Cortex Gray Matter Volume Correlates with Loss of Economic Rationality in Aging." *The Journal of Neuroscience* 37 (49): 12068–77. 10.1523/JNEUROSCI.1171-17.2017.. [PubMed: 28982708]
- Cikara Mina, and Van Bavel Jay J.. 2014. "The Neuroscience of Intergroup Relations." *Perspectives on Psychological Science* 9 (3): 245–74. 10.1177/1745691614527464. [PubMed: 26173262]
- Crawford Jarret T., and Brandt Mark J.. 2020. "Ideological (A)Symmetries in Prejudice and Intergroup Bias." *Current Opinion in Behavioral Sciences* 34 (759320): 40–45. 10.1016/j.cobeha.2019.11.007.
- Cristofori Irene, Viola Vanda, Chau Aileen, Zhong Wanting, Krueger Frank, Zamboni Giovanna, and Grafman Jordan. 2015. "The Neural Bases for Devaluing Radical Political Statements Revealed by Penetrating Traumatic Brain Injury." *Social Cognitive and Affective Neuroscience* 10 (8): 1038–44. 10.1093/scan/nsu155. [PubMed: 25656509]
- Curry Oliver Scott, Mullins Daniel Austin, and Whitehouse Harvey. 2019. "Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies." *Current Anthropology* 60 (1): 47–69. 10.1086/701478.
- Decety Jean, and Cowell Jason M.. 2018. "Interpersonal Harm Aversion as a Necessary Foundation for Morality: A Developmental Neuroscience Perspective." *Development and Psychopathology* 30 (01): 153–64. 10.1017/S0954579417000530. [PubMed: 28420449]
- Decety Jean, and Lamm Claus. 2007. "The Role of the Right Temporoparietal Junction in Social Interaction: How Low-Level Computational Processes Contribute to Meta-Cognition." *The Neuroscientist* 13 (6): 580–93. 10.1177/1073858407304654. [PubMed: 17911216]
- Decety Jean, Michalska Kalina J. J., and Kinzler Katherine D. D.. 2012. "The Contribution of Emotion and Cognition to Moral Sensitivity: A Neurodevelopmental Study." *Cerebral Cortex* 22 (1): 209–20. 10.1093/cercor/bhr111. [PubMed: 21616985]
- Decety Jean, Pape Robert, and Workman Clifford I.. 2018. "A Multilevel Social Neuroscience Perspective on Radicalization and Terrorism." *Social Neuroscience* 13 (5): 511–29. 10.1080/17470919.2017.1400462. [PubMed: 29091541]
- Decety Jean, and Porges Eric C.. 2011. "Imagining Being the Agent of Actions That Carry Different Moral Consequences: An FMRI Study." *Neuropsychologia* 49 (11): 2994–3001. 10.1016/j.neuropsychologia.2011.06.024. [PubMed: 21762712]
- Decety Jean, Skelly Laurie R., and Kiehl Kent A.. 2013. "Brain Response to Empathy-Eliciting Scenarios Involving Pain in Incarcerated Individuals With Psychopathy." *JAMA Psychiatry* 70 (6): 638. 10.1001/jamapsychiatry.2013.27. [PubMed: 23615636]
- Decety Jean, and Yoder Keith J. 2016. "Empathy and Motivation for Justice: Cognitive Empathy and Concern, but Not Emotional Empathy, Predict Sensitivity to Injustice for Others." *Social Neuroscience* 11 (1): 1–14. 10.1080/17470919.2015.1029593. [PubMed: 25768232]
- . 2017. "The Emerging Social Neuroscience of Justice Motivation." *Trends in Cognitive Sciences* 21 (1): 6–14. 10.1016/j.tics.2016.10.008. [PubMed: 27865787]
- Delgado Mauricio R, Beer Jennifer S, Fellows Lesley K, Huettel Scott A, Platt Michael L, Quirk Gregory J, and Schiller Daniela. 2016. "Viewpoints: Dialogues on the Functional Role of the Ventromedial Prefrontal Cortex." *Nature Neuroscience* 19 (12): 1545–52. 10.1038/nn.4438. [PubMed: 27898086]
- Delgado Mauricio R, Li Jian, Schiller Daniela, and Phelps Elizabeth A. 2008. "The Role of the Striatum in Aversive Learning and Aversive Prediction Errors." *Philosophical Transactions of the Royal Society B: Biological Sciences* 363 (1511): 3787–3800. 10.1098/rstb.2008.0161.
- Ditto Peter H., Liu Brittany S., Clark Cory J., Wojcik Sean P., Chen Eric E., Rebecca Hofstein Grady Jared B. Celniker, and Zinger Joanne F. 2018. "At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives." *Perspectives on Psychological Science*, 5, 174569161774679. 10.1177/1745691617746796.
- Domínguez Juan F., Félice van Nunspeet Ayushi Gupta, Eres Robert, Louis Winnifred R., Decety Jean, and Molenberghs Pascal. 2018. "Lateral Orbitofrontal Cortex Activity Is Modulated by Group

- Membership in Situations of Justified and Unjustified Violence.” *Social Neuroscience* 13 (6): 739–55. 10.1080/17470919.2017.1392342. [PubMed: 29022771]
- Feinberg Matthew, Kovacheff Chloe, Teper Rimma, and Inbar Yoel. 2019. “Understanding the Process of Moralization: How Eating Meat Becomes a Moral Issue.” *Journal of Personality and Social Psychology* 117 (1): 50–72. 10.1037/pspa0000149. [PubMed: 30869989]
- FeldmanHall Oriël, Mobbs Dean, Evans Davy, Hiscox Lucy, Navrady Lauren, and Dalgleish Tim. 2012. “What We Say and What We Do: The Relationship between Real and Hypothetical Moral Choices.” *Cognition* 123 (3): 434–41. 10.1016/j.cognition.2012.02.001. [PubMed: 22405924]
- Fessler Daniel M T, Pisor Anne C, and Holbrook Colin. 2017. “Political Orientation Predicts Credulity Regarding Putative Hazards.” *Psychological Science* 28 (5): 651–60. 10.1177/0956797617692108. [PubMed: 28362568]
- Fido Dean, and Harper Craig A. 2018. “The Potential Role of Empathy in Reducing Political Polarization: The Case of Brexit.” *OSF Preprints*, 1–25. <https://osf.io/sw637/>.
- Fiske Alan Page, and Rai Tage Shakti. 2014. *Virtuous Violence. Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships*. Cambridge: Cambridge University Press. 10.1017/CBO9781316104668.
- Forman Steven D., Cohen Jonathan D., Fitzgerald Mark, Eddy William F., Mintun Mark A., and Noll Douglas C.. 1995. “Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (fMRI): Use of a Cluster-Size Threshold.” *Magnetic Resonance in Medicine* 33 (5): 636–47. 10.1002/mrm.1910330508. [PubMed: 7596267]
- Geuter Stephan, Qi Guanghao, Welsh Robert C, Wager Tor D, and Lindquist Martin A. 2018. “Effect Size and Power in fMRI Group Analysis.” *BioRxiv Preprints*, 1–23. 10.1101/295048.
- Ginges Jeremy. 2019. “The Moral Logic of Political Violence.” *Trends in Cognitive Sciences* 23 (1): 1–3. 10.1016/j.tics.2018.11.001. [PubMed: 30497936]
- Ginges Jeremy, and Atran Scott. 2009. “What Motivates Participation in Violent Political Action.” *Annals of the New York Academy of Sciences* 1167 (1): 115–23. 10.1111/j.1749-6632.2009.04543.x. [PubMed: 19580558]
- Gintis Herbert. 2011. “Gene-Culture Coevolution and the Nature of Human Sociality.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1566): 878–88. 10.1098/rstb.2010.0310.
- Gollwitzer Mario, and Rothmund Tobias. 2011. “What Exactly Are Victim-Sensitive Persons Sensitive To?” *Journal of Research in Personality* 45 (5): 448–55. 10.1016/j.jrp.2011.05.003.
- Gourley Shannon L, Zimmermann Kelsey S, Allen Amanda G, and Taylor Jane R. 2016. “The Medial Orbitofrontal Cortex Regulates Sensitivity to Outcome Value.” *Journal of Neuroscience* 36 (16): 4600–4613. 10.1523/JNEUROSCI.4253-15.2016.. [PubMed: 27098701]
- Haidt Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Reprint. New York, NY: Vintage Books.
- Halevy Nir, Kreps Tamar A., Weisel Ori, and Goldenberg Amit. 2015. “Morality in Intergroup Conflict.” *Current Opinion in Psychology* 6 (December): 10–14. 10.1016/j.copsyc.2015.03.006.
- Haxby James V., Connolly Andrew C., and Swaroop Guntupalli J. 2014. “Decoding Neural Representational Spaces Using Multivariate Pattern Analysis.” *Annual Review of Neuroscience* 37 (1): 435–56. 10.1146/annurev-neuro-062012-170325.
- Henson R. 2006. “Forward Inference Using Functional Neuroimaging: Dissociations versus Associations.” *Trends in Cognitive Sciences* 10 (2): 64–69. 10.1016/j.tics.2005.12.005. [PubMed: 16406759]
- Hiser Jaryd, and Koenigs Michael. 2018. “The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology.” *Biological Psychiatry* 83 (8): 638–47. 10.1016/j.biopsych.2017.10.030. [PubMed: 29275839]
- Janak Patricia H, and Tye Kay M. 2015. “From Circuits to Behaviour in the Amygdala.” *Nature* 517 (7534): 284–92. 10.1038/nature14188. [PubMed: 25592533]
- Kable Joseph W., and Glimcher Paul W.. 2007. “The Neural Correlates of Subjective Value during Intertemporal Choice.” *Nature Neuroscience* 10 (12): 1625–33. 10.1038/nn2007. [PubMed: 17982449]

- Kalmoe Nathan P. 2020. "Uses and Abuses of Ideology in Political Psychology." *Political Psychology* 41 (4): 771–93. 10.1111/pops.12650.
- Kelly Meagan, Ngo Lawrence, Chituc Vladimir, Huettel Scott, and Sinnott-Armstrong Walter. 2017. "Moral Conformity in Online Interactions: Rational Justifications Increase Influence of Peer Opinions on Moral Judgments." *Social Influence* 12 (2–3): 57–68. 10.1080/15534510.2017.1323007.
- Krajbich Ian, Adolphs Ralph, Tranel Daniel, Denburg Natalie L, and Camerer Colin F. 2009. "Economic Games Quantify Diminished Sense of Guilt in Patients with Damage to the Prefrontal Cortex." *Journal of Neuroscience* 29 (7): 2188–92. 10.1523/JNEUROSCI.5086-08.2009. [PubMed: 19228971]
- Krastev Sekoul, McGuire Joseph T., McNeney Denver, Kable Joseph W., Stolle Dietlind, Gidengil Elisabeth, and Fellows Lesley K.. 2016. "Do Political and Economic Choices Rely on Common Neural Substrates? A Systematic Review of the Emerging Neuropolitics Literature." *Frontiers in Psychology* 7 (February): 264. 10.3389/fpsyg.2016.00264. [PubMed: 26941703]
- Krueger Frank, Kevin McCabe Jorge Moll, Kriegeskorte Nikolaus, Zahn Roland, Strenziok Maren, Heinecke Armin, and Grafman Jordan. 2007. "Neural Correlates of Trust." *Proceedings of the National Academy of Sciences* 104 (50): 20084–89. 10.1073/pnas.0710103104.
- Kurzban Robert, Dukes Amber, and Weeden Jason. 2010. "Sex, Drugs and Moral Goals: Reproductive Strategies and Views about Recreational Drugs." *Proceedings of the Royal Society B: Biological Sciences* 277 (1699): 3501–8. 10.1098/rspb.2010.0608.
- Lebreton Maël, Jorge Soledad, Michel Vincent, Thirion Bertrand, and Pessiglione Mathias. 2009. "An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging." *Neuron* 64 (3): 431–39. 10.1016/j.neuron.2009.09.040. [PubMed: 19914190]
- Levendusky Matthew S., and Malhotra Neil. 2016. "(Mis)Perceptions of Partisan Polarization in the American Public." *Public Opinion Quarterly* 80 (S1): 378–91. 10.1093/poq/nfv045.
- Levy Dino, and Glimcher Paul. 2015. "Common Value Representation—a Neuroeconomic Perspective." In *Handbook of Value*, edited by Brosch T and Sander D, 85–118. Oxford, UK: Oxford University Press. 10.1093/acprof:oso/9780198716600.003.0005.
- Mazaika PK, Hoefl F, Glover GH, and Reiss AL. 2009. "Methods and Software for FMRI Analysis of Clinical Subjects." *NeuroImage* 47 (July): S58. 10.1016/S1053-8119(09)70238-1.
- Miller Ryan, and Cushman Fiery A.. 2018. "Moral Values and Motivations: How Special Are They?" In *The Atlas of Moral Psychology: Mapping Good and Evil*, edited by Gray K and Graham J, 1–13. New York, NY, NY: Guilford Press. <http://cushmanlab.fas.harvard.edu/docs/MoralValues.pdf>.
- Molenberghs Pascal, Gapp Joshua, Wang Bei, Louis Winnifred R., and Decety Jean. 2016. "Increased Moral Sensitivity for Outgroup Perpetrators Harming Ingroup Members." *Cerebral Cortex* 26 (1): 225–33. 10.1093/cercor/bhu195. [PubMed: 25183886]
- Molenberghs Pascal, Ogilvie Claudette, Louis Winnifred R., Decety Jean, Bagnall Jessica, and Bain Paul G.. 2015. "The Neural Correlates of Justified and Unjustified Killing: An FMRI Study." *Social Cognitive and Affective Neuroscience* 10 (10): 1397–1404. 10.1093/scan/nsv027. [PubMed: 25752904]
- Moll Jorge, Bado Patricia, Ricardo de Oliveira-Souza Ivanei E. Bramati, Lima Debora O., Paiva Fernando F., Sato João R., Tovar-Moll Fernanda, and Zahn Roland. 2012. "A Neural Signature of Affiliative Emotion in the Human Septohypothalamic Area." *Journal of Neuroscience* 32 (36): 12499–505. 10.1523/JNEUROSCI.6508-11.2012.. [PubMed: 22956840]
- Moll Jorge, and de Oliveira-Souza Ricardo. 2007. "Response to Greene: Moral Sentiments and Reason: Friends or Foes?" *Trends in Cognitive Sciences* 11 (8): 323–24. 10.1016/j.tics.2007.06.011.
- Moll Jorge, Zahn Roland, Ricardo de Oliveira-Souza Frank Krueger, and Grafman Jordan. 2005. "The Neural Basis of Human Moral Cognition." *Nature Reviews Neuroscience* 6 (10): 799–809. 10.1038/nrn1768. [PubMed: 16276356]
- Ongür D, Price JL, and Ngur D. 2000. "The Organization of Networks within the Orbital and Medial Prefrontal Cortex of Rats, Monkeys and Humans." *Cerebral Cortex* 10 (3): 206–19. 10.1093/cercor/10.3.206. [PubMed: 10731217]

- Pretus Clara, Hamid Nafees, Sheikh Hammad, Ángel Gómez Jeremy Ginges, Adolf Tobeña Richard Davis, Vilarroya Oscar, and Atran Scott. 2019. "Ventromedial and Dorsolateral Prefrontal Interactions Underlie Will to Fight and Die for a Cause." *Social Cognitive and Affective Neuroscience*, 5, 1–23. 10.1093/scan/nsz034. [PubMed: 3122289]
- Pryor Campbell, Perfors Amy, and Howe Piers D. L.. 2019. "Even Arbitrary Norms Influence Moral Decision-Making." *Nature Human Behaviour* 3 (1): 57–62. 10.1038/s41562-018-0489-y.
- Rai Tage Shakti, and Fiske Alan Page. 2011. "Moral Psychology Is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality." *Psychological Review* 118 (1): 57–75. 10.1037/a0021867. [PubMed: 21244187]
- Rangel Antonio, and Clithero John A.. 2018. "The Computation of Stimulus Values in Simple Choice." In *Neuroeconomics: Decision Making and the Brain: Second Edition*, edited by Glimcher PW and Fehr E, 2nd ed., 125–48. London, UK: Academic Press. 10.1016/B978-0-12-416008-8.00008-5.
- Ritchie J. Brendan, Kaplan David Michael, and Klein Colin. 2017. "Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience." *The British Journal for the Philosophy of Science*, no. November (September): 1–33. 10.1093/bjps/axx023.
- Rouault Marion, Drugowitsch Jan, and Koechlin Etienne. 2019. "Prefrontal Mechanisms Combining Rewards and Beliefs in Human Decision-Making." *Nature Communications* 10 (1): 301. 10.1038/s41467-018-08121-w.
- Ruff Christian C., and Fehr Ernst. 2014. "The Neurobiology of Rewards and Values in Social Decision Making." *Nature Reviews Neuroscience* 15 (8): 549–62. 10.1038/nrn3776. [PubMed: 24986556]
- Ruff Christian C., Ugazio Giuseppe, and Fehr Ernst. 2013. "Changing Social Norm Compliance with Noninvasive Brain Stimulation." *Science* 342 (6157): 482–84. 10.1126/science.1241399. [PubMed: 24091703]
- Russell Gordon W. 2004. "Sport Riots: A Social–Psychological Review." *Aggression and Violent Behavior* 9 (4): 353–78. 10.1016/S1359-1789(03)00031-4.
- Schmitt Manfred, Baumert Anna, Gollwitzer Mario, and Maes Jürgen. 2010. "The Justice Sensitivity Inventory: Factorial Validity, Location in the Personality Facet Space, Demographic Pattern, and Normative Data." *Social Justice Research* 23 (2–3): 211–38. 10.1007/s11211-010-0115-2.
- Seara-Cardoso Ana, Sebastian Catherine L., Eamon McCrory Lucy Foulkes, Buon Marine, Roiser Jonathan P., and Viding Essi. 2016. "Anticipation of Guilt for Everyday Moral Transgressions: The Role of the Anterior Insula and the Influence of Interpersonal Psychopathic Traits." *Scientific Reports* 6 (1): 36273. 10.1038/srep36273. [PubMed: 27808160]
- Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, and Greicius MD. 2007. "Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control." *Journal of Neuroscience* 27 (9): 2349–56. 10.1523/JNEUROSCI.5587-06.2007. [PubMed: 17329432]
- Simas Elizabeth N., Clifford Scott, and Kirkland Justin H.. 2019. "How Empathic Concern Fuels Political Polarization." *American Political Science Review*, 10, 1–12. 10.1017/S0003055419000534.
- Skitka Linda J. 2010. "The Psychology of Moral Conviction." *Social and Personality Psychology Compass* 4 (4): 267–81. 10.1111/j.1751-9004.2010.00254.x.
- Skitka Linda J., Bauman Christopher W., and Sargis Edward G.. 2005. "Moral Conviction: Another Contributor to Attitude Strength or Something More?" *Journal of Personality and Social Psychology* 88 (6): 895–917. 10.1037/0022-3514.88.6.895. [PubMed: 15982112]
- Skitka Linda J., and Conway Paul. 2019. "Morality." In *Advances Social Psychology*, edited by Finkel EJ and Baumeister RF, 299–323. New York, NY: Oxford University Press.
- Skitka Linda J., and Scott Morgan G. 2014. "The Social and Political Implications of Moral Conviction." *Political Psychology* 35 (SUPPL.1): 95–110. 10.1111/pops.12166.
- Smith Kristopher M., and Kurzban Robert. 2019. "Morality Is Not Always Good." *Current Anthropology* 60 (1): 61–62. 10.1086/701478.

- Stallen Mirre, Smidts Ale, and Sanfey Alan G. 2013. "Peer Influence: Neural Mechanisms Underlying in-Group Conformity." *Frontiers in Human Neuroscience* 7 (March): 1–7. 10.3389/fnhum.2013.00050. [PubMed: 23355817]
- Thomas E.F. Emma F., Bury Simon M., Louis Winnifred R. W.R., Amiot Catherine E. C.E., Molenberghs Pascal, Crane Monique F. M.F., and Decety Jean. 2019. "Vegetarian, Vegan, Activist, Radical: Using Latent Profile Analysis to Examine Different Forms of Support for Animal Welfare." *Group Processes & Intergroup Relations* 22 (6): 836–57. 10.1177/1368430218824407.
- Tooby John, and Cosmides Leda. 2010. "Groups in Mind: The Coalitional Roots of War and Morality." In *Human Morality and Sociality: Evolutionary and Comparative Perspectives*, edited by Høgh-Olesen H, 191–234. New York, NY: Palgrave Macmillan UK.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, and Joliot M. 2002. "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain." *NeuroImage* 15 (1): 273–89. 10.1006/nimg.2001.0978. [PubMed: 11771995]
- Villasenor John. 2017. "Views among College Students Regarding the First Amendment: Results from a New Survey." Washington, DC. <https://www.brookings.edu/blog/fixgov/2017/09/18/views-among-college-students-regarding-the-first-amendment-results-from-a-new-survey/>.
- Volz Kirsten G, Kessler Thomas, and Yves von Cramon D. 2009. "In-Group as Part of the Self: In-Group Favoritism Is Mediated by Medial Prefrontal Cortex Activation." *Social Neuroscience* 4 (3): 244–60. 10.1080/17470910802553565. [PubMed: 19085561]
- Wager Tor D., Davidson Matthew L., Hughes Brent L., Lindquist Martin A., and Ochsner Kevin N.. 2008. "Prefrontal-Subcortical Pathways Mediating Successful Emotion Regulation." *Neuron* 59 (6): 1037–50. 10.1016/j.neuron.2008.09.006. [PubMed: 18817740]
- Waytz Adam, Iyer Ravi, Young Liane, Haidt Jonathan, and Graham Jesse. 2019. "Ideological Differences in the Expanse of the Moral Circle." *Nature Communications* 10 (1): 4389. 10.1038/s41467-019-12227-0.
- Wiech Katja, Kahane Guy, Shackel Nicholas, Farias Miguel, Savulescu Julian, and Tracey Irene. 2013. "Cold or Calculating? Reduced Activity in the Subgenual Cingulate Cortex Reflects Decreased Emotional Aversion to Harming in Counterintuitive Utilitarian Judgment." *Cognition* 126 (3): 364–72. 10.1016/j.cognition.2012.11.002. [PubMed: 23280149]
- Wilson Michael L., Boesch Christophe, Fruth Barbara, Furuichi Takeshi, Gilby Ian C., Hashimoto Chie, Hobaiter Catherine L., et al. 2014. "Lethal Aggression in Pan Is Better Explained by Adaptive Strategies than Human Impacts." *Nature* 513 (7518): 414–17. 10.1038/nature13727. [PubMed: 25230664]
- Wittmann Marco K., Lockwood Patricia L., and Rushworth Matthew F.S.. 2018. "Neural Mechanisms of Social Cognition in Primates." *Annual Review of Neuroscience* 41 (1): 99–118. 10.1146/annurev-neuro-080317-061450.
- Workman Clifford I, Lythe Karen E, McKie Shane, Moll Jorge, Gethin Jennifer A, Deakin John FW, Elliott Rebecca, and Zahn Roland. 2016. "Subgenual Cingulate–Amygdala Functional Disconnection and Vulnerability to Melancholic Depression." *Neuropsychopharmacology* 41 (8): 2082–90. 10.1038/npp.2016.8. [PubMed: 26781519]
- Yarkoni Tal. 2009. "Big Correlations in Little Studies: Inflated FMRI Correlations Reflect Low Statistical Power—Commentary on Vul et Al. (2009)." *Perspectives on Psychological Science* 4 (3): 294–98. 10.1111/j.1745-6924.2009.01127.x. [PubMed: 26158966]
- Yoder Keith J, and Decety Jean. 2014. "The Good, the Bad, and the Just: Justice Sensitivity Predicts Neural Response during Moral Evaluation of Actions Performed by Others." *Journal of Neuroscience* 34 (12): 4161–66. 10.1523/JNEUROSCI.4648-13.2014. [PubMed: 24647937]
- Zahn Roland, Moll Jorge, Paiva Mirella, Garrido Griselda, Krueger Frank, Huey Edward D., and Grafman Jordan. 2009. "The Neural Basis of Human Social Values: Evidence from Functional MRI." *Cerebral Cortex* 19 (2): 276–83. 10.1093/cercor/bhn080. [PubMed: 18502730]
- Zamboni Giovanna, Gozzi Marta, Krueger Frank, Duhamel Jean-René, Sirigu Angela, and Grafman Jordan. 2009. "Individualism, Conservatism, and Radicalism as Criteria for Processing Political Beliefs: A Parametric FMRI Study." *Social Neuroscience* 4 (5): 367–83. 10.1080/17470910902860308. [PubMed: 19562629]

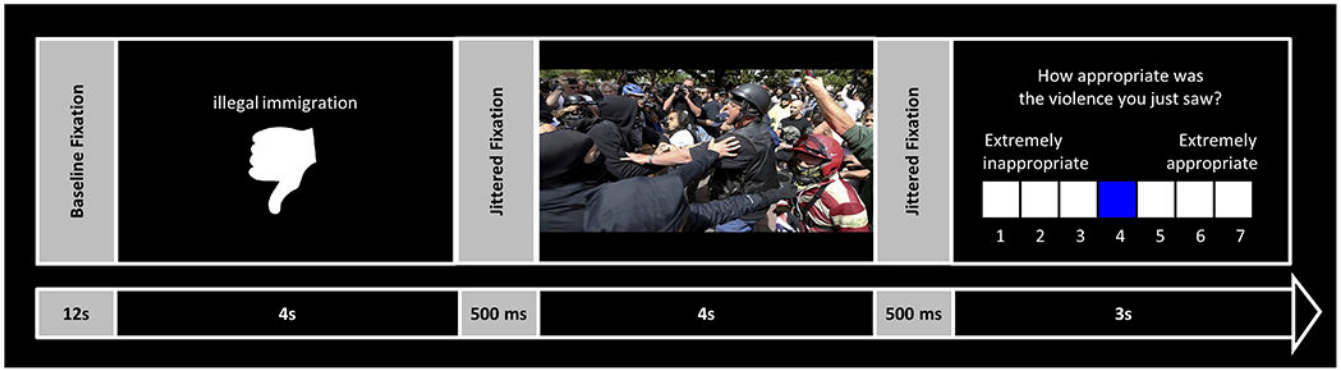


Figure 1 |. Schematic of an individual trial from the novel functional MRI task that assessed beliefs about violence that was either congruent or incongruent with participants' views. After baseline fixation, participants saw a text description of a sociopolitical issue purportedly motivating a violent protest (e.g., illegal immigration). This issue was paired with either a “thumbs up” or “thumbs down” to indicate the protestors’ support or opposition. Next, after jittered fixation, participants saw a photograph ostensibly depicting the violent protest. Following an additional period of jittered fixation, participants then indicated the appropriateness of the violence in the photograph.

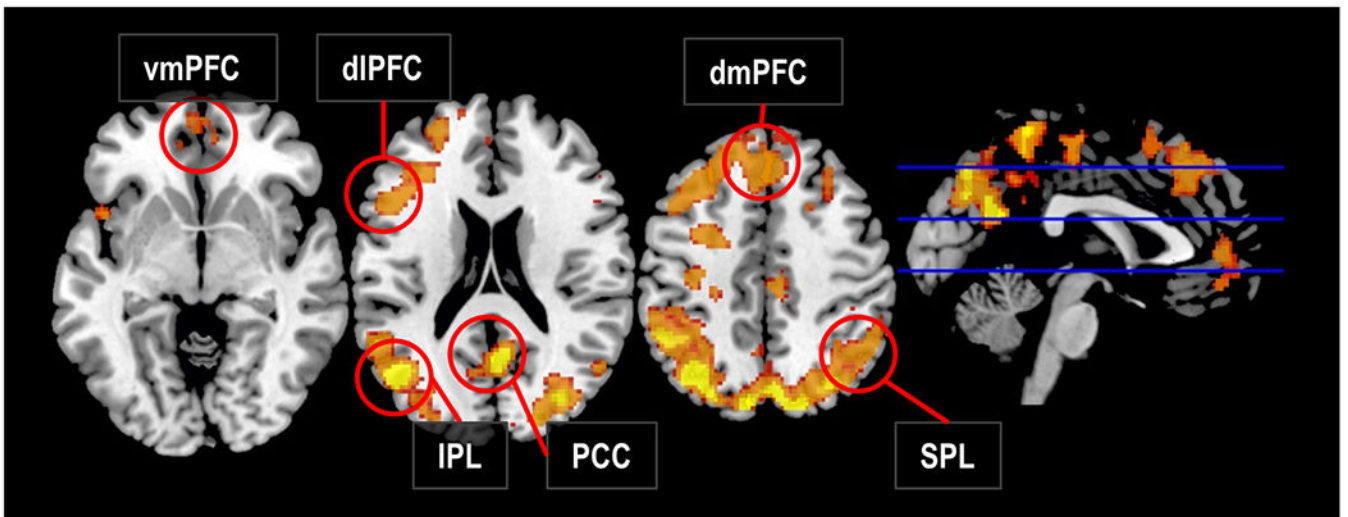


Figure 2 | Neural responses sensitive to the congruency of violent sociopolitical protests. Brain regions in which above-chance classification accuracy was achieved using Support Vector Machine (SVM) searchlights to distinguish between responses to photographs of violent protests that were ostensibly congruent compared to incongruent with participants' sociopolitical views. dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; PCC, posterior cingulate cortex; SPL, superior parietal lobule; vmPFC, ventromedial prefrontal cortex.

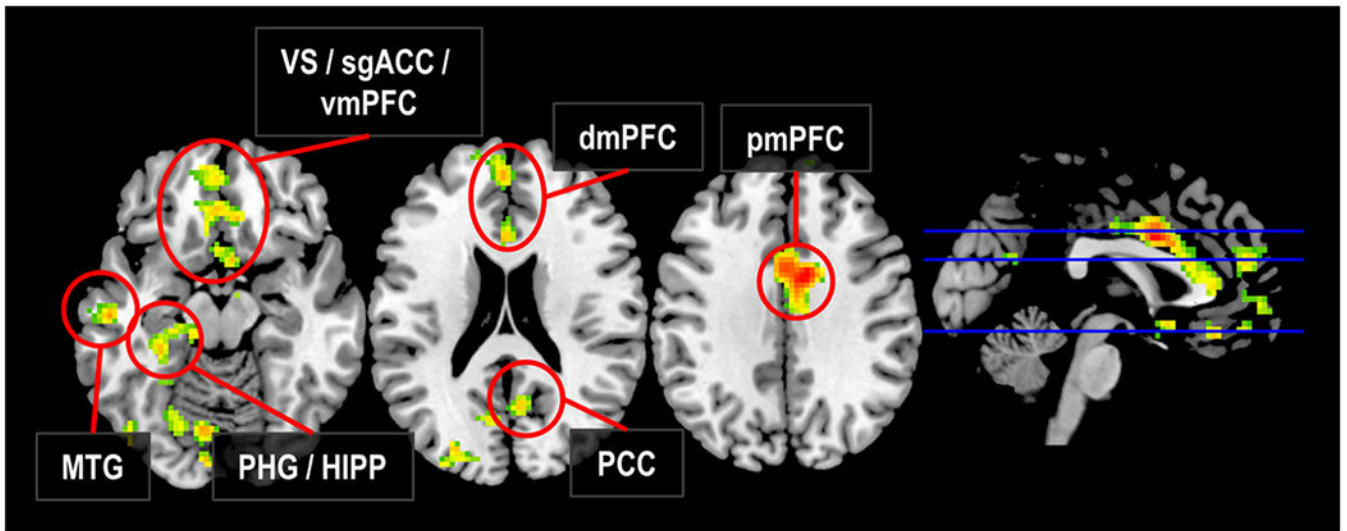


Figure 3 |. Neural responses sensitive to the congruency of violent sociopolitical protests and relations between these responses and beliefs about violence.

Regions in which the strength of neural response to perceiving congruent relative to incongruent acts of sociopolitical violence showed a positive parametric relationship with in-scanner judgments regarding the appropriateness of these acts. dmPFC, dorsomedial prefrontal cortex; HIPP, hippocampus; MTG, middle temporal gyrus; PCC, posterior cingulate cortex; PHG, parahippocampal gyrus; pmPFC, posterior medial prefrontal cortex; sgACC, subgenual cingulate cortex; vmPFC, ventromedial prefrontal cortex; VS, ventral striatum.

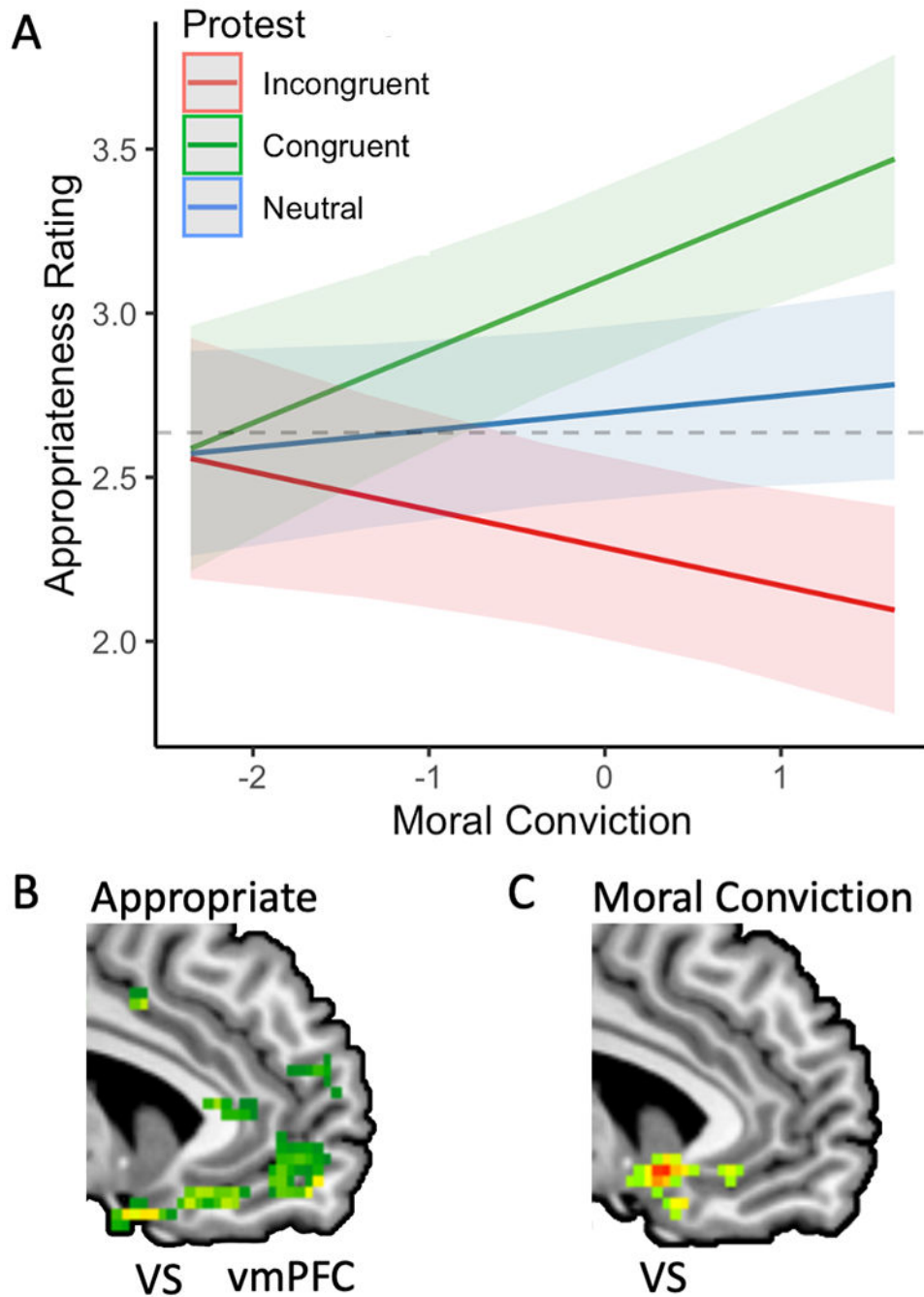


Figure 4 | Moral convictions and judgments about violence ostensibly carried out in service of specific sociopolitical issues converge in implicating nodes of the reward circuit.

A. Estimated marginal means showing opposing effects of moral conviction on judgments about the appropriateness of political violence. Those with the strongest moral convictions were the most tolerant of congruent political violence while also the least tolerant of incongruent violence. **B.** Moral judgments of appropriateness positively modulated neural signal in the ventral striatum (VS) and ventromedial prefrontal cortex (vmPFC). **C.** Moral

convictions about specific issues, assessed prior to scanning, positively modulated hemodynamic response amplitudes in VS.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Brain areas demonstrating significantly greater hemodynamic response to photographs depicting acts of sociopolitical violence that are purportedly congruent compared to incongruent with participants' sociopolitical views. The opposite contrast did not reveal regions in which hemodynamic response was increased for incongruent compared to incongruent sociopolitical violence.

Brain Regions by Cluster	Cluster Size	Peak MNI Coordinates			Peak <i>t</i> statistic
		<i>x</i>	<i>y</i>	<i>z</i>	
Bilateral dmPFC Bilateral pre-SMA	157	-7	18	47	4.496
Right dlPFC	461	43	46	14	4.474
Left dlPFC	113	-42	28	29	4.258
Right aINS Right latOFC	170	48	16	-4	4.110
Bilateral PCC	134	6	-37	23	4.101
Left precuneus	159	-20	-67	50	3.909

aINS, anterior insula; dACC, dorsal anterior cingulate cortex; dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; latOFC, lateral orbitofrontal cortex; MNI, Montreal Neurological Institute; PCC, posterior cingulate cortex; SMA, supplementary motor area.

Table 2.

Brains areas where SVM searchlights obtained above-chance classification accuracy for Congruent vs. Incongruent scenes.

Brain Regions by Cluster	Cluster Size	Peak MNI Coordinates			Peak z statistic
		x	y	z	
Left angular gyrus	14241	-46	-66	24	2.327
PCC/Precuneus		2	-60	22	2.267
Left angular gyrus		-46	-66	24	2.327
Right occipital		22	-84	22	2.071
Right superior parietal		32	-68	50	2.160
Left inferior frontal gyrus	6188	-46	16	6	2.038
Left dlPFC		-44	18	34	2.038
Left vlPFC		-44	14	4	2.036
dmPFC		-2	30	44	2.038
vmPFC		-4	53	0	2.036
Right middle frontal gyrus	253	40	24	28	1.981
Right superior frontal gyrus	25	28	58	10	1.970

dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; MNI, Montreal Neurological Institute; PCC, posterior cingulate cortex; SVM, support vector machine; vlPFC, ventrolateral prefrontal cortex.

Table 3.

Brains areas in which the amplitude of hemodynamic response to congruent compared to incongruent sociopolitical violence varied parametrically with ratings made during scanning of the appropriateness of the violence depicted in each photograph.

Brain Regions by Cluster	Cluster Size	Peak MNI Coordinates			Peak <i>t</i> statistic
		<i>x</i>	<i>y</i>	<i>z</i>	
Bilateral pmPFC Bilateral dACC	510	11	3	35	5.31
Left parahippocampal gyrus Left hippocampus	154	-25	-50	-16	4.65
Bilateral VS Bilateral sgACC Left vmPFC	344	-10	6	-19	4.45
Left medial occipitotemporal gyrus	211	-20	-70	-10	4.38
Left middle occipital gyrus	249	-45	-72	-10	4.16
Left middle temporal gyrus	105	-55	-17	-13	4.02
Bilateral dmPFC	165	-2	51	20	3.91
Bilateral PCC	100	6	-60	20	3.57

dACC, dorsal anterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; MNI, Montreal Neurological Institute; PCC, posterior cingulate cortex; pmPFC, posterior medial prefrontal cortex; sgACC, subgenual cingulate cortex; vmPFC, ventromedial prefrontal cortex; VS, ventral striatum.

Table 4.

Brains areas where the amplitude of hemodynamic response for congruent compared to incongruent sociopolitical violence varied parametrically with moral conviction scores corresponding to each of the issues that purportedly motivated the violent protests.

Brain Regions by Cluster	Cluster Size	Peak MNI Coordinates			Peak <i>t</i> statistic
		<i>x</i>	<i>y</i>	<i>z</i>	
Left middle occipital gyrus Left fusiform gyrus	871	-45	-55	-19	5.54
Right middle occipital gyrus Right fusiform gyrus	1054	26	-77	5	5.49
Left VS Left sgACC	94	-7	16	-7	5.24

MNI, Montreal Neurological Institute; sgACC, subgenual cingulate cortex; VS, ventral striatum.