Cognitive, Affective, & Behavioral Neuroscience (2025) 25:291–310
https://doi.org/10.3758/s13415-024-01243-3

RESEARCH ARTICLE

Moral conviction interacts with metacognitive ability in modulating neural activity during sociopolitical decision-making

Qiongwen Cao[1] · Michael S. Cohen[1] · Akram Bakkour[1] · Yuan Chang Leong[1] · Jean Decety[1,2]

Accepted: 6 November 2024 / Published online: 19 December 2024

Abstract

The extent to which a belief is rooted in one's sense of morality has significant societal implications. While moral conviction can inspire positive collective action, it can also prompt dogmatism, intolerance, and societal divisions. Research in social psychology has documented the functional characteristics of moral conviction and shows that poor metacognition exacerbates its negative outcomes. However, the cognitive and neural mechanisms underlying moral conviction, their relationship with metacognition, and how moral conviction is integrated into the valuation and decision-making process remain unclear. This study investigated these neurocognitive processes during decision-making on sociopolitical issues varying in moral conviction. Participants (N = 44) underwent fMRI scanning while deciding, on each trial, which of two groups of political protesters they supported more. As predicted, stronger moral conviction was associated with faster decision times. Hemodynamic responses in the anterior insula (aINS), anterior cingulate cortex (ACC), and lateral prefrontal cortex (lPFC) were elevated during decisions with higher moral conviction, supporting the emotional and cognitive dimensions of moral conviction. Functional connectivity between lPFC and vmPFC was greater on trials higher in moral conviction, elucidating mechanisms through which moral conviction is incorporated into valuation. Average support for the two displayed groups of protesters was positively associated with brain activity in regions involved in valuation, particularly vmPFC and amygdala. Metacognitive sensitivity, the ability to discriminate one's correct from incorrect judgments, measured in a perceptual task, negatively correlated with parametric effects of moral conviction in the brain, providing new evidence that metacognition modulates responses to morally convicted issues.

Keywords Attitudes · Decision making · Metacognition · Moral conviction · Political attitudes · Valuation

Introduction

Morally convicted attitudes pertain to beliefs regarding what is fundamentally right and wrong. These attitudes reflect core moral values, which are perceived as culturally universal absolutes, stable over time, and particularly resistant to authority (Luttrell & Togans, 2021; Skitka et al., 2021). Moral convictions can inspire benevolent forms of collective action, such as the American civil rights movement, but they can also incite dogmatism, division, and authoritarianism (Decety, 2024; Garrett & Bankert, 2020). Other harmful consequences include aggressive attitudes, hate speech, justification of prejudice, vigilantism, and physical violence against people or groups who share different values or practices (Wright & Pözler, 2022; Workman et al., 2020; Yoder & Decety, 2022), all of which can have a corrosive effect on democracy (Finkel et al., 2020).

Moral conviction incorporates cognitive and affective dimensions (Wright et al., 2008). The cognitive dimension refers to the distinction between beliefs that are or are not moralized and accounts for the fact that the former are treated almost as objective and universal truth (Goodwin & Darley, 2008). The affective dimension, in contrast, reflects the emotional intensity associated with these beliefs. Research in social psychology and political science suggests that when people moralize their opinions or attitudes, they are less likely to take in new information or consider arguments based on cost/benefit analysis (Ryan, 2019). This heightened moral conviction not only affects individuals'

✉ Jean Decety
decety@uchicago.edu

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

[2] Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA

cognition, but also scales up and leads to actions that influence others. For example, moral conviction has been linked to greater chances of sharing politically congruent partisan news, regardless of its veracity (Marie et al., 2023). Another study analyzed more than 500,000 tweets and demonstrated that messages increase their reach by 20% with each additional moral-emotional word (Brady et al., 2017). People are also more willing to engage in normative and nonnormative collective action when confronted with attitudes that strongly, rather than weakly, violate their moral beliefs (Pauls et al., 2022; Thomas et al., 2019). Meanwhile, encountering others holding opposing beliefs elicits negative emotions, particularly if these beliefs are perceived as moral obligations (Ryan, 2014; Zaal et al., 2017). Reading about more morally convicted political information predicts stronger physiological arousal measured by skin conductance, further supporting the contention that moral conviction evokes strong emotions (Garrett, 2019). Overall, these behavioral findings support the idea that morally convicted beliefs, compared with beliefs held without moral conviction, are cognitively perceived as more absolute and universal, and emotionally as higher in salience, motivating people to exert considerable effort to persuade others and achieve their moral objectives.

Because moral conviction both elicits strong emotions and leads to the cognitive characterization of attitudes as largely absolute, judgments and behaviors about highly morally convicted issues may be faster and more consistent across time and scenarios. Indeed, research has found that moral evaluations of actions are faster than pragmatic evaluations, and individuals who perceived a greater moral basis of their attitudes showed a stronger correlation between their attitudes and behavioral intentions (Luttrell et al., 2016; Van Bavel et al., 2012). It remains underexplored; however, how moral conviction affects sociopolitical decisions. The current study introduces a novel task where participants choose between two groups of protesters and decide which group they support more. This design enables us to investigate how moral conviction levels influence decision time and consistency between attitudes and decisions.

Prior research on the cognitive neuroscience of morality suggests a pivotal role for lateral prefrontal cortex (lPFC) in the flexible implementation of social norms and the pursuit of moral goals (Carlson & Crockett, 2018; Yoder & Decety, 2018). Proper functioning of the dorsolateral prefrontal cortex (dlPFC) seems necessary for individuals to act cooperatively, synthesize the intentions of wrongdoers and the perceived harm to the victim, and determine punishments for moral and norm violations (Decety & Cowell, 2018; Krueger & Hoffman, 2016; Soutschek et al., 2015). Meanwhile, behavioral research has found that categorizing a belief as moral, relying on a cognitive judgment, leads to less tolerance toward those holding divergent views, while holding a belief with greater emotional intensity alone does not (Wright et al., 2008). Thus, punitive decisions studied in previous cognitive neuroscience studies can be understood as extensions of this intolerance and as consequences arising from the cognitive recognition of goals as moral. Other cognitive neuroscience studies have documented the implication of the anterior insula (aINS), anterior cingulate cortex (ACC), and amygdala in assessing the outcomes of interpersonal actions, especially harm, in a moral context (FeldmanHall & Mobbs, 2015; Hesse et al., 2016). The salience network, a suite of interconnected cortical and subcortical regions, including the ACC, aINS, amygdala, ventral striatum, periaqueductal gray, and ventral tegmental area, has a crucial role in detecting behaviorally relevant information and coordinating neural resources (Uddin, 2015). Specifically, the amygdala plays an essential role in directing attention to motivationally relevant and emotionally arousing stimuli regardless of valence (Cunningham & Brosch, 2012; Lin et al., 2020; Wang et al., 2017). The aINS, a key region in the salience network, serves a related function in tracking overall emotional appraisals of moral situations (Hutcherson et al., 2015; Shenhav & Greene, 2014). Another major node in the salience network, the ACC, monitors the strength of emotional reactions to specific events (Seamans & Floresco, 2022). Together, it is reasonable to predict that the lPFC, amygdala, aINS, and ACC might be among the brain regions in which activity during decision-making tracks the levels of moral conviction.

Functional neuroimaging studies have consistently identified a set of interconnected regions that underly moral decision-making. This circuit includes the ventromedial prefrontal cortex (vmPFC), the ventral striatum (VS), the lPFC, and aINS (Qu et al., 2022; Yoder & Decety, 2018). Consistent with the common-currency hypothesis, some studies have demonstrated that the valuation system, including the vmPFC and VS, tracks the subjective value of voluntary donations and the appropriateness ratings of sociopolitical violence (Clithero et al., 2011; Hare et al., 2010; Workman et al., 2020). Other studies have shown that distinct brain regions, including the temporoparietal junction (TPJ), lPFC, and insula, track moral values (Crockett et al., 2017; Qu et al., 2020; Ugazio et al., 2022). However, most past research has used heterogeneous tasks whose moral relevance has not been systematically measured. Hence, it is difficult to determine whether activity in the reported regions is specifically due to the moral significance of the issues being tested or because of related features, such as empathy and mentalizing. In the present study, while fMRI data were being recorded, participants made decisions on which of two protester groups they supported more. Prior to scanning, the support and moral conviction levels of each issue had been measured. Thus, it was possible to quantitatively link these measures to neural mechanisms during the decision-making

process and relate the strength of these neural responses to behavioral measures such as decision time and attitude-decision consistency.

Cognitive inflexibility plays a role in moral conviction and is generally seen as an outcome of the moralization process (Skitka et al., 2005). Relatedly, evidence from multiple studies suggests that strong and dogmatic opinions are associated with a cognitive style that includes low metacognitive sensitivity (Rollwage et al., 2018; Yoder & Decety, 2022; Zmigrod et al., 2020). Metacognitive sensitivity reflects the degree to which confidence differentiates between correct and incorrect responses. Notably, the radical and dogmatic opinions examined are sociopolitical in nature, whereas metacognitive sensitivity is measured by an unrelated perceptual task. These findings are consistent with work showing that metacognition is at least partly domain-general (Mazancieux et al., 2020). While the precise mechanistic commonalities between perceptual metacognition and formation of sociopolitical beliefs are not fully understood, the domain-generality of metacognition helps to make sense of this perhaps-surprising relationship that has emerged across several distinct studies. Note that metacognitive sensitivity should not be confused with metacognitive bias, which is the overall level of overconfidence or underconfidence (Fleming & Lau, 2014). Low metacognitive sensitivity refers to a reduced capacity to distinguish between an individual's own correct and incorrect decisions; in such cases, after answering a question, individuals' confidence judgments are not well-aligned with their actual performance. Importantly, in a study in which both metacognitive sensitivity and overconfidence were measured, only metacognitive sensitivity showed a significant negative association with dogmatism, whereas overconfidence did not (Rollwage et al., 2018). Furthermore, in that study, for individuals holding politically radical beliefs, poor metacognitive sensitivity is driven by unreasonably high confidence particularly with *incorrect* decisions, compared to moderates. Another previous study found that individuals with lower metacognitive sensitivity show a more robust tendency for stronger moral conviction to be associated with decreased social conformity (Yoder & Decety, 2022). That study used electrophysiological measures and found that metacognitive sensitivity moderates the effects of moral conviction on midfrontal negativity (MFN), a signal thought to originate in the ACC, when participants evaluate how much they support violent political protesters. Cognitive neuroscience research has found that the spatiotemporal organization of the salience network is predictive of cognitive flexibility (Chen et al., 2016) and the lPFC plays an important role in metacognition (Fleming & Dolan, 2012; Lapate et al., 2020). These regions overlap with some of the regions that we expect to be associated with moral conviction. Thus, it was predicted that individual differences in metacognitive ability may modulate the effects of moral conviction on neural activity during social decision-making.

The current study was designed to determine: 1) the extent to which moral conviction and associated neural activity influence decision time and decision consistency; 2) where and how brain responses are modulated by one's support for and moral conviction about sociopolitical protests during decisions choosing which one of two protests to support; and 3) how individual differences in metacognitive sensitivity influence these neural responses. Based on previous research on moral cognition and value-based decision-making, blood oxygen level-dependent (BOLD) responses in the vmPFC and VS were expected to encode the mean support rating of the two issues presented in each decision, signaling the overall subjective value of the decision. Moreover, the neural activity in the lPFC, amygdala, ACC and aINS was anticipated to be associated with the moral conviction level of a decision. Metacognitive sensitivity, as an individual-level trait, was expected to moderate the impact of moral conviction on hemodynamic responses in these regions.

Methods

Participants

Eighty adult U.S. citizens (45 females, 34 males, 1 nonbinary; age range 18–48 years; $M_{age} = 23.84$, $SD_{age} = 5.98$) from the Chicago metropolitan area were compensated $10 to complete an online survey to collect their demographic information and assess their views on current sociopolitical issues. Only those who had 15 or more unique combinations of support and moral conviction scores on the sociopolitical issues were invited to participate in the fMRI study. Forty-nine healthy adult participants completed the fMRI study and were paid an additional $40 compensation. All participants provided informed written consent, and all procedures were approved by the Institutional Review Board at the University of Chicago. No part of the study procedures or analyses were pre-registered prior to the research being conducted.

Three participants were excluded from both behavioral and fMRI analyses involving the in-scanner task owing to a below-chance (less than 50%) level of consistency. A choice was considered consistent when a participant chose the protesters who supported the issue to which they had given a higher support score in the initial survey and was inconsistent if the opposite was true. Framewise displacement (FD), calculated using MRIQC (Esteban et al., 2017), was used to identify runs and participants with excessive motion. Functional MRI runs for which greater than 5% of volumes had

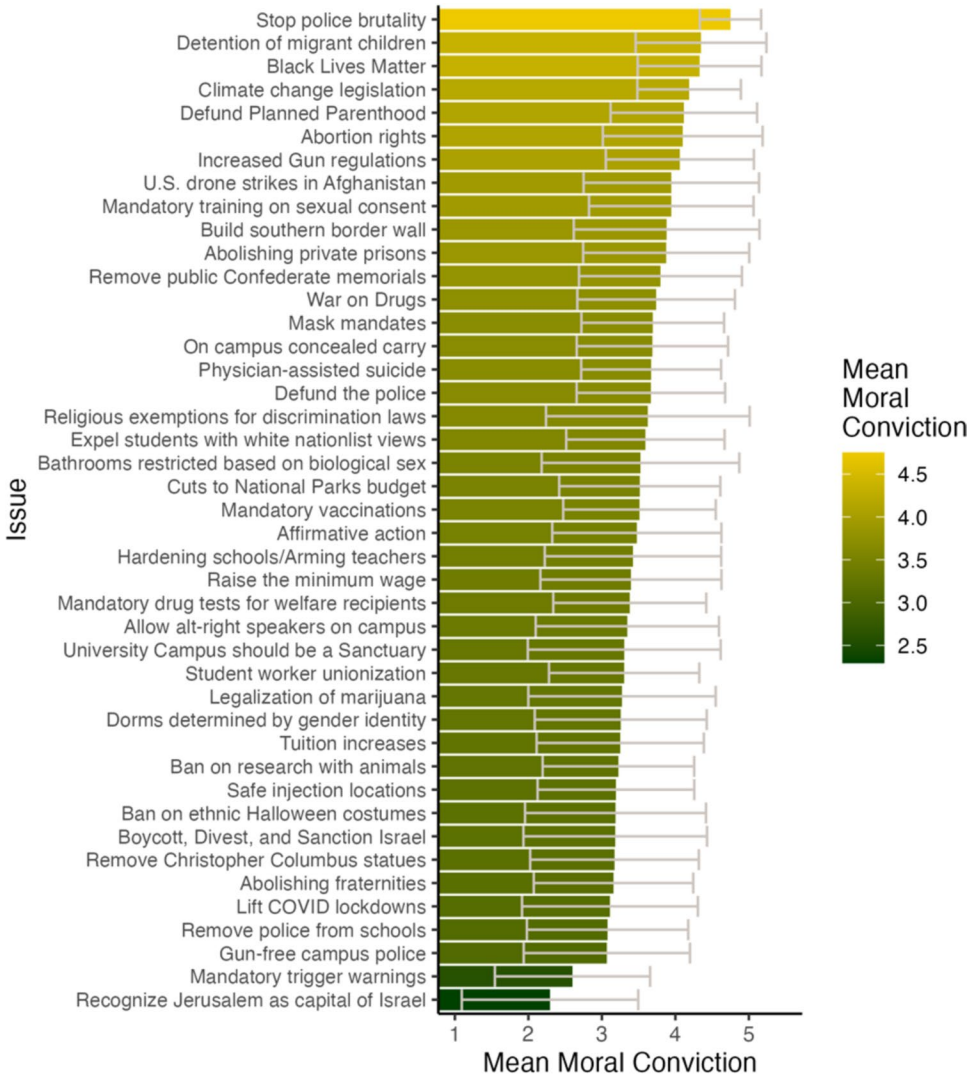

Fig. 1 Mean moral conviction ratings for each sociopolitical issue from the initial online survey. Error bars represent 1 standard deviation from the mean

FD above 0.9 mm were excluded from the MRI analyses, as were runs in which the mean FD was at least two standard deviations worse than the mean from all runs. If three or more runs from the same participant met these criteria, all runs from the participant were excluded. Accordingly, one participant and five individual runs from four participants were excluded from the MRI analyses because of excessive movement. While we did allow participants who were taking a single SSRI antidepressant to participate in the study, one participant was erroneously recruited despite reporting that they were taking Adderall; this participant was removed from behavioral and fMRI analyses involving the in-scanner task. Metacognitive sensitivity data from six participants were lost owing to technical issues. Thus, the final behavioral sample for the decision-making task had 45 participants, with fMRI data included for 44 of them (27 females, 16 males, 1 nonbinary; age range 18–48 years; $M_{age} = 22.27$, $SD_{age} = 5.06$, 38 of whom provided data on the metacognitive sensitivity measure).

Procedures

The initial survey was completed online at least 1 week prior to the fMRI study. Participants answered four questions about a series of sociopolitical issues (Fig. 1). For each issue, participants indicated their degrees of familiarity (from not at all familiar to very familiar) on a 5-point scale and support (from strongly oppose to strongly support) on a 7-point scale. Moral conviction for each issue was indexed by the average score of two questions using a 5-point scale: "To what extent is your position on ___ a reflection of your core moral beliefs and convictions?" and "To what extent

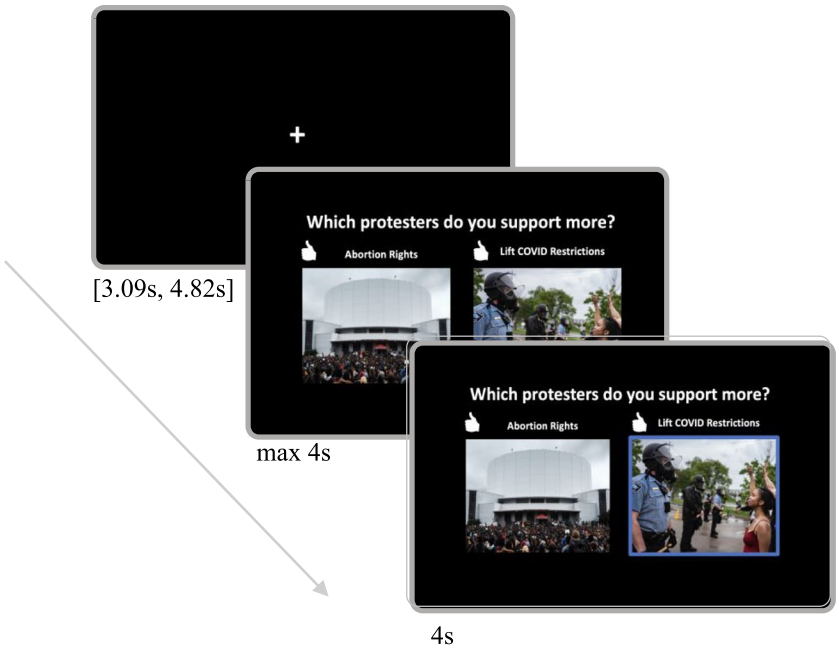

Fig. 2 Example of stimulus sequence in a trial during fMRI scanning

is your position on ___ connected to your beliefs about fundamental right and wrong?" (Skitka & Morgan, 2014). Demographics, including age, gender, household income, level of education, and religiosity, were obtained. Participants provided their political engagement, party registration, party alignment, and political orientation regarding social and economic issues. Religiosity and justice sensitivity were measured by the Duke University Religion Index (DUREL) (Koenig & Büssing, 2010) and the Justice Sensitivity Short Scales (Baumert et al., 2014), respectively.

Prior to functional MRI scanning, participants' metacognitive sensitivity was assessed with a perceptual confidence task designed by Fleming and Lau (2014). The task was programmed by using PsychToolbox and run within MATLAB. In each trial, participants saw two circles that contained different numbers of dots and determined which circle had more dots. After each decision, participants rated their confidence level regarding their choice using a 5-point scale. The task was self-paced. After four example stimuli, difficulty of the task was customized for each participant during an initial calibration phase in which instead of indicating confidence, each participant received feedback (correct or incorrect) on their judgments. The number of trials in the calibration phase was dependent on each participant. The first calibration trial started with one circle containing 50 dots and the other containing at least 1 and at most 100 dots. The algorithm then flexibly increased difficulty (after 2 correct judgments) or decreased difficulty (after 1 incorrect judgment) by manipulating the number of dots presented in the two circles in a trial. The calibration ended when the difficulty level had reversed eight times. Supposedly at this point, the accuracy level reached approximately 71% for the specific participant doing the task (Levitt, 1971). Data corresponding to the exact numbers of trials in the calibration phase for each participant were not available. The minimum number in theory, based on the algorithm used for calibration, is 12 trials, but often the number of trials was substantially larger than this. After calibration, participants completed 10 practice trials that included confidence judgments, and then two blocks of 25 trials for the main task. Metacognitive sensitivity was computed with a hierarchical Bayesian framework (Fleming, 2017; Maniscalco & Lau, 2012). Each participant's meta-d' was estimated based on their responses, modeled using Markov Chain Monte Carlo (MCMC) as implemented in JAGS (version 3.4.0) within MATLAB.

During fMRI scanning, participants completed a decision-making task (Fig. 2). In each trial, two photographs showing protesters for or against various sociopolitical issues were presented and participants decided which of the two protest groups they supported more. All issues had been rated previously in the online survey and unfamiliar issues (familiarity rated *not at all familiar*) were excluded.

A total of 120 trials (5 runs * 24 trials per run) were included. The thumbs-up or thumbs-down icons next to the photos indicated whether the protesters supported or opposed the issue. In other words, a thumbs-down icon reverses the direction of an issue. Both support (thumbs-up) and opposition (thumbs-down) trials were included to achieve a larger range of overall support levels across trials

and to reduce the potential collinearity between support and moral conviction ratings. Support or opposition was always consistent for the two issues presented in a single trial and was randomized evenly between trials. Issues shown for each participant were tailored according to their prior ratings so that the support ratings of the two issues within a trial always differed. If there were more than 120 possible pairs of issues, a randomly selected subset of pairs was used, whereas if fewer than 120 distinct pairs were possible, some trials were repeated. All familiar issues were presented at least once during the decision-making task. The total number of familiar issues ranged from 24 to 40 with a mean of 36 and SD of 3.75. Trials were configured so that the issue with higher previously rated support was evenly distributed between the right and left sides of the screen. The first trial of each run started with a jittered fixation ranging from 3.09 s to 3.91 s, whereas other trials started with a jittered fixation ranging from 3.18 s to 4.82 s. Participants had 4 s to respond before the program would automatically proceed to the next trial. All stimuli were presented in E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA).

Behavioral analysis

To examine whether participants' demographic characteristics and dispositional traits affect their general tendency to moralize the sociopolitical issues presented, a mixed-effects linear regression model was fitted with moral conviction as the outcome variable. Intensity of support (calculated as |support—3| to capture deviation from the midpoint of the support scale), familiarity with the issue, religiosity, justice sensitivity, gender, age, education, income, party alignment, and political engagement were all included as fixed effects. For all behavioral analyses, gender was coded with two levels: 0 for nonfemale and 1 for female. Age was standardized. Education was converted to years of education: 12 for *high school diploma or GED* or *associates or technical degree*; 14 for *some college, but no degree*; 16 for *bachelor's degree*; and 18 for *graduate or professional degree (MA, MS, MBA, PhD, JD, MD, DDS,* etc.*)*. Income was coded as a continuous variable: 1 for *less than $25,000*; 2 for *$25,000–$49,000*; 3 for *$50,000–$74,999*; 4 for *$75,000-$99,999*; 5 for *$100,000-$149,999*; 6 for *$150,000 or more,* and participants who responded *prefer not to say* were treated as NA. Party alignment was treated as continuous from 0 (*Strong Democrat*) to 6 (*Strong Republican*). Political engagement ranged from 0 (*never*) to 4 (*always*), representing the extent to which participants followed politics and public affairs. Participants and issues were modeled with random intercepts respectively. Issues with which participants were not at all familiar were excluded from the analysis. While this analysis was conducted with data from 60 participants who completed the initial online survey and provided complete demographic data, other behavioral analyses relied on data collected during the scan session; for these analyses, data were available for the 38 participants who took part in the scanning part of the study and met the inclusion criteria for behavioral analyses including providing complete demographic data. Behavioral results from the latter set of analyses remained consistent in direction and significance when demographic covariates were removed from the analyses and the full sample of 45 participants was included.

A second behavioral analysis was a mixed-effects logistic regression used to test whether the relative protest support rating for the issue presented on the lefthand side of the screen (compared with the one presented on the righthand side) predicted a corresponding in-scanner choice of that issue. If an issue was presented in a thumbs-down form, support ratings were multiplied by − 1. The relative support was calculated as support rating (left) - support rating (right). The mixed-effects model included relative support as the primary fixed effect, with gender, age, education, income, party alignment, and political engagement entered as fixed-effect covariates of no interest. Participants were modeled with random intercepts.

Response time was analyzed as well with a mixed-effects linear regression. The moral conviction level of a trial was operationalized as the higher moral conviction rating of the two issues within a trial and denoted as the maximum moral conviction. In the model, fixed effects included maximum moral conviction, familiarity with the issue having maximum moral conviction, support difference (chosen – unchosen), the protesters' position (thumbs-up vs. thumbs-down), and the interaction between maximum moral conviction and support difference. As above, gender, age, education, income, party alignment, and political engagement were included as covariates of no interest, and participants were modeled with random intercepts. Only consistent choices were included in the analysis (3961 of 4823 trials that had responses) to better capture the cognitive processes underlying decisions that reflect participants' stated preferences, as inconsistent trials may involve different cognitive processes, such as increased uncertainty, attentional lapses, or decision error. To confirm the results, an additional mixed-effects linear regression model with the same specifications was fitted to the data from all trials that had responses.

MRI acquisition and analysis

MRI scanning was conducted with a 3.0 T Philips Achieva MRI scanner (Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel SENSE head coil at the University of Chicago MRI Research Center. First, high-resolution T1-weighted anatomical scans were acquired using a 3D MP-RAGE sequence (TR = 8 ms; TE = 3.5 ms; voxel size = $0.85 \times 0.85 \times 0.85$ mm^3; matrix = 284×260). Then, functional images were collected in ascending order and transverse slices using a single-shot EPI sequence with the

following parameters: voxel size = 3.0 × 3.1 × 3.0 mm^3, flip angle = 80°, matrix = 64 × 62, TR = 2000 ms, TE = 28 ms, field-of-view = 192 × 192 mm^2, slice gap = 0 mm). Each of the 5 runs acquired 135 volumes and lasted 4 min and 42 s. Participant attention to the task was monitored throughout the scan using an EyeLink 1000 Plus Eye Tracker (SR Research, Ontario, Canada), but eye-tracking data are not reported here.

Preprocessing

Data were preprocessed using fmriprep v22.1.1 (Esteban et al., 2020). The following four paragraphs are excerpted and adapted from the documentation distributed with fmriprep. The T1-weighted (T1w) image was corrected for intensity nonuniformity (INU) with N4BiasFieldCorrection (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM), and gray-matter (GM) was performed on the brain-extracted T1w using fast (part of FSL 6.0.5.1; Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.2.0; Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical GM of Mindboggle (Klein et al., 2017). Volume-based spatial normalization to standard space was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was used for spatial normalization: FSL's MNI ICBM 152 nonlinear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model (TemplateFlow ID: MNI152Nlin6Asym; Evans et al., 2012). A B0-nonuniformity map (or fieldmap) was estimated based on two echo-planar imaging (EPI) references with topup (FSL 6.0.5.1; Andersson et al., 2003).

For each BOLD run, the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1; Jenkinson et al., 2002). The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. BOLD runs were slice-time corrected to 0.975 s (0.5 of slice acquisition range 0–1.95 s) using 3dTshift from AFNI (Cox & Hyde, 1997). The BOLD reference was then coregistered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Coregistration was configured with 6 degrees of freedom.

A set of physiological regressors was extracted to allow for component-based noise correction (aCompCor) (Behzadi et al., 2007; Muschelli et al., 2014). Principal components were estimated after high-pass filtering of the preprocessed BOLD time-series (using a discrete cosine filter with 128 s cutoff). Probabilistic masks for CSF and WM were generated in anatomical space, and components were calculated separately within each mask.

The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152Nlin6Asym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. All resamplings can be performed with a single interpolation step by composing all pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and coregistrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed by using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Nongridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Functional MRI analyses

Univariate analyses

Data preprocessed in fMRIPrep were then analyzed by using FSL. Smoothing was applied using a 5-mm full-width half-maximum (FWHM) Gaussian kernel. At the first level, a general linear model (GLM) containing 5 to 6 primary regressors and the temporal derivative of each was applied to data from each scan run. The following primary regressors were modeled: 1) all trials with a response, 2) support mean, 3) support difference, 4) maximum moral conviction, 5) RT, and 6) trials with no recorded response (if present on a given run). Support mean and maximum moral conviction were chosen to represent the overall support level and moral conviction level, respectively, in each trial. The reasoning for using the mean of support ratings for the two causes is to reflect the overall subjective value of a trial. This design allows us to examine whether the support rating a participant assigns to the cause represented by a group of protesters is akin to assigning a valuation to that group of protesters. In other words, if support is akin to value, considering two sets of protesters who the participant on average agrees with would be processed as more rewarding than considering two sets of protesters whose cause is less aligned with the participants' views. The parametric

regressor for mean support quantifies this relationship. Cognitive neuroscience studies have found that brain activity in the valuation system, especially the vmPFC, tracks the sum/mean of the subjective value of multiple options in a decision (e.g., Hunt et al., 2012; Levy & Glimcher, 2011). Moral conviction is expected to show a somewhat different pattern: as soon as the decision-maker recognizes one highly morally convicted issue, heightened neural activity in brain regions that play a role in the detection of morally relevant information will be elicited, regardless of the moral conviction level of the other issue. The length of the boxcar was set to 4 s for all trials, and each trial was convolved with a double-gamma hemodynamic response function (HRF). Confound regressors representing the following variables, computed in fMRIPrep, were also included in the GLM: 6 motion parameters, 5 aCompCor parameters from WM, 5 aCompCor parameters from CSF (or fewer components if sufficient to account for 50% of variance in the CSF tissue compartment), and 3 cosine basis functions for high-pass filtering. This analysis was intended to identify the brain areas that track the average support ratings and the maximum moral conviction of the two issues shown.

Regressors computed at the first level were then combined across all runs in a fixed-effects analysis in FEAT, and contrasts were computed, yielding a single regressor for each contrast in each individual. Group-level results were computed in FEAT by combining data from each individual in a FLAME 1 and 2 mixed-effects analysis with automatic outlier detection. A voxel threshold of $z > 3.1$ was applied together with a cluster threshold of $p < 0.05$ implemented using Gaussian random field theory in FEAT.

Region of interest analyses

To examine whether regions of the valuation system are keeping track of the degrees of support and/or moral conviction, region of interest (ROI) analyses were implemented. The ROIs were defined as the vmPFC and VS ROIs identified in a meta-analysis of domain-general reward signal (Bartra et al., 2013). For each participant, activity was calculated by averaging the parameter estimate (COPE) of the parametric effect of support mean and maximum moral conviction from the second level FEAT, across all voxels within the two ROIs.

Functional connectivity analyses

A generalized psychophysiological interaction (gPPI) analysis (McLaren et al., 2012) was conducted to compare functional coupling between specific brain regions for trials varying in maximum moral conviction. Regressors for the gPPI analysis were constructed using AFNI. The average time series in the seed region was extracted from the data after fMRIPrep preprocessing and smoothing in FSL (using a 5-mm FWHM Gaussian kernel). This time series was then detrended and up-sampled by a factor of 20. A gamma function HRF was then deconvolved from the time series using the AFNI 3dTfitter command. This approach used a combination of penalty functions based on the raw time series and its first and second derivatives, a penalty weight of −2, and Lasso regularization with lambda = −6, following the suggested parameters from the 3dTfitter help file. To create the physiological regressor, the deconvolved time series of the ROI was reconvolved with a gamma HRF and down-sampled back to the TR length. The psychological regressor for gPPI (maximum moral conviction) and four to five additional regressors (all trials with a response, support mean, support difference, RT, and missed trials, if present) were computed by up-sampling the raw regressors, convolving them with a gamma HRF and down-sampling back to the TR length. The gPPI regressors were created by multiplying the up-sampled raw psychological regressor for maximum moral conviction with the deconvolved time series of the seed region, then convolving with a gamma HRF, and finally down-sampling back to the TR length.

For the first-level analysis in FSL, data that were preprocessed in fMRIPrep and smoothed by FSL (using a 5-mm FWHM Gaussian kernel) served as input. No additional preprocessing was done. The primary psychological regressor, physiological regressor, gPPI regressor and the four to five additional control psychological regressors listed above were modeled to fit a general linear model. The temporal derivatives for all psychological regressors were calculated and included in the model and temporal filtering was also applied to these regressors. Additional confound regressors computed in fMRIPrep (same as specified in the univariate analysis) were included. A second-level fixed effects analysis combined data from five runs and contrasts of interests were calculated for each participant. Finally, outputs from the second-level analysis were entered as inputs for the group level analysis, which was a FLAME1 and FLAME2 mixed-effects model using automatic outlier deweighting. Cluster thresholding was used with a $z > 3.1$ voxel threshold, and a $p < 0.05$ cluster threshold based on Gaussian random field theory.

Univariate effects and behavioral analyses

The univariate analyses mentioned above identified the brain network that was responsive to moral conviction level on sociopolitical issues. In analyses relating behaviors to brain activity in this network, for each participant, the parametric effects of maximum moral conviction were extracted from each cluster found in the whole-brain parametric modulation analyses. These effects were then averaged (after multiplying activity estimates by −1 in clusters showing negative effects) to form a single measure of moral conviction-related brain activity. Specifically, a mixed-effects logistic linear regression was fitted to test whether participants who showed a stronger effect

of maximum moral conviction in the univariate fMRI analysis also showed more consistency between the initial support rating and in-scanner decisions. Relative support, brain activity related to moral conviction, and the interaction of these two variables were included as the primary fixed effects. Another mixed-effects regression was used to examine whether higher maximum moral conviction led to greater reduction in decision time in those individuals who showed a stronger parametric effect of moral conviction in the univariate fMRI analysis. In this model, support difference (chosen – unchosen), the protesters' position (thumbs-up vs. thumbs-down), familiarity with the issue having maximum moral conviction, maximum moral conviction for the trial, moral conviction-related brain activity, and the interaction between the latter two variables were included as fixed effects. Only trials where the response in the scanner was consistent with earlier support ratings were included for the response time analysis. In both models, participants' gender, age, education, income, party alignment, and political engagement were entered as fixed-effect covariates of no interest. Participants were modeled with random intercepts.

Univariate effects and metacognition analyses

A series of Pearson correlation analyses was conducted to examine the relationships between metacognitive sensitivity and brain activity related to moral conviction and support levels for issues presented in the sociopolitical decisions. Brain activity related to maximum moral conviction or support mean was quantified in the same way described in the previous session such that the parametric effects of maximum conviction and support mean on fMRI BOLD signal, averaged across all significant clusters, were used as measures of moral conviction and support-related brain activity, respectively. Additional correlation analyses were implemented to test the relationships between metacognitive bias (general tendency to have higher confidence) and the brain activity associated with moral conviction. The Benjamini–Hochberg false discovery rate method was employed in cases where control for multiple comparisons was necessary (Benjamini & Hochberg, 1995).

Results

Behavioral

Results from the mixed-effects linear regression examining effects of demographic characteristics and dispositional traits on moral conviction rating indicated a significant positive association between religiosity and moral conviction [$B = 0.02$, 95% confidence interval [CI] (0.003, 0.044), $p = 0.04$] (Fig. 3). Additionally, regardless of whether the participant supported or opposed the issue, when their position about the issue in question was more extreme, they were more likely to have a higher level of moral conviction, while holding the other variables constant [$B = 0.56$, 95% CI (0.52, 0.61), $p < 0.001$]. Familiarity positively contributed to moral conviction as well [$B = 0.21$, 95% CI (0.17, 0.26), $p < 0.001$].

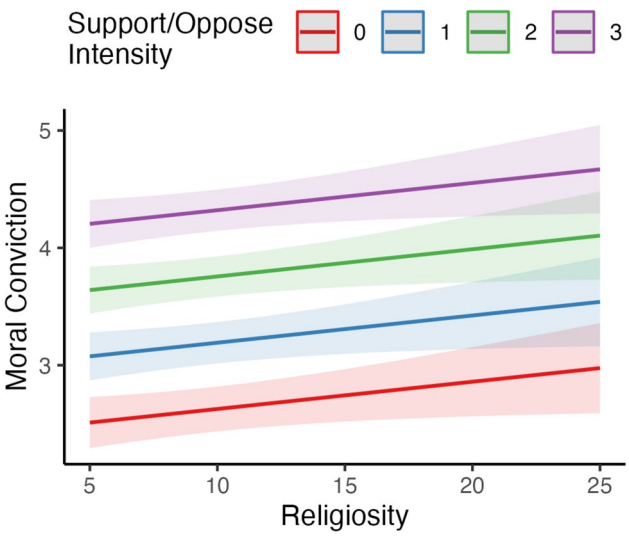

Fig. 3 Effect of religiosity on moral conviction. Shades illustrate 95% high density intervals

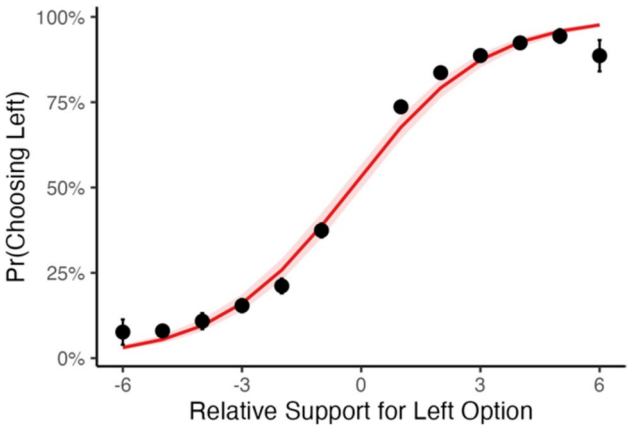

Fig. 4 Relationship between probability of choosing to support protesters on the left and the relative support rating of the issue on the left. Shade illustrates the 95% high density interval

Results from the logistic model examining the relationship between relative support and in-scanner choice showed that higher relative support for the issue on the left side of the screen led to a higher chance of choosing to support the protesters for that issue ($p < 0.001$, OR = 1.82, 95% CI = [1.76, 1.88]; Fig. 4). Furthermore, while support ratings from the pre-scan survey were not always perfectly consistent with choices made in the scanner, the pattern of results in Fig. 4 shows that inconsistent choices were more likely to occur on trials where the support rating difference between the two issues was

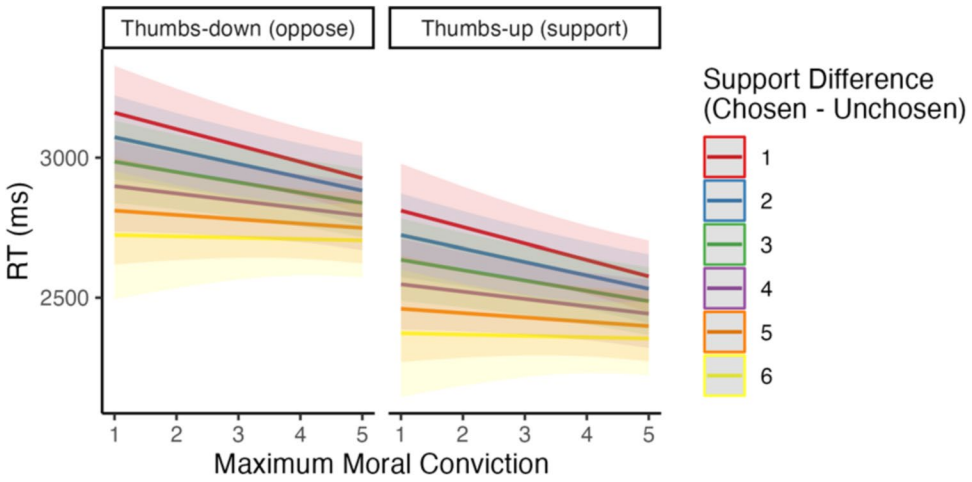

Fig. 5 Effects of maximum moral conviction, support difference, and protesters' position on response time. Shades illustrate 95% high density intervals

smaller rather than larger, supporting our assumption that in-scanner choices broadly align with earlier ratings. In a separate mixed-effects logistic regression, the main effects of relative support, maximum moral conviction, and their interaction were included as fixed effects, with the same covariates and the random effect. The interaction between relative support and maximum moral conviction was not significant ($p = 0.46$, OR = 1.02, 95% CI = [0.98, 1.06]).

The mixed-effects linear regression analysis demonstrated significantly shorter response times with both higher maximum moral conviction [$B = -69.33$, 95% CI ($-119.36, -19.24$), $p = 0.007$] and higher support difference [$B = -98.32$, 95% *CI* ($-164.48, -30.51$), $p = 0.004$] (Fig. 5). Participants also made decisions faster when the protesters supported the issue in question (thumbs-up) compared with when they opposed the issue (thumbs-down) [$B = -350.11$, 95% CI ($-389.09, -310.99$), $p < 0.001$]. Rated familiarity with the more morally convicted issue was associated with quicker decisions as well [$B = -26.98$, 95% CI ($-50.02, -4.65$), $p = 0.02$]. Numerically, the interaction between support difference and maximum moral conviction trended in the predicted direction, because the association between stronger

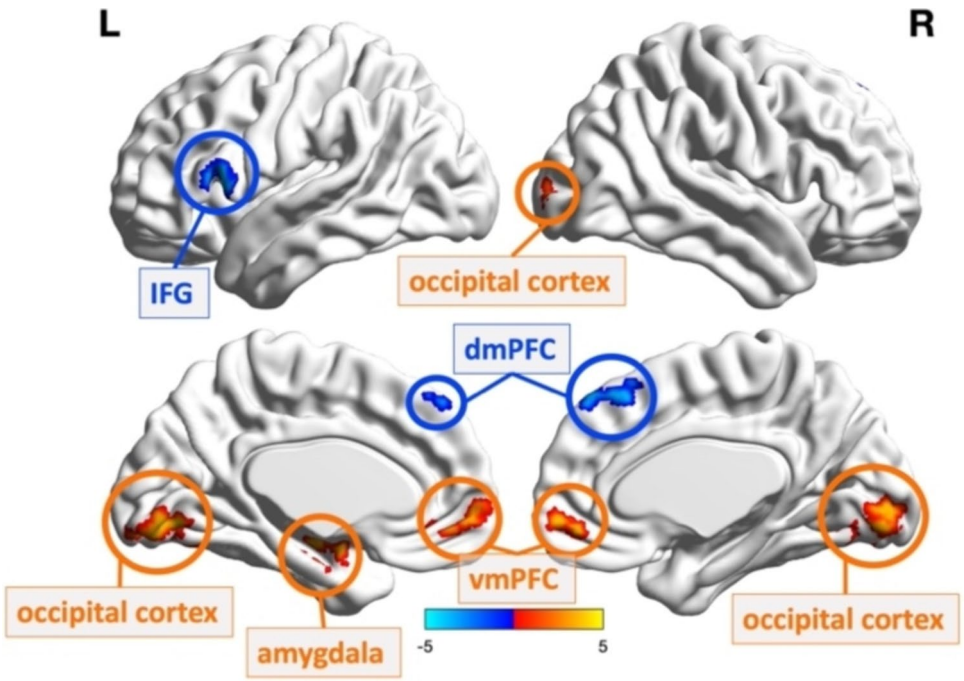

Fig. 6 Parametric effects of support of sociopolitical issues during decision-making

Table 1 Brain regions showing significant parametric effects of mean support ratings across the two issues within a given trial on hemodynamic responses

Brain region	Cluster size (Voxels)	Peak MNI coordinates			Peak *z* statistic
		x	*Y*	*z*	
Greater response					
Left occipital	760	−10	−74	−6	4.85
vmPFC	255	−2	50	−10	4.1
Left amygdala	120	−20	−4	−24	4.56
Hippocampus		−28	−10	−28	3.91
Weaker response					
dmPFC	112	0	44	48	4.27
Left IFG	88	−54	20	8	4.27

maximum moral conviction and faster response time was stronger when the support difference was smaller. However, this interaction was not significant [$B = 10.77$, 95% CI (−4.82, 26.11), $p = 0.17$]. In the model where all trials were included, effects of maximum moral conviction [$B = -36.75$, 95% CI (−64.60, −9.55), $p = 0.009$], support difference [$B = -40.55$, 95% CI (−77.19, −3.91), $p = 0.03$], protesters' position [$B = -327.30$, 95% CI (−362.74, −291.63), $p < 0.001$], and familiarity [$B = -34.27$, 95% CI (−55.10, −13.89), $p = 0.001$] were significant and in the same direction as the effects shown in the model with only consistent choices included.

Neural activity related to support for sociopolitical issues during decision-making

Elevated hemodynamic responses to decisions that had higher mean support ratings were found in the left occipital cortex, vmPFC, and left amygdala (Table 1; Fig. 6). The reverse contrast, which identified increased responses to lower mean support ratings, showed significant effects in the dorsal medial prefrontal cortex (dmPFC) and left inferior frontal gyrus (IFG) (Table 1; Fig. 6). All fMRI results in this paper were visualized with the BrainNet Viewer (Xia et al., 2013). An ROI analysis averaging across VS and vmPFC, using regions defined by a meta-analysis of domain-general reward (Bartra et al., 2013), also showed a significant parametric effect of mean support (t (43) = 3.00, $p = 0.004$). The analyses using VS and vmPFC as separate ROIs indicated a significant parametric effect of mean support in the vmPFC (t (43) = 3.35, $p = 0.002$, $p_{adjusted} = 0.003$) and not in the VS (t (43) = 1.32, $p = 0.19$, $p_{adjusted} = 0.19$), controlling for the false discovery rate using the Benjamini–Hochberg method.

Neural activity related to the moral conviction level of sociopolitical issues during decision-making

This parametric modulation analysis showed brain regions within which BOLD signal was greater or weaker based on the maximum moral conviction in a given trial. Left inferior frontal cortex, presupplementary motor area (SMA), lPFC, and bilateral aINS showed increased hemodynamic responses to trials with higher maximum moral conviction, whereas the left precuneus displayed decreased activity on trials with higher maximum moral conviction (Table 2; Fig. 7). An ROI analysis examining the role of VS and vmPFC in moral conviction (parametric effect of maximum moral conviction averaged across these regions) was not significant (t (43) = −0.79, $p = 0.43$). ROI analyses examining the VS and vmPFC separately showed a marginal trend toward a negative effect of maximum moral conviction on activity in the VS (t (43) = −2.06, $p = 0.046$, $p_{adjusted} = 0.09$) and no significant impact in the vmPFC (t (43) = 0.40, $p = 0.69$, $p_{adjusted} = 0.69$).

Moral conviction and functional connectivity using lPFC as the seed region

The cluster in lPFC that showed a significant response to maximum moral conviction in the univariate analysis was

Table 2 Brain regions showing significant parametric effects of maximum moral conviction ratings on hemodynamic response

Brain region	Cluster size (voxels)	Peak MNI coordinates			Peak z statistic
		X	y	z	
Greater response					
pre-SMA	234	0	16	58	4.39
Left inferior frontal cortex	230	−50	10	32	4.37
Left aINS	130	−42	20	−4	4.44
ACC	87	0	36	30	3.86
Left lPFC	70	−34	46	24	4.41
Weaker response					
Left precuneus	79	−10	−64	22	4.28

Table 3 Brain regions showing significant increases in functional connectivity with lPFC during decisions with higher maximum moral conviction

		Peak MNI coordinates			
Brain region	Cluster size (voxels)	*x*	*y*	*z*	Peak *z* statistic
vmPFC	224	0	52	−6	4.62
mPFC	141	4	52	14	3.97

selected as the seed region for a gPPI analysis. This choice was based on lPFC's pivotal role in complex cognitive functioning and value-based decision-making. Prior studies have suggested that lPFC interacts with brain regions including the vmPFC, ACC, aINS, and the posterior cingulate cortex (PCC) to achieve various goals, including those are morally relevant (Carlson & Crockett, 2018; Dixon & Christoff, 2014; Duverne & Koechlin, 2017). Thus, psychophysiological interaction analysis was conducted to examine whether functional connectivity with lPFC—specifically with the cluster responsive to moral conviction—varied depending on the moral conviction level of a sociopolitical decision. The analysis demonstrated that, during decisions with higher maximum moral conviction, there is significantly stronger functional connectivity between the lPFC seed cluster and the vmPFC and mPFC (Fig. 8, Table 3).

Neural activity related to moral conviction, choice consistency, and decision time

The logistic regression model used to examine how relative support and the parametric effect of moral conviction on brain activity together predict in-scanner choices showed that the relative level of support for an issue in the pre-scan survey was associated with a higher chance of choosing to support the protesters corresponding to that issue ($p < 0.001$, OR = 1.70, 95% CI = [1.63, 1.78]). A significant interaction was found between this effect and the participant's parametric effect of moral conviction on brain activity during these choices ($p < 0.001$, OR = 1.02, 95% CI = [1.008, 1.024]; Fig. 9A). Specifically, individuals whose neural responses were more strongly modulated by moral conviction showed greater alignment between the pre-scan support ratings and in-scanner decisions.

The mixed-effects linear regression model predicting decision time showed that individuals with a greater parametric effect of moral conviction on brain activity had slower response times overall [$B = 34.91$, 95% CI (4.67, 64.30), $p = 0.03$], but this effect was modulated by a significant interaction with trial-level maximum moral conviction

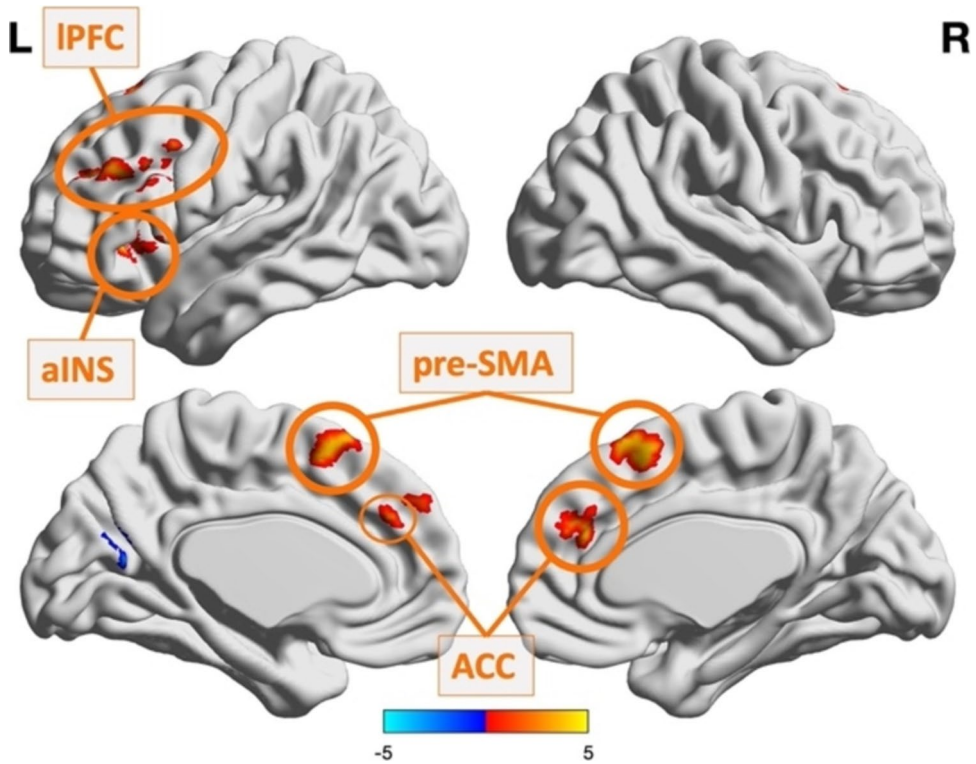

Fig. 7 Parametric effects of moral conviction of sociopolitical issues during decision-making

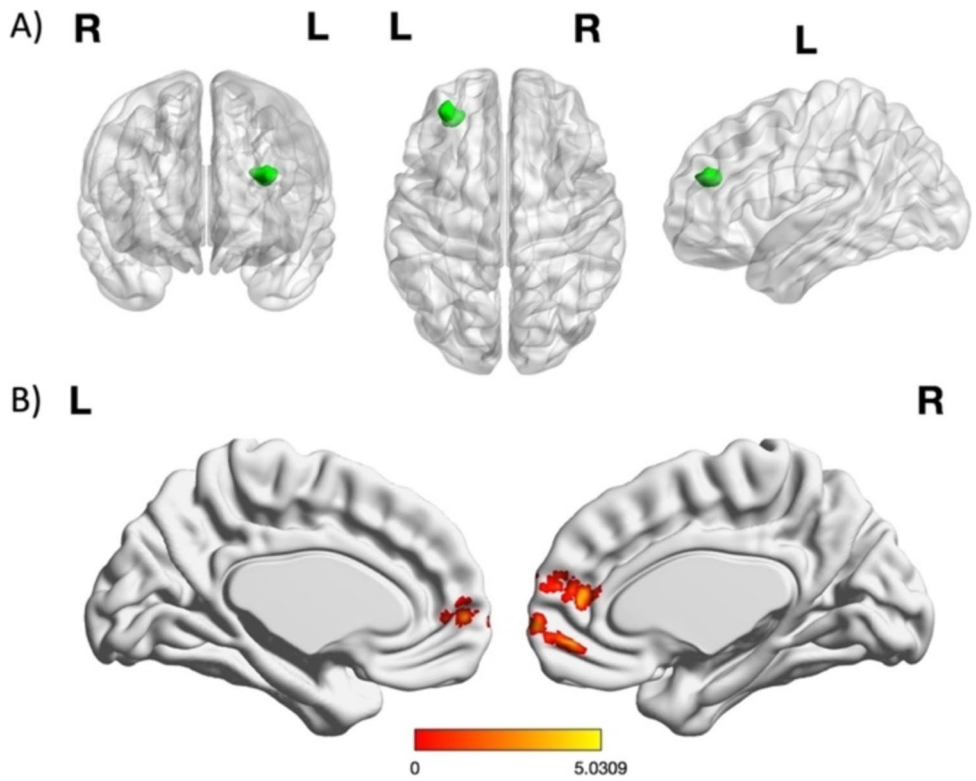

Fig. 8 **A**) lPFC seed region and **B**) gPPI connectivity higher maximum moral conviction

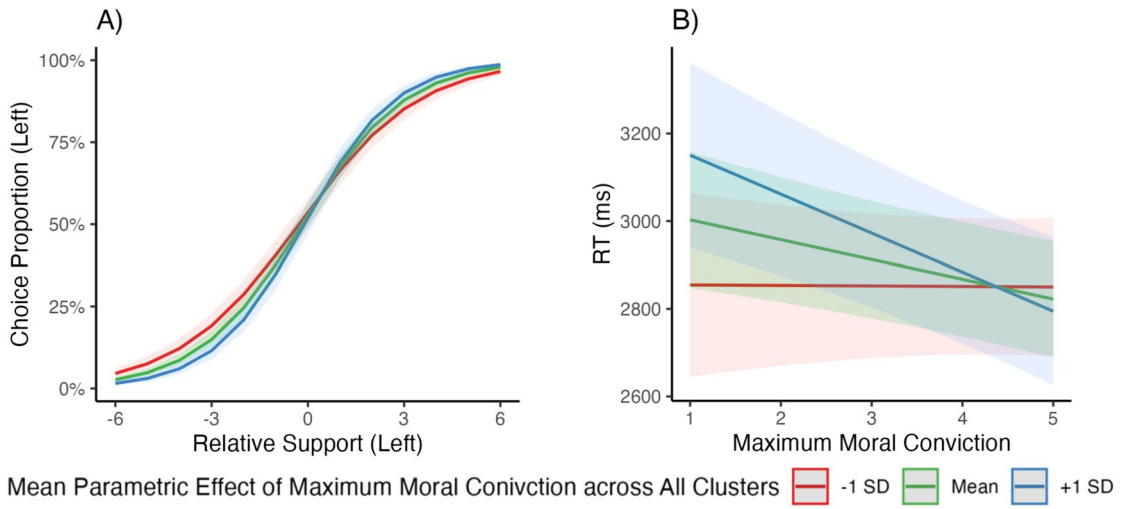

Fig. 9 **A**) Relationship between probability of choosing to support protesters on the left side of the screen and the relative support rating of the issue from the pre-scan survey, moderated by the strength of an individual's parametric effect of moral conviction on brain response in regions shown in Fig. 7. **B**) Relationship between decision time and maximum moral conviction on each trial, moderated by the individual's parametric effect of moral conviction on brain response in regions shown in Fig. 7. Shadings indicate 95% high-density intervals

[$B = -7.99$, 95% CI ($-14.08, -1.70$), $p = 0.01$]. Specifically, in participants with stronger parametric modulation of neural responses to moral conviction, higher maximum moral conviction was associated with faster decisions, whereas this effect was not observed in those with weaker parametric modulation in response to moral conviction (Fig. 9B). Shorter response times were also associated with a larger difference in support between the chosen and nonchosen choice options [$B = -50.11$, 95% CI ($-63.20, -36.39$), $p < 0.001$], greater familiarity with the issue having higher

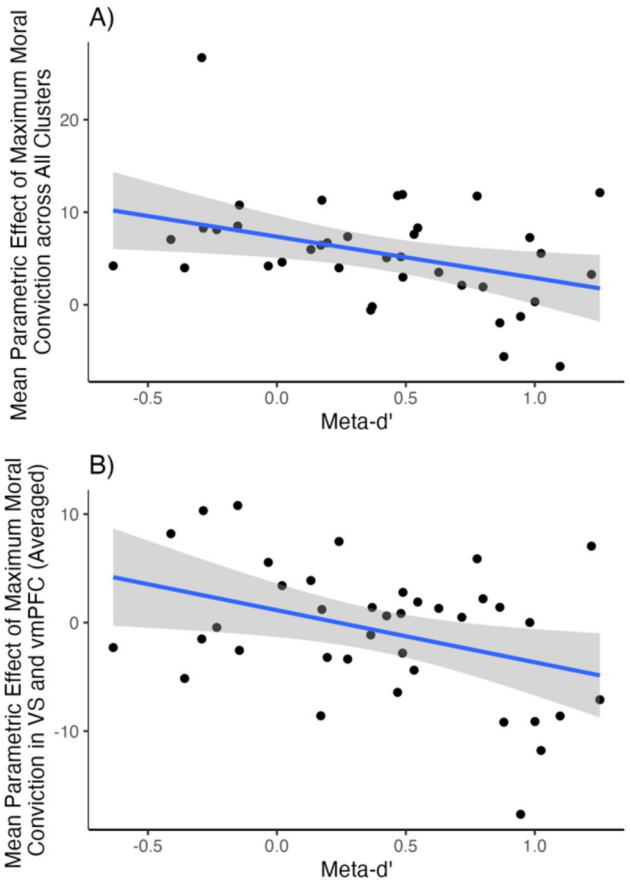

Fig. 10 Negative correlations between meta-d' and the parametric effects of moral conviction. **A**) Across all responsive regions **B**) In cortical valuation regions (VS and vmPFC)

moral conviction [B = −30.41, 95% CI (−53.67, −8.01), $p = 0.009$], and when protesters are in support (thumbs-up) rather than opposed to (thumbs-down) the issue in question [B = −345.73, 95% CI (−385.14, −306.13), $p < 0.001$].

Meta-d' and parametric effects of moral conviction and support of sociopolitical issues

No significant correlation was found between an individual's metacognitive sensitivity (measured by meta-d') and support-related brain activity [r (36) = 0.19, $p = 0.24$], but a significant correlation was observed between metacognitive sensitivity and brain activity associated with maximum moral conviction (Fig. 10A) [r (36) = −0.38, $p = 0.02$]. To confirm that this observed effect was not being driven by a relationship between overconfidence and dogmatic belief, the mean confidence rating on the perceptual confidence task was calculated for each participant as a measure of metacognitive bias and was correlated with brain activity related to maximum moral conviction. No significant correlation was found [r (36) = −0.14, $p = 0.39$]. An effect driven by overconfidence would imply a positive correlation, which clearly is not apparent here. Finally, to clarify whether the neural response in specific clusters drove the negative association between meta-d' and the mean parametric effect of moral conviction, Pearson correlation analyses were performed to test the relationship between metacognitive sensitivity and the parametric effect of moral conviction in each of the six clusters. Results showed that meta-d' was negatively associated with the neural activity in the left lPFC [r_{lPFC} (36) = −0.44, $p = 0.005$, $p_{adjusted} = 0.03$] and ACC [r_{ACC} (36) = −0.35, $p = 0.03$, $p_{adjusted} = 0.09$], although the latter effect was not significant after controlling for the false discovery rate. No significant effects (uncorrected $p < 0.05$) were found in any other areas. These results suggest that lPFC, and potentially ACC, are driving the effects across the full network.

The role of valuation regions was further examined by correlating the parametric effects of moral conviction and support in the valuation network (by averaging responses from the VS and vmPFC) with metacognitive sensitivity. A significant negative correlation was found between metacognitive sensitivity and the parametric effect of moral conviction in reward-sensitive regions [r (36) = −0.38, $p = 0.02$] (Fig. 10B). When examining the two reward-sensitive regions separately, a significant negative association between metacognitive sensitivity and the parametric effects of moral conviction in the VS was found [r (36) = −0.32, $p = 0.0496$, $p_{adjusted} = 0.06$], although this effect was not significant after controlling for false discovery rate. The association between metacognitive sensitivity and the parametric effect of moral conviction in the vmPFC also trended in the negative direction [r (36) = −0.31, $p = 0.06$, $p_{adjusted} = 0.06$]. No significant correlations were found between metacognitive sensitivity and the parametric effect of mean support in the valuation network (either for VS and vmPFC individually or for their average) (combined: r (36) = −0.13, $p = 0.42$; VS: r (36) = −0.14, $p = 0.41$, $p_{adjusted} = 0.59$; vmPFC: r (36) = −0.09, $p = 0.59$, $p_{adjusted} = 0.59$).

Discussion

Moral convictions are perceived as absolute, universal, and definite beliefs or principles. They can operate as moral imperatives that delineate which opinions, actions, and policies are right or wrong, as well as motivate collective actions (Decety, 2024). By integrating fMRI data and behavioral measurements, this study provides new evidence about the neural and cognitive underpinnings of moral conviction, including its relationship with metacognitive abilities.

In keeping with previous research in social psychology (Van Bavel et al., 2012), the behavioral data suggest that moral conviction serves as an indicator of choice significance, because it

leads participants to make faster decisions about highly moralized items. These effects persisted when controlling for the difference between the support levels of the two protest groups and the level of familiarity of the highly moralized issue, demonstrating that moral conviction is more than just attitude strength or familiarity. While previous research has shown that moral evaluations are faster than nonmoral ones (Van Bavel et al., 2012; Van Berkum et al., 2009), the current study addresses how different degrees of moral conviction affect decision-making about timely sociopolitical issues that have strong real-life implications and shows that content that evokes high levels of moral conviction are processed faster.

A positive association was found between participants' religiosity and their moral conviction ratings of the issues presented, controlling for participants' support level, justice sensitivity, age, gender, education, income, party alignment, and political engagement. This finding is consistent with past work that shows that religion promotes *righteous* morality over *prosocial* morality and with prior findings that religious morality is primarily deontological and nonconsequentialist (Saroglou & Craninx, 2021). While some studies have found evidence for a distinction between moral and religious convictions (Skitka et al., 2018), others suggest that they do not differ categorically (Brownlee, 2017; Cousar et al., 2021). Results from the current study support that participants with high religiosity potentially have a greater tendency to think about sociopolitical issues from a righteous point of view, which leads to higher overall moral conviction. Further investigation is needed to clarify the relationships between moral conviction and religious attitudes.

Consistent with our theoretical perspective, whole-brain univariate analyses suggest a twofold neurocognitive process underlying moral conviction. An emotional component is reflected by increased neural activity in the salience network, which includes the insula and ACC, in response to high moral conviction. This component signals the salience of morally convicted items, which may subsequently regulate downstream cognitive functions that underlie moral reasoning and decisions. This result aligns with studies showcasing the importance of the salience network in the detection of morally charged information, as well as with the role played by the ACC and the inferior frontal gyrus in distinguishing moral versus conventional norms (Eres et al., 2018; Sevinc et al., 2017; White et al., 2017). The results also indicate that the cluster in lPFC is more activated in the context of high levels of moral conviction, possibly supporting the cognitive dimension of moral conviction. Previous research showed the implication of lPFC in various functions, including cognitive control, executive function, planning, social cognition, and moral judgment (Forbes & Grafman, 2010; Miller & Cohen, 2001). When it comes to moral cognition, one study reported that disruption of the right dlPFC by transcranial magnetic stimulation causes participants to act less fairly toward a social partner (Knoch et al., 2009). Another study showed that patients with dlPFC damage are less likely to cooperate in a public goods game (Wills et al., 2018). Together, these studies suggest that the dlPFC exerts self-control and inhibits selfish behaviors. In other contexts, however, increased neural response in the lPFC has been associated with dishonest and selfish behaviors (FeldmanHall et al., 2012; Greene & Paxton, 2009). Thus, results from the present study add to our knowledge of the role of the lPFC, demonstrating that it may not be limited to inhibiting impulsive behavior but instead flexibly engages in upholding and enforcing goals (Carlson & Crockett, 2018; Tusche & Hutcherson, 2018). The cognitive dimension of moral conviction reflects the ability to distinguish morally convicted beliefs from beliefs lacking moral conviction and underscores their perceived objectivity grounded by universal and unalterable facts that transcend personal and social boundaries (Wright et al., 2008). This cognitive aspect shapes moral goals and serves as the key to aligning actions with these aims. This conceptual overlap between achieving goals and the imperative nature of moral conviction suggests that the lPFC is a compelling candidate region for implementing the cognitive dimension of moral conviction.

While behavioral ratings of moral conviction showed no significant effect on the consistency between pre-scan support ratings and in-scanner decisions, neural activity associated with moral conviction did moderate the relationship between pre-scan support ratings and these same in-scanner decisions. Specifically, individuals who exhibited a stronger parametric relationship with moral conviction in the brain regions identified across the sample (Fig. 7) also showed greater consistency between their pre-scan survey responses and their choices in the scanner. The strength of the same parametric brain response to moral conviction also modulated the relationship between moral conviction and decision time. Higher moral conviction was associated with shorter decision times particularly in those individuals whose brain responses were more strongly modulated by moral conviction. These results suggest that individual differences in the neural encoding of moral conviction reflect moral significance to a degree that behavioral ratings cannot. They highlight the unique contribution that neural measures can make toward explaining judgment and decision-making, beyond what is possible with purely behavioral and self-report measures.

A gPPI analysis using as the seed region the lPFC cluster parametrically modulated by maximum moral conviction in the univariate analysis demonstrates stronger functional connectivity with vmPFC and mPFC during decisions with higher moral conviction. In cognitive tasks, the functional coupling between dlPFC and vmPFC has been linked to the successful exertion of self-control in choosing larger-delayed rewards over smaller-immediate ones and in choosing healthier over tastier food items (Hare et al., 2009, 2014). Dorsolateral PFC and vmPFC interactions also have been shown to contribute to

adaptive value calculations in different contexts, in the absence of a requirement for self-control (Rudorf & Hare, 2014). Therefore, it is likely that the lPFC-vmPFC connectivity signals an increase in the importance of domain-general goals (e.g., health goals, moral goals) and greater incorporation of these goals into the value-based decision-making process (Brocas & Carrillo, 2021). Taken together with the univariate results that the lPFC tracks moral conviction while vmPFC and amygdala track overall support, the increase in functional coupling between lPFC and vmPFC during higher moral conviction decisions suggests an increased integration of moral considerations and sociopolitical opinions. In line with this interpretation, one previous study examining fairness and costly punishment found increased connectivity between right dlPFC and posterior vmPFC in people who more frequently decided to exert costly punishment (Baumgartner et al., 2011), indicating that such neural connectivity may underlie the enforcement of moral conviction related to fairness.

Metacognitive sensitivity moderated neural responses to moral conviction level during decision-making, particularly in left lPFC and potentially in ACC. Our data provide no evidence of a positive correlation between metacognitive bias and neural responses related to moral conviction, indicating that the relationship with moral conviction is specific to low metacognitive sensitivity and not overconfidence. In previous literature, the lPFC has been implicated in metacognitive judgments (Fleming & Dolan, 2012; Lapate et al., 2020). The present study is the first to demonstrate that metacognition and moral conviction may rely on overlapping circuitry in lPFC. Although the finding that metacognitive sensitivity moderated neural response to moral conviction in ACC was no longer significant after correction for multiple comparisons, such an effect would be consistent with previous findings. Specifically, reduced metacognitive sensitivity has been associated with stronger medial frontal negativity (MFN) elicited as a function of moral conviction level for highly moralized issues (Yoder & Decety, 2022), and the MFN is thought to originate in the ACC (Gehring & Willoughby, 2002). Together, these findings contribute to a growing body of evidence highlighting the role of the lPFC and ACC in metacognitive abilities and moral decision-making. Future research is needed to investigate the underlying neural mechanisms at a more granular level.

Moral conviction showed no significant main effect on activity in the valuation system (vmPFC and VS). Interestingly, though, the magnitude of this effect across individuals was negatively associated with metacognitive sensitivity. In other words, compared with individuals with higher metacognitive sensitivity, those with lower metacognitive sensitivity exhibited greater activity in the vmPFC and VS during trials with higher maximum moral conviction. Thus, moral conviction may be seen as rewarding specifically for those with poor metacognitive abilities. Meanwhile, results from the whole-brain univariate analyses and ROI analyses demonstrated that greater mean support for social issues was associated with greater activity in vmPFC and amygdala. These are key regions in the valuation system, suggesting that moral issues with which one agrees are among the many domains in which liking is indexed in the reward circuit (Hare et al., 2008; Mormann et al., 2019). The effects of mean support in the reward system showed no significant correlation with metacognitive sensitivity. Overall, these results provide evidence that the extent to which the valuation system is involved in moral conviction may depend on individuals' metacognitive abilities while its role in tracking level of support is more universal. Past behavioral studies have linked lower metacognition to dogmatism, political radicalism, and reduced social conformity (Osorio & Reyes, 2023; Rollwage et al., 2018; Yoder & Decety, 2022). The current study suggests a possible neural mechanism through which metacognition regulates moral conviction's impact.

One limitation of the current study is that multiple cognitive processes may have occurred simultaneously, increasing the difficulty of unambiguously relating brain activity with specific cognitive processes. For instance, to what extent is attitude strength distinct from moral conviction? Does the brain's valuation system simply track attitude strength and how rewarding it is to act on a morally convicted attitude, or does it also distinguish between actions that are morally neutral and those driven by strong moral conviction? Moreover, higher moral conviction toward an issue may lead to differences in key subprocesses of decision-making, such as working memory load, attention allocation, and evidence accumulation—all of which are important for metacognition. Future studies should be designed to clarify the basic cognitive processes that play a role in moral conviction, to better understand the neurocognitive mechanisms by which metacognitive sensitivity influences morally charged sociopolitical decisions. In addition, while emotion is a characteristic of moral beliefs and moral conviction, emotional salience does not necessitate moral conviction. Additional questions remain, to be more thoroughly addressed in future work, about this distinction (Avramova & Inbar, 2013; Wright et al., 2008).

Moral values and social norms are crucial for fostering cooperation, which is the cornerstone of human civilization. Moral conviction specifically influences opinions, attitudes, and behaviors, leading to a range of outcomes from positive collective action to dogmatism, intolerance, and violence (Decety, 2024). Past studies have established a link between strong moral conviction and social and political intolerance, which exacerbates the distinction between "us" and "them" and potentially leads to the approval of sociopolitical violence aligned with ones' own ideological views (Garrett & Bankert, 2020; Pretus et al., 2023; Workman et al., 2020; Yoder & Decety, 2022). This study builds on existing knowledge by delineating the mechanisms through which support, moral conviction, and metacognitive sensitivity guide decisions about sociopolitical protests and provides a basis for future research investigating the psychological roots of political and social action.

Conclusions

The study sheds light on the neural mechanisms underlying how moral conviction guides decisions on sociopolitical issues and interacts with metacognitive abilities. Moral conviction is associated with greater neural activity in core regions of the salience network (i.e., ACC and aINS), as well as in regions associated with executive functioning (lPFC and pre-SMA). These neural mechanisms support the emotional and cognitive dimensions of moral conviction, respectively. We propose that regions in the salience network encode the emotional intensity of morally convicted issues while the lPFC plays an important role in recognizing moral goals and signaling the objectivity of moral conviction. Increased coupling between the lPFC and vmPFC as a function of moral conviction, as revealed by a gPPI functional connectivity analysis, suggests a potential decision mechanism by which moral goals tracked by the lPFC are incorporated into the valuation process to a greater extent when a decision involves sociopolitical issue(s) with stronger moral conviction. The neural responses associated with moral conviction are also stronger in individuals who score lower on metacognitive sensitivity. Activity in vmPFC and VS associated with moral conviction, although not significant in the aggregate, is stronger in individuals with lower metacognitive sensitivity as well. These findings may provide a mechanistic explanation for the increasingly documented observation in cognitive science that beliefs tend to be more rigid in people with poor metacognitive performance. Activity in reward circuitry tracks the value of the decision options and is associated with the overall level of support for the two decision options. This effect is uncorrelated with metacognitive sensitivity, suggesting that activity in the valuation system tracks support regardless of metacognitive abilities. Overall, the study provides new insight into the cognitive and neural mechanisms of moral conviction and provides a basis for further examination of the interplay between moral conviction and metacognition.

Funding No funding was received for conducting this study.

Data availability Raw fMRI data for this project is available at https://openneuro.org/datasets/ds005040/. Behavioral data and materials can be accessed at OSF https://osf.io/rw6y4/.

Code availability Codes for analyses can be accessed at OSF https://osf.io/rw6y4/.

Declarations

Ethics approval This experiment was approved by the Institutional Review Board at the University of Chicago.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication All participants have consented to publish the results of this study and share information from this study.

Conflicts of interest/Competing interests All authors declare not having any conflict of interest concerning this work.

References

Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage, 20*, 870–888.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis, 12*, 26–41.

Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment. *Wires Cognitive Science, 4*(2), 169–178.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427.

Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., & Rammstedt, B. (2014). Measuring four perspectives of justice sensitivity with two items each. *Journal of Personality Assessment, 96*(3), 380–390.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience, 14*(11), 1468–1474.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage, 37*, 90–101.

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

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313–7318.

Brocas, I., & Carrillo, J. D. (2021). Value computation and modulation: A neuroeconomic theory of self-control as constrained optimization. *Journal of Economic Theory, 198*, 105366.

Brownlee, K. (2017). Is religious conviction special? In C. Laborde & A. Bardon (Eds.), *Religion in liberal political philosophy* (pp. 309–320). Oxford Academic.

Carlson, R. W., & Crockett, M. J. (2018). The lateral prefrontal cortex and moral goal pursuit. *Current Opinion in Psychology, 24*, 77–82.

Chen, T., Cai, W., Ryali, S., Supekar, K., & Menon, V. (2016). Distinct global brain dynamics and spatiotemporal organization of the salience network. *PLOS Biology, 14*(6), e1002469.

Clithero, J. A., Smith, D. V., Carter, R. M., & Huettel, S. A. (2011). Within- and cross- participant classifiers reveal different neural coding of information. *NeuroImage, 56*(2), 699–708.

Cousar, K. A., Carnes, N. C., & Kimel, S. Y. (2021). Morality as fuel for violence? Disentangling the role of religion in violent conflict. *Social Cognition, 39*(1), 166–182.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine, 10*, 171–178.

Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P., & Dolan, R. J. (2017). Moral transgressions corrupt neural representations of value. *Nature Neuroscience, 20*(6), 879–885.

Cunningham, W. A., & Brosch, T. (2012). Motivational salience: Amygdala tuning from traits, needs, values, and goals. *Current Directions in Psychological Science, 21*(1), 54–59.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. *Segmentation and Surface Reconstruction. Neuroimage, 9*, 179–194.

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, or violence. *Proceedings of the Paris Institute for Advanced Study, 1*, 1–80.

Decety, J., & Cowell, J. M. (2018). Interpersonal harm aversion as a necessary foundation for morality: A developmental neuroscience perspective. *Development and Psychopathology, 30*(1), 153–164.

Dixon, M. L., & Christoff, K. (2014). The lateral prefrontal cortex and complex value-based learning and decision making. *Neuroscience & Biobehavioral Reviews, 45*, 9–18.

Duverne, S., & Koechlin, E. (2017). Rewards and cognitive control in the human prefrontal cortex. *Cerebral Cortex, 27*(10), 5024–5039.

Eres, R., Louis, W. R., & Molenberghs, P. (2018). Common and distinct neural networks involved in fMRI studies investigating morality: An ALE meta-analysis. *Social Neuroscience, 13*(4), 384–398.

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS ONE, 12*, e0184661.

Esteban, O., Ciric, R., Finc, K., et al. (2020). Analysis of task-based functional MRI data preprocessed with fMRIPrep. *Nature Protocols, 15*, 2186–2202.

Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases. *NeuroImage, 62*, 911–922.

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*(7), 743–751.

FeldmanHall, O., & Mobbs, D. (2015). A neural network for moral decision making. In *Brain Mapping* (pp. 205–210). Elsevier.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science, 370*(6516), 533–536.

Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness, 2017*(1), 1–14.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1338–1349.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*, 1–9.

Forbes, C. E., & Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review of Neuroscience, 33*(1), 299–324.

Garrett, K. N. (2019). Fired up by morality: The unique physiological response tied to moral conviction in politics. *Political Psychology, 40*(3), 543–563.

Garrett, K. N., & Bankert, A. (2020). The moral roots of partisan division: How moral conviction heightens affective polarization. *British Journal of Political Science, 50*(2), 621–640.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279–2282.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition, 106*(3), 1339–1366.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences, 106*(30), 12506–12511.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage, 48*, 63–72.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience, 30*(2), 583–590.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*(5927), 646–648.

Hare, T. A., Hakimi, S., & Rangel, A. (2014). Activity in dlPFC and its effective connectivity to vmPFC are associated with temporal discounting. *Frontiers in Neuroscience, 8*(50), 1–14.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *The Journal of Neuroscience, 28*(22), 5623–5630.

Hesse, E., Mikulan, E., Decety, J., Sigman, M., Garcia, M. D. C., Silva, W., Ciraolo, C., Vaucheret, E., Baglivo, F., Huepe, D., Lopez, V., Manes, F., Bekinschtein, T. A., & Ibanez, A. (2016). Early detection of intentional harm in the human amygdala. *Brain, 139*(1), 54–61.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience, 15*(3), 470–476.

Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *Journal of Neuroscience, 35*(36), 12593–12605.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17*, 825–841.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., et al. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology, 13*, e1005350.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Sciences, 106*(49), 20895–20899.

Koenig, H. G., & Büssing, A. (2010). The Duke university religion index (DUREL): A five-item measure for use in epidemological studies. *Religions, 1*(1), 78–85.

Krueger, F., & Hoffman, M. (2016). The emerging neuroscience of third-Party punishment. *Trends in Neurosciences, 39*(8), 499–501.

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis, 1*, 76–85.

Lapate, R. C., Samaha, J., Rokers, B., Postle, B. R., & Davidson, R. J. (2020). Perceptual metacognition of human faces is causally supported by function of the lateral prefrontal cortex. *Communications Biology, 3*(1), 360.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2B), 467–477.

Levy, D. J., & Glimcher, P. W. (2011). Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain. *The Journal of Neuroscience, 31*(41), 14693–14707.

Lin, H., Müller-Bardorff, M., Gathmann, B., Brieke, J., Mothes-Lasch, M., Bruchmann, M., Miltner, W. H. R., & Straube, T. (2020). Stimulus arousal drives amygdalar responses to emotional expressions across sensory modalities. *Scientific Reports, 10*(1), 1898.

Luttrell, A., Petty, R. E., Briñol, P., & Wagner, B. C. (2016). Making it moral: Merely labeling an attitude as moral increases its strength. *Journal of Experimental Social Psychology, 65*, 82–93.

Luttrell, A., & Togans, L. J. (2021). The stability of moralized attitudes over time. *Personality and Social Psychology Bulletin, 47*(4), 551–564.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430.

Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020). Is there a g factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *Journal of Experimental Psychology: General, 149*, 1788–1799.

Marie, A., Altay, S., & Strickland, B. (2023). Moralization and extremism robustly amplify myside sharing. *PNAS Nexus*, *2*(4), pgad078.

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage, 61*, 1277–1286.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167–202.

Mormann, F., Bausch, M., Knieling, S., & Fried, I. (2019). Neurons in the human left amygdala automatically encode subjective value irrespective of task. *Cerebral Cortex, 29*(1), 265–272.

Muschelli, J., Nebel, M. B., Caffo, B. S., Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2014). Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage, 96*, 22–35.

Osorio T., H., & Reyes M., G. (2023). Decision making in moral judgment context is modulated by individual metacognition. *Psychological Reports*, 00332941231191067.

Pauls, I. L., Shuman, E., Van Zomeren, M., Saguy, T., & Halperin, E. (2022). Does crossing a moral line justify collective means? Explaining how a perceived moral violation triggers normative and nonnormative forms of collective action. *European Journal of Social Psychology, 52*(1), 105–123.

Pretus, C., Ray, J. L., Granot, Y., Cunningham, W. A., & Van Bavel, J. (2023). The psychology of hate: Moral concerns differentiate hate from dislike. *European Journal of Social Psychology, 53*, 336–353.

Qu, C., Bénistant, J., & Dreher, J.-C. (2022). Neurocomputational mechanisms engaged in moral choices and moral learning. *Neuroscience & Biobehavioral Reviews, 132*, 50–60.

Qu, C., Hu, Y., Tang, Z., Derrington, E., & Dreher, J.-C. (2020). Neurocomputational mechanisms underlying immoral decisions benefiting self or others. *Social Cognitive and Affective Neuroscience, 15*(2), 135–149.

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology, 28*(24), 4014-4021.e8.

Rudorf, S., & Hare, T. A. (2014). Interactions between dorsolateral and ventromedial prefrontal cortex underlie context-dependent stimulus valuation in goal-directed choice. *The Journal of Neuroscience, 34*(48), 15988–15996.

Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics, 76*(2), 380–397.

Ryan, T. J. (2019). Actions versus consequences in political arguments: Insights from moral psychology. *The Journal of Politics, 81*(2), 426–440.

Saroglou, V., & Craninx, M. (2021). Religious moral righteousness over care: A review and a meta-analysis. *Current Opinion in Psychology, 40*, 79–85.

Seamans, J. K., & Floresco, S. B. (2022). Event-based control of autonomic and emotional states by the anterior cingulate cortex. *Neuroscience & Biobehavioral Reviews, 133*, 104503.

Sevinc, G., Gurvit, H., & Spreng, R. N. (2017). Salience network engagement with the detection of morally laden information. *Social Cognitive and Affective Neuroscience, 12*(7), 1118–1127.

Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of Neuroscience, 34*(13), 4741–4749.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895–917.

Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The Psychology of moral conviction. *Annual Review of Psychology, 72*, 347–366.

Skitka, L. J., Hanson, B. E., Washburn, A. N., & Mueller, A. B. (2018). Moral and religious convictions: Are they the same or different things? *PLoS ONE, 13*(6), e0199311.

Skitka, L. J., & Morgan, G. S. (2014). The social and political implications of moral conviction. *Political Psychology, 35*(S1), 95–110.

Soutschek, A., Sauter, M., & Schubert, T. (2015). The importance of the lateral prefrontal cortex for strategic decision making in the prisoner's dilemma. *Cognitive, Affective, & Behavioral Neuroscience, 15*(4), 854–860.

Thomas, E. F., Bury, S. M., Louis, W. R., Amiot, C. E., Molenberghs, P., Crane, M. F., & Decety, J. (2019). Vegetarian, vegan, activist, radical: Using latent profile analysis to examine different forms of support for animal welfare. *Group Processes & Intergroup Relations, 22*(6), 836–857.

Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. *eLife*, *7*, e31185.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging, 29*, 1310–1320.

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience, 16*(1), 55–61.

Ugazio, G., Grueschow, M., Polania, R., Lamm, C., Tobler, P., & Ruff, C. (2022). Neuro-computational foundations of moral preferences. *Social Cognitive and Affective Neuroscience, 17*(3), 253–265.

Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PLoS ONE, 7*(11), e48693.

Van Berkum, J. J., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science, 20*(9), 1092–1099.

Wang, S., Yu, R., Tyszka, J. M., Zhen, S., Kovach, C., Sun, S., Huang, Y., Hurlemann, R., Ross, I. B., Chung, J. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2017). The human amygdala parametrically encodes the intensity of specific facial emotions and their categorical ambiguity. *Nature Communications, 8*(1), 14821.

White, S. F., Zhao, H., Leong, K. K., Smetana, J. G., Nucci, L. P., & Blair, R. J. R. (2017). Neural correlates of conventional and harm/welfare-based moral decision-making. *Cognitive, Affective, & Behavioral Neuroscience, 17*(6), 1114–1128.

Wills, J., et al. (2018). Dissociable contributions of the prefrontal cortex in group-based cooperation. *Social Cognitive and Affective Neuroscience, 13*(4), 349–356.

Workman, C. I., Yoder, K. J., & Decety, J. (2020). The dark side of morality—neural mechanisms underpinning moral convictions and support for violence. *AJOB Neuroscience, 11*(4), 269–284.

Wright, J., Cullum, J., & Schwab, N. (2008). The cognitive and affective dimensions of moral conviction: Implications for attitudinal and behavioral measures of interpersonal tolerance. *Personality and Social Psychology Bulletin, 34*(11), 1461–1476.

Wright, J. C., & Pölzler, T. (2022). Should morality be abolished? An empirical challenge to the argument from intolerance. *Philosophical Psychology, 35*(3), 350–385.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE, 8*(7), e68910.

Yoder, K. J., & Decety, J. (2018). The neuroscience of morality and social decision-making. *Psychology, Crime & Law, 24*(3), 279–295.

Yoder, K. J., & Decety, J. (2022). Moral conviction and metacognitive ability shape multiple stages of information processing during social decision-making. *Cortex, 151*, 162–175.

Zaal, M. P., Saab, R., O'Brien, K., Jeffries, C., Barreto, M., & Van Laar, C. (2017). You're either with us or against us! Moral conviction determines how the politicized distinguish friend from foe. *Group Processes & Intergroup Relations, 20*(4), 519–539.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging, 20*, 45–57.

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2020). The partisan mind: Is extreme political partisanship related to cognitive inflexibility? *Journal of Experimental Psychology: General, 149*(3), 407–418.