Neuropsychologia 127 (2019) 195–203

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

When morality opposes the law: An fMRI investigation into punishment judgments for crimes with good intentions

Qun Yang[a,*], Robin Shao[b], Qian Zhang[a], Chun Li[c], Yu Li[a], Haijiang Li[d,*], Tatia Lee[b]

[a] Institutes of Psychological Sciences, Hangzhou Normal University, Hangzhou 311121, China
[b] The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China
[c] Intel Mobile Communications Technology Ltd., Hangzhou 310053, China
[d] Department of Psychology, Shanghai Normal University, Shanghai 200234, China

ARTICLE INFO

Keywords:
Legal judgments
Moral judgments
Intentions
Temporoparietal junction
Prefrontal cortex

ABSTRACT

In judicial practice, morally right but legally wrong instances usually pose significant challenges for legal decision makers. To examine the cognitive and neural foundations of legal judgments in criminal cases involving apparent moral conflicts, we scanned 30 female participants during punishment judgments for crimes committed with good intentions. The behavioral results confirmed that moral acceptability was significantly correlated with the punishment ratings only in the good-intentioned crimes. The fMRI data mainly revealed that the right temporoparietal junction (rTPJ) plays special roles in processing criminal offenders' state of mind and that the right dorsal lateral prefrontal cortex (rDLPFC) plays roles in resolving moral conflicts involved in legal judgments. Specifically, we found that compared to the bad-intentioned scenarios, the good-intentioned scenarios evoked greater activities during the postreading stage in the brain area of the rTPJ and that a signal increase in the rTPJ was associated with more lenient penalty judgments in the good-intentioned scenarios. Furthermore, reading crime scenarios with good intentions elicited stronger activation in the rdlPFC, which showed enhanced functional connectivity with the medial prefrontal cortex (mPFC). Overall, our study sheds some light on the neurocognitive underpinnings of legal judgments in special criminal cases and enhances our understanding of the relationship between legal and moral judgments.

1. Introduction

In modern society, the capacity of humankind to morally condemn and legally punish social norm violators has been particularly effective at enabling high-level long-term cooperation. Both normative judgments are related to thinking about right and wrong (Goodenough, 2001; Goodenough and Prehn, 2004); however, they serve to prevent the degeneration of social norms in different manners. Moral judgments refer to people's evaluation of an act or a person as good or bad in light of certain moral values (Haidt, 2001), which may vary among people or cultures. A moral decision usually involves condemnation of a wrongdoer without inflicting any actual punishment upon him/her. However, the purpose of legal judgments extends beyond determining blameworthiness. A distinctive feature of legal judgments is the enforcement of punishment for actions against the criminal law established by the national legislature (Buckholtz et al., 2008). Therefore, a legal decision determines how criminals should be punished based on the law.

In criminal punishment judgments, a legal decision maker (usually a judge or juror) first decides whether a suspect is guilty of a crime by conviction and then imposes a specific punishment by choosing a sentence that fits the crime. An appropriate legal decision requires the combined evaluation of the intent of the offender while committing the wrongful act and the consequential harm inflicted on the victim (Alter et al., 2007). Most research efforts aiming to unravel the mechanisms of legal judgments have focused on decoding the interactive process during which subjects weigh intent and harm information.

It has been well established that evaluations of intent or mental states in legal judgment primarily recruit a core brain network involved in theory of mind (ToM), such as the temporoparietal junction (TPJ), medial prefrontal cortex (mPFC), and posterior cingulate cortex (PCC), while harm assessment typically engages neural circuitry associated with socioemotional processing, such as the salience network (SN), which is anchored in the amygdala, insula and dorsal anterior cingulate cortex (dACC) (Buckholtz et al., 2008; Ginther et al., 2016; Krueger and Hoffman, 2016; Krueger et al., 2013; Treadway et al., 2014; Yamada et al., 2012). The integration of intent and harm information involves

* Corresponding authors.
E-mail addresses: qunyang@hznu.edu.cn (Q. Yang), haijiangli@shnu.edu.cn (H. Li).

https://doi.org/10.1016/j.neuropsychologia.2019.01.020
Received 19 August 2018; Received in revised form 30 October 2018; Accepted 20 January 2019
Available online 22 February 2019

the default mode network (DMN), which is anchored in the medial prefrontal cortex (mPFC) (Buckholtz and Marois, 2012; Ginther et al., 2016; Krueger and Hoffman, 2016). Finally, a decision about the appropriate punishment based on the integrated information depends on the central executive network (CEN) (Krueger and Hoffman, 2016). The dorsal lateral prefrontal cortex (dlPFC), which is a critical node in the CEN, has been commonly suggested to be involved in executive functions, such as inhibitory control, working memory, planning, and abstract thinking, which are necessary for reasoning, problem solving, and decision making (e.g., Alvarez and Emory, 2006; Diamond, 2011; Greene et al., 2004; Knoch et al., 2006). However, the specific role of the dlPFC in legal decision making remains controversial. Some studies have revealed that the dlPFC is engaged in legal judgments to inhibit prepotent responses (Schleim et al., 2011), whereas other studies have associated this region with the selection of a specific punishment response after integrating all necessary information (Bellucci et al., 2017; Buckholtz et al., 2008, 2015; Ginther et al., 2016; Treadway et al., 2014).

The neural circuits involved in legal judgments partially overlap with those involved in moral judgments. Thus far, no brain areas have been found to be uniquely associated with morality. Instead, it has been consistently shown that moral judgments are fueled by domain-general processes, such as mental states comprehension, emotional arousal, attention, and cognitive control (Young and Dungan, 2012). Determining right from wrong morally generally requires the dual process system of cognition and emotion, which involves brain regions, such as the mPFC, TPJ, dlPFC, ventromedial prefrontal cortex (vmPFC), amygdala, and insula (Bzdok et al., 2012; Yoder and Decety, 2014).

Similar neural processes should be expected during legal and moral judgments because they conceptually overlap (Goodenough, 2001; Goodenough and Prehn, 2004). Moral judgments are inherently embedded in the legal judgment process such that the severity of punishment in legal circumstances has been found to heavily rely on the moral wrongfulness of the offense (Alter et al., 2007). In particular, lay people tended to sentence a criminal based on the “just-deserts” motives (Carlsmith et al., 2002). The moral values of different cultures are often reflected in their legal systems, particularly in criminal justice (Alter et al., 2007). Therefore, in most cases, the decision of the judicial ruling should be in accordance with our moral intuition about the offenses. However, morally right but legally wrong instances are not uncommon in real lives. In 2017, in a case called "*Yuhuan v. Shandong*" in China, a 22-year- old young man named "Yuhuan" was sentenced to life imprisonment for his crime of intentional injury, leading to one death and two other seriously injured victims. The judicial ruling results elicited strong public opinion as people learned from the media that Yuhuan committed the crime to protect his mother from being severely harassed by 11 debt collectors for hours. The traditional Chinese cultural moral value of filial piety earned Yuhuan great sympathy from the society. Legal decision makers are sometimes indeed greatly challenged by the conflicts between moral values and criminal justice.

The intent of an act legally varies from purposeful, knowing, and reckless to negligent and blameless (Shen et al., 2011). A crime committed purposefully or knowingly with good intentions might be morally blameless but would usually be subjected to criminal justice. In modern criminal law in China, the malignancy level of criminal intent may not affect the conviction but could be considered mitigating circumstances in the sentencing in cases in which judges are granted discretionary power. A few imaging studies have already addressed the issue of the neural underpinnings underlying legal judgments in the presence of mitigating circumstances (e.g., Buckholtz et al., 2008; Yamada et al., 2012); however, no study has directly examined moral conflicts in the legal judgments process. In the study conducted by Buckholtz et al. (2008), the mental states of the criminal in the diminished-responsibility scenarios varied from blameless, negligent, and reckless to purposeful under the threat of danger. The criminal acts may be less blameworthy than those in the full-responsibility circumstances, but they are by no means morally justified. However, Yamada et al. (2012) simply manipulated the mitigating circumstances by introducing some pitiful experiences of an offender, such as suffering from a fatal disease.

In the present study, we investigated the cognitive and neural basis of penalty judgments for purposeful criminal acts with good intentions in an attempt to determine how people resolve apparent moral conflicts in their legal judgments to obtain a better understanding of the relationship between the two types of normative judgments. To address this issue, we recruited 30 participants who were asked to assign an appropriate punishment to a hypothetical perpetrator who caused different levels of harm to victims out of either good or bad intentions. The types of offenses varied from property crimes to violent murders under both the good or bad conditions but were matched between the two conditions. We focused on the sentencing stage while maintaining the conviction consistent across the two conditions with or without mitigating circumstances.

We hypothesized that lay people might mainly rely on the "just-deserts" motives in their legal judgments and, thus, that their moral acceptability ratings of acts should determine the magnitudes of the assigned punishment. As indicated by previous research, adult participants tend to place much more weight on intentions than outcomes in making moral or legal judgments, particularly for harm-related violations (Alter et al., 2007; Young and Saxe, 2011). Additionally, the inflicted harm was matched across the good- and bad-intentioned scenarios. Therefore, we expected that the main difference in the cognitive processes supporting the punishment judgments between the two types of scenarios should be mental states evaluation. The TPJ, which is typically associated with mentalization in legal judgments, was hypothesized to be more engaged in the good-intentioned scenarios than in the bad-intentioned scenarios. Normally, crimes are expected to be committed maliciously. Therefore, we believe that the participants may find it easier to process the intent information in the bad-intentioned crimes but that they may focus more on mentalization in punishment decisions against their moral intuitions for crimes that caused actual harm but were committed for good purposes. Thus, the moral conflicts under the good condition should lead the participants to rely more on cognitive control to regulate their decision responses, which should be evidenced by increased activities in some nodes in the CEN, such as the dlPFC, under the good condition compared with those under the bad condition. However, if the dlPFC mainly plays a punishment selective role in legal judgments as indicated by previous research (e.g., Buckholtz et al., 2015; Ginther et al., 2016), we expected this region to be more activated in the bad-intentioned scenarios than the good-intentioned scenarios because the protagonist in the former scenario should be punished more. In addition, given the *a priori* hypothesis regarding the role of the rdlPFC in legal judgments (for a review, see Krueger and Hoffman, 2016), this region was selected as a seed region of interest for the functional connectivity analyses. We intended to identify brain areas in which the neural connectivity with the dlPFC could be modulated by the two types of legal scenarios. Specifically, we hypothesized that greater connectivity would be observed between the dlPFC and brain areas associated with mental states processing in the good-bad contrast during legal judgments. If this functional connectivity serves the role of effective cognitive control over resolving moral conflicts involved in legal judgment, we expected to observe enhanced neural coupling with harsher penalty judgments in the good-intentioned scenarios.

2. Materials and methods

2.1. Participants

As we focused on how laypersons made legal judgments involving moral conflicts, thirty healthy nonlaw college students were recruited for the experiment (aged 21.2 ± 2.28 years). Only females were

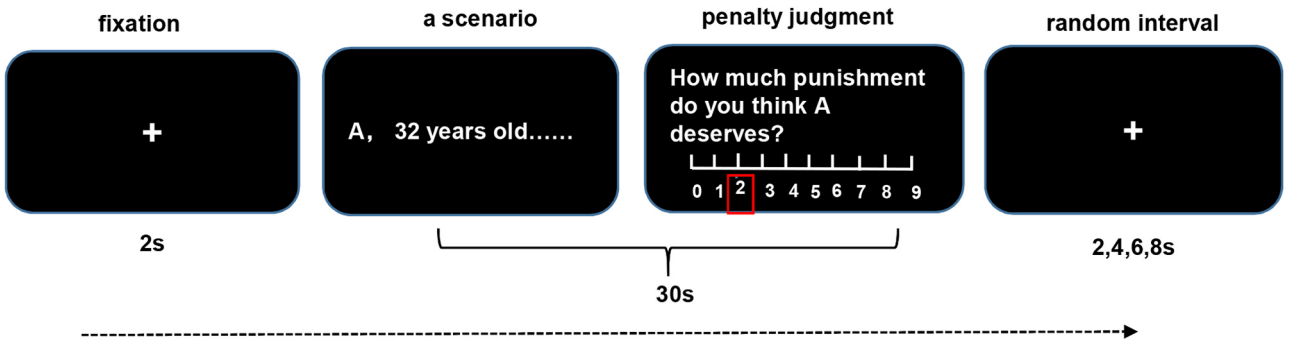

Fig. 1. An illustration of the sequence of a trial.

included in the sample due to a shortage of voluntary male participants. All participants had normal or corrected to normal vision. No physical or mental illnesses were reported by any subject. The ethics committee of the Center for Cognitive and Brain Disorders of Hangzhou Normal University approved the experimental protocol. Each participant signed informed consent before the experiment and was paid 120 RMB per hour (roughly equal to 18.5 US dollars). The participants were allowed to quit the experiment at any time due to discomfort with the fMRI procedure.

2.2. Stimuli

In total, 100 short written scenarios of criminal acts were adapted from legal cases reported in the media or stories from other sources. Each scenario depicted a hypothetical protagonist, “A”, who intentionally (either with good or bad intention) conducted a criminal act and inflicted harm on the victim. The criminal intent was manipulated in such a way that the criminal acts were well-intended in half of the scenarios and bad-intended in the other half. Therefore, the 100 scenarios consisted of 50 pairs, each of which differed only in the criminal intent (good vs. bad). Each pair of scenarios was identical in the identities of the criminals and the victims and the features of the criminal methods and the severity of harm. According to the Chinese criminal law, the malignancy level of criminal intent should be considered a discretionary circumstance of sentencing. Thus, each pair of scenarios was designed to have the same conviction, but the sentencing may differ.

Another 20 nonlaw subjects aged between 18 and 25 years (10 females) participated in a pilot survey to evaluate the materials. The participants were asked to make decisions regarding the conviction of the criminals and make judgments about the clarity of the intent. Additionally, the participants rated each scenario using scales from 0 to 9 concerning how much punishment “A” deserved, how malignant the criminal intent was and how likely it was for the story to occur in real lives. Meanwhile, we invited two legal professionals from law schools and a professional judge from the court to provide suggestions regarding the ecological validity of and possible convictions and sentences for each scenario.

Fourteen pairs of scenarios were excluded either because the differences in the malignancy level of the criminal intent were insignificant or the convictions were likely to differ between the good and bad conditions (see supplementary methods and results for more details regarding the stimuli screening process). Eventually, 72 scenarios (36 good-intentioned and 36 bad-intentioned scenarios) were selected as the final set of experiment stimuli (see supplementary materials). The three legal professionals agreed on the ecological validity of the final stimulus set. Each of the 36 pairs of scenarios was equated in terms of the length of words. The average word counts for all scenarios were 130 Chinese characters (132 was the maximum, and 129 was the minimum).

Because the participants may have guessed the purpose of the experiment after being presented with good-bad pair scenarios under the same premises, two groups of the 36 pairs of scenarios were constructed. Thus, we assigned a number to each pair of scenarios. Group A consisted of 18 scenarios with good intentions from the first half of the 36 pairs and another 18 scenarios with bad intentions from the remaining half; Group B consisted of 18 good-intentioned scenarios from the latter half and 18 bad-intentioned scenarios from the first half. Therefore, the basic stories of the scenarios presented in the good condition to Group A were the same as those in the bad condition presented to Group B, while the basic stories of the scenarios in the bad condition presented to Group A were identical to those in the good condition presented to Group A. The presentation of the two groups of stimuli were counterbalanced across participants.

2.3. Task and procedures

The participants were instructed to imagine that they were judges and determine the appropriate punishment for the criminal (“A”) described in each scenario on a scale ranging from 0 to 9, with "0" indicating no punishment and 9 indicating the most serious penalty. The participants were randomly assigned to either of the two experimental stimuli groups. Each participant received written instructions and four practice trials before entering the scanner. The participants were presented with the instruction again after entering the scanner, followed by another two practice trials to familiarize the participants with the experimental procedure before the formal experiment. Six good-intentioned scenarios and six bad-intentioned scenarios were selected for each of the three fMRI runs. The good-intentioned and bad-intentioned trial presentations were randomized within each run. As shown in Fig. 1, each trial started with a fixation for 2 s, followed by a scenario. The participants were asked to use their left index finger to press a key after they have finished reading the scenario to proceed to the next screen for the penalty judgment question. Immediately after the key was pressed, the question “How much punishment do you think A deserves?” was presented with a 10-point scale. The participants were required to respond as soon as possible once they had reached a decision by moving a red square designed to randomly appear on any of the numbers on the scale to the desired number. The participants responded with either their right index finger (to move left) or their right middle finger (to move right). The scenario and the question were presented for a total of 30 s. The response screen did not disappear until the set time elapsed. The whole trial ended with a jittered interstimulus interval, and the duration of this interval was randomized among 2, 4, 6 or 8 s.

After the scanning session, the participants were asked to rate the same scenarios by answering the following questions: “How much

anger, disgust or sympathy did you feel for the criminal?", "how much conflict did you feel or how difficult was it to make the decision?", and "how morally acceptable was it for A to conduct the act?".

2.4. fMRI data acquisition

The imaging was performed on a 3.0 T GE Discovery MR-750 Scanner (GE Medical Systems, Milwaukee, WI) with an 8-channel receive-only head coil located at the Center for Cognitive and Brain disorders, Hangzhou Normal University. High-resolution whole-brain anatomical images (T1-Weighted inversion prepared 3D spoiled gradient echo sequence: TR/TE = 8.1/3.1 ms; number of sagittal slices = 176; slice thickness = 1 mm; flip angle = 8°; FOV = 256 × 256 mm^2, and inversion time = 450 ms) were obtained prior to the functional images (T2 * weighted echo-planar imaging sequence: TR/TE = 2000/30 ms; flip angle = 90°; number of axial slices = 43; slice thickness = 3.2 mm; matrix size = 64 × 64; voxel size = 3.4 × 3.4 × 3.2 m^3; and FOV = 220 × 220 mm^2). The first five images of each run were discarded for equilibration. The participants responded manually in the scanner with two-button keypads (one for each hand; Sinorad, Shenzhen, China). The participants performed three fMRI runs. The duration of each run varied from 7.6 min to 7.7 min.

2.5. Behavioral data analyses

First, the rating scores of the criminal punishment, moral acceptability, difficulties in punishment and emotional feelings toward the offenders were compared between the good-intentioned and bad-intentioned crimes using paired-sample *t*-tests. Second, correlational analyses were performed to reveal how moral acceptability and emotional feelings were related to the criminal punishment judgments in the good-intentioned and bad-intentioned scenarios separately. The Bonferroni correction method was applied for the multiple-testing analyses.

2.6. Image preprocessing

The functional data were preprocessed using dpabi software (version 2.3) (Yan et al., 2016). The preprocessing of the functional data included slice timing correction, three-dimensional motion correction by realigning all volumes to the first volume, normalization to the standardized MNI space (Montreal Neurology Institute; resampling voxel size was 3 × 3 × 3 mm^3) and spatial smoothing using a Gaussian kernel with a FWHM 6.0 mm. The functional images were coregistered to each individual's T1 weighted anatomical images. The data were temporally filtered using a high-pass filter with a width of 128 s.

2.7. fMRI data analyses

The preprocessed MRI images were further analyzed using a general linear model with SPM8 (Wellcome Department of Imaging Neuroscience, London) in MATLAB (MathWorks). First, we separately modeled the "reading phase" and "postreading phase" occurring while the participants read the circumstances and approached the final decision, respectively. These two processes might be associated with different brain activities examined in previous studies (Bellucci et al., 2017; Ginther et al., 2016; Schleim et al., 2011). Although the decision-making process might have started while the participants were reading the circumstances, it is worthy to identify brain activities during the postreading phase as this stage is the nearest to the final decision after the participants had all the available information. The duration of the reading phase was defined as the time between the scenario onset and scenario offset, while the postreading phase was modeled with a duration of 5 s beginning with the disappearance of the scenarios (due to technical reasons, we failed to record the time when the participants started their first press, and thus, the actual reaction time of the decision responses is unknown. The time window between the time the participants finished reading the circumstances and the time when the participants finished moving the square that appeared on the desired number was 6.6 s. Considering that the participants needed time to move the square, we selected 5 s for the postreading data analyses. See supplementary results for more details). The 5-second postreading duration was close to the 6-second time window used in the decision phase analyses in the study conducted by Bellucci et al. (2017). The good and bad circumstances were modeled as separate conditions using a general linear model (GLM). The six head movement correction parameters were included as regressors of no interest. Each regressor was convolved with a canonical hemodynamic response function (HRF). For each participant, we first performed the first-level contrasts of "good > bad" and "bad > good" using one-sample *t*-tests. Then, the resulting statistical maps of each contrast for the whole group (n = 30) were submitted to second-level random-effect analyses.

As the crime scenarios in either the good or bad condition differed in the severity of the criminal behaviors, ranging from property crimes to murder, we carried out additional analyses to identify brain activities associated with the severity of the crimes (the results are reported in the supplementary information). We adopted the median split method previously used by Buckholtz et al. (2008) to divide each set of scenarios from the good and bad conditions into high (above the median) and low (below the median) groups based on the participants' actual punishment scores. Thus, a separate GLM model was performed by including 4 regressors, i.e., the two types of scenarios (high and low) for each of the two intentions (good and bad).

Based on previous findings regarding commonly observed neural circuits involved in legal judgments, small volume corrections (within an 8-mm radius sphere) were conducted in the following *a priori* regions of interest: rTPJ and rdlPFC (Bellucci et al., 2017; Buckholtz et al., 2008, 2015; Ginther et al., 2016; Krueger and Hoffman, 2016; Yamada et al., 2012). The coordinates of the rTPJ (x = 57, y = -56, and z = 24) were obtained from Young et al. (2007), and the coordinates of the rdlPFC (x = 40, y = 21, and z = 27) were derived from Ott et al. (2011). For each ROI, the mean activities within an 8-mm-radius sphere were extracted. For multiple comparisons in the above whole-brain and ROI analyses in brain regions reaching cluster-level significance at $p < 0.05$, the FWE corrected values (with a cluster-forming threshold of $p = 0.001$ and a cluster size ≥ 10 voxels) are reported.

Finally, generalized psychophysiological interactive analyses (gPPI) were carried out with the gPPI toolbox (McLaren et al., 2012) to examine the functional connectivity during legal judgments. The PPI seed region was created by placing an 8-mm-radius sphere around the local maxima of the rdlPFC identified in the good-bad contrast while the participants read the legal scenarios. The threshold of the PPI results was set at $p < 0.05$ and FWE corrected (with a cluster-forming threshold of $p = 0.001$ and a cluster size ≥ 10 voxels). We extracted the mean functional connectivity value of the 8-mm sphere centered on the peak coordinate of the region that showed a significant effect in the PPI analysis. Then, we conducted exploratory analyses of the relationship between the brain functional connectivity and behavioral measures. Outlier data beyond 3 standard deviations from the mean value were excluded from all analyses.

3. Results

3.1. Behavioral data

The paired-sample *t*-tests showed that there were significant differences between the good-intentioned and bad-intentioned scenarios

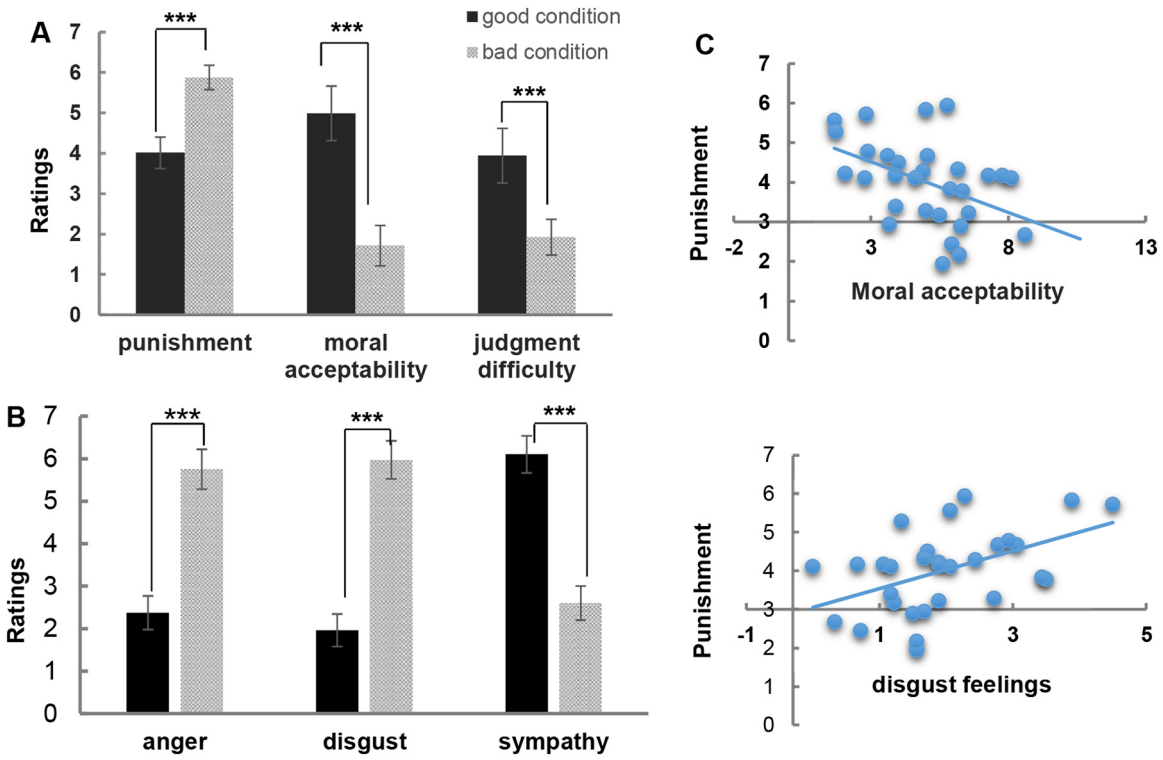

Fig. 2. (A) The average ratings of criminal punishment, moral acceptability and judgment difficulty. (B) The average ratings of the emotions anger, disgust and sympathy toward the criminals. (C) Correlation between moral acceptability and punishment scores in the good-intentioned condition. Correlation between the disgust emotion and punishment scores in the good-intentioned condition. "***"$p < 0.001$, error bars: ± 2 SE.

Table 1
Significant brain activation associated with the main effects in the whole brain analyses.

Brain regions	MNI coordinates			Max t-value	p (FWE-corr.)	Voxel size
	X	Y	Z			
Good > bad						
Reading phase						
Inferior frontal gyrus_R	45	33	−12	5.640	0.062	55
Middle frontal gyrus_R	39	21	33	4.670	0.053	58
Precuneus_L	−6	−66	24	4.580	0.009	92
Postreading phase	No significant results					
Bad > good						
Reading phase						
Anterior cingulate cortex_L	−15	6	39	4.850	0.028	70
Insula_R	48	−21	21	4.340	0.025	72
Postreading phase	No significant results					

in the ratings of the decision-related variables (criminal punishment, moral acceptability, and difficulties in punishment) and emotional feelings (disgust, anger and sympathy) toward the offenders. Crimes with good intentions were assigned less severe punishment and found morally more acceptable than those with bad intentions ($t = -11.640$, $p < 0.001$; $t = 14.137$, $p < 0.001$, Bonferroni correction for the three tests under the decision-related variables). The participants reported that making a judgment decision was more difficult for these crimes ($t = 8.032$, $p < 0.001$, Bonferroni correction for the three tests). Furthermore, compared to the bad-intentioned scenarios, the participants felt remarkably less anger and disgust and more sympathy toward the criminals in the good-intentioned scenarios ($t = -16.511$, $p < 0.001$; $t = -19.126$, $p < 0.001$; $t = 17.120$, $p < 0.001$, Bonferroni correction for the three tests under the emotional variables) (see Fig. 2).

The correlational analyses of the decision-related variables showed that moral acceptability was significantly associated with the criminal punishment scores in the good condition ($r = -0.483$, $p = 0.007$, Bonferroni correction of the pairwise correlations among the three decision-related variables). Regarding the three emotions, only disgust was significantly correlated with the penalty judgments in the good condition ($r = 0.508$, $p = 0.004$, Bonferroni correction of the pairwise correlations between the punishment ratings and the three emotions). We did not observe any association between moral acceptability and the penalty judgments or any significant correlation between the penalty judgments and any of the three emotion in the bad-intentioned condition.

3.2. fMRI data

3.2.1. Brain activation

A small volume correction was applied to the two ROIs in this study (rTPJ and rdlPFC). The results showed that the rdlPFC ($t = 4.67$, $k_E = 14$, $p_{(FWE\text{-}corr.)} = 0.005$) was more strongly activated while the participants read the good-intentioned scenarios compared to that while they read the bad-intentioned scenarios, while the rTPJ ($t = 4.88$, $k_E = 18$, $p_{(FWE\text{-}corr.)} = 0.004$) was more engaged under the good condition than under the bad condition during the postreading phase.

The whole brain analyses showed that the SPM of the "good > bad" contrast during the reading phase yielded stronger activation in the left precuneus, which extends to the PCC, and tended to have greater activation in the inferior frontal gyrus cortex and middle frontal gyrus. In the "bad > good" contrast, we found greater activation in the left ACC and right insula (see Table 1 and Fig. 3).

The whole brain analyses of the postreading phase revealed no

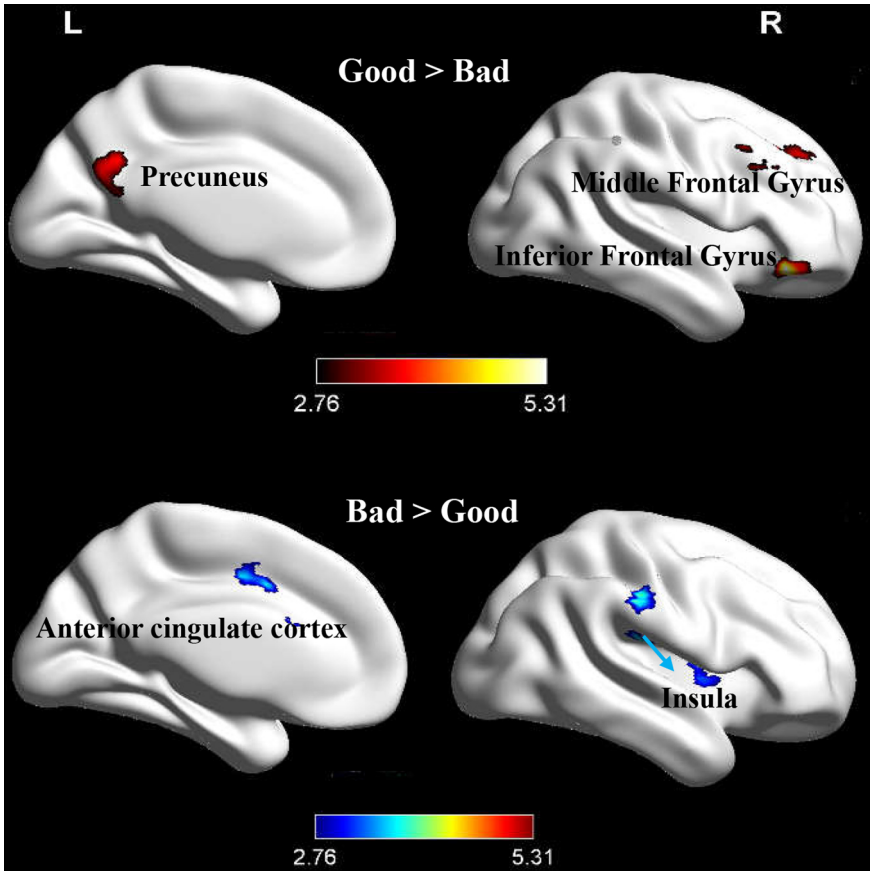

Fig. 3. Significant brain activation in the "good > bad" and "bad > good" contrasts in the whole brain analyses (a cluster-forming threshold of $p < 0.005$ was used for viewing).

significant results in either the "good > bad" contrast or the "bad > good" contrast.

3.2.2. Functional connectivity

Compared to the bad-intentioned scenarios, the PPI analyses seeded in the rdlPFC (x = 39, y = 21, and z = 33) revealed significant positive coupling with the mPFC (x = 21, y = 24, and z = 42; t = 5.750, $p_{(FWE\text{-}corr.)} = 0.001$, $k_E = 135$) and marginally significant connectivity with the rTPJ (x = 54, y = -51, and z = 24; t = 5.300, $p_{(FWE\text{-}corr.)} = 0.052$, $k_E = 55$) while the participants read the good-intentioned scenarios.

3.2.3. Brain regions and functional connectivity associated with behavioral ratings

We conducted further exploratory analyses to examine the brain-behavioral data relationship. Using the median punishment scores to divide the participants into the high and low punishment groups, the results showed that the participants who scored lower in the punishment ratings had significantly stronger activation in the area of the rTPJ in the good-intentioned scenarios ($M_{high} = 0.446$, $M_{low} = 1.697$; t = 2.753, p = 0.013, two-tailed test, Bonferroni correction of the two ROIs).

In addition, we found that the participants who scored above the median scores in the punishment ratings had marginally significantly greater rdlPFC-mPFC connectivity than those who scored below the median scores in the good-intentioned crime scenarios ($M_{high} = 0.130$, $M_{low} = 0.059$; t = 2.367, p = 0.025, two-tailed test, Bonferroni correction of the rdlPFC-mPFC and the rdlPFC-rTPJ connectivity) (see Fig. 4).

4. Discussion

In this study, we examined the cognitive and neural processes of penalty judgments related to sentencing and aimed to reveal how laypersons resolve moral conflicts in determining punishment magnitudes for crimes committed with good intentions.

4.1. Contributions of the rTPJ and rdlPFC to legal judgments

Our findings showed that moral acceptability was significantly correlated with sentencing in legal circumstances in which moral intuitions conflict with legal norms. Higher moral acceptability was associated with a lower penalty magnitude. In particular, the fMRI data revealed that the rTPJ plays an important role in translating the perpetrators' mental states into the punishment ratings.

It has been well documented that the TPJ is involved in theory of mind (ToM) and mentalizing (for meta-analyses, see Molenberghs et al., 2016; Schurz et al., 2014). In particular, the neural activity in the right TPJ has been suggested to be associated with analyzing mental states information throughout the process of moral judgments from the initial encoding to post hoc justification (Jeurissen et al., 2014; Krall et al., 2015; Sellaro et al., 2015; Koster-Hale et al., 2013; Young and Dungan, 2012). The increased BOLD signal in the rTPJ led participants to assign more weights to the innocent intentions than the resulting harm in moral scenarios involving accidental harm (Young and Saxe, 2009). Disruption of the rTPJ shifted participants from relying on intentionality to relying on accidental outcomes in moral judgments (Young et al., 2010).

In the current study, we found that compared to the bad-intentioned scenarios, the rTPJ was more active in the good-intentioned scenarios during the time the participants approached their final decision. We

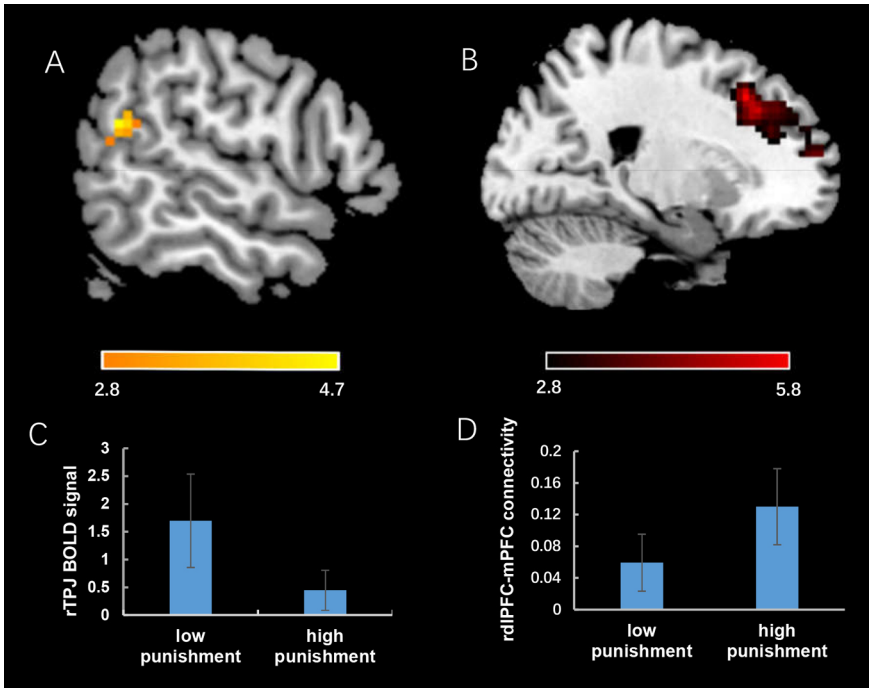

Fig. 4. (A) Clusters in the rTPJ (peak: x = 57, y = -54, and z = 21) in the "good > bad" contrast during the post-reading phase. (B) Significant functional connectivity with the mPFC (x = 21, y = 24, and z = 42) in PPI seeded in the rdlPFC in the "good > bad" contrast. (C) rTPJ BOLD signal in the high and low punishment groups in the good-intentioned condition using a median grouping method, error bars: ± 2 SE. (D) Functional connectivity between the rdlPFC and mPFC in the high and low punishment groups in the good-intentioned condition using the median split method, error bars: ± 2 SE (a cluster-forming threshold of $p < 0.005$ was used in the brain images for viewing).

assumed that the participants might weigh their evaluations of the offenders' mental states more heavily during their decision for crimes committed with good intentions compared to those committed maliciously. Consistent with our assumption, we did not find any significant neural differences between the high and low punishment scenarios in the good condition, whereas the high punishment circumstances evoked greater neural activity in the inferior parietal lobule than the low punishment scenarios in the bad condition (see supplementary results). The exploratory brain-behavioral data analyses showed that the neural responses in the rTPJ were related to moral acceptability and punishment magnitudes in an opposing manner. In particular, the greater the differences in moral acceptability between the good- and bad-intentioned scenarios, the larger the observed rTPJ signal changes (see supplementary results); the signal increase in the rTPJ was associated with more lenient penalty judgments in legal circumstances. These findings were tentative but consistent with our assumption.

The region of the prefrontal cortex has been commonly associated with the role of cognitive control (Dixon, 2015; Duverne and Koechlin, 2017; Miller and Cohen, 2001). Numerous studies have suggested that the dlPFC is engaged in inhibiting intuitive responses and resolving possible conflicts in moral judgments and other norm-related behaviors. For example, enhanced dlPFC activity has been associated with greater difficulty in moral judgments (Greene et al., 2004). The involvement of this region was evidenced by participants' increase in norm compliance under punishment conditions to inhibit prepotent selfish responses (Spitzer et al., 2007). Disrupting the area of the rdlPFC weakened the individuals' ability to override their selfish impulses and, thus, reduced second-party altruistic punishment implementation (Knoch et al., 2006). However, the cognitive control hypothesis has gained much less evidence in the third-party punishment field than the integration-and-selection hypothesis, which claims that the rdlPFC should be more activated under conditions in which the participants decide to assign harsher punishment during legal decision making (Buckholtz et al., 2015; Ginther et al., 2016). One of the strong pieces of evidence was that disrupting the region of the rdlPFC was found to decrease the punishment magnitude in full-responsibility, but not diminished-responsibility, crimes (Buckholtz and Marois, 2012).

The proposed integration-and-selection hypothesis has mostly been tested in research exclusively adopting bad-intentioned crimes involving little conflict during the participants' decision making. In the current study, we were particularly interested in determining how people make legal decisions in scenarios that involve apparent moral conflicts. Our results showed that the rdlPFC was more engaged in the good-intentioned crimes relative to that in the bad-intentioned crimes, which is difficult to reconcile with the integration-and-selection hypothesis. Instead, this finding could be better interpreted in terms of the cognitive control hypothesis. Usually, moral intuition could facilitate and motivate punishment decisions in situations in which the moral norms are consistent with the legal codes. However, a moral conflict in the legal context makes it more difficult for participants to determine punishments for a crime and, thus, may require them to recruit brain areas, such as the rdlPFC, to inhibit their moral intuition, which acts against the legal norms. Apparently, even people without much professional legal knowledge may not purely rely on moral intuition in making criminal punishment judgments.

Moreover, we observed increased functional connectivity between the rdlPFC and mPFC in the "good > bad" contrast. We proposed that the enhanced rdlPFC-mPFC connectivity during the good-intentioned crimes demonstrates increased cognitive control over the integration of intention and harm information in legal circumstances in which moral intuitions conflict with the legal norms. In addition, further exploratory analyses demonstrated that the increased neural coupling between the rdlPFC and mPFC tended to be linked to more severe punishment and lower ratings of moral acceptability (see supplementary results) in the good-intentioned scenarios. These tendencies were relatively consistent with the cognitive control role of the rdlPFC in criminal punishment judgments.

4.2. Role of aversive emotions in legal judgments

Growing scientific evidence has revealed the role of emotions in legal judgments. Aversive emotions, such as anger or disgust, toward criminals have been known to both directly and indirectly motivate legal judgments in legal contexts (Bright and Goodman-Delahunty, 2006; Buckholtz and Marois, 2012; Salerno and Bottoms, 2009; Treadway et al., 2014). Emotional states may serve as a source of information guiding punishment judgments (Buckholtz and Marois, 2012;

Capestany and Harris, 2014). The brain regions involved in threat detection and affective processing, such as the amygdala, insula, and precuneus, have been found to be actively engaged in punishment judgments (Buckholtz et al., 2008; Ginther et al., 2016; Treadway et al., 2014; Yamada et al., 2012). However, some prosocial emotions, such as sympathy for criminals, may reduce moral culpability and, thus, motivate the mitigation of punishment. Brain areas, such as the precuneus, left TPJ and DMPFC, have been found to be associated with sympathy processing in legal judgments in the presence of mitigating circumstances (Yamada et al., 2012).

Our findings provide some support for the role of negative emotions in legal judgments. The participants reported less anger and disgust toward the criminals in the good-intentioned scenarios, whereas more anger and disgust toward the criminals were reported in the bad-intentioned scenarios. The behavioral results showed that the disgust emotion was correlated with the punishment ratings. Sympathy for the offenders was not related to the penalty judgments. The fMRI findings demonstrated that the "good > bad" contrast during the reading phase yielded greater activities in the left precuneus. The precuneus has been commonly suggested to be involved in self-referential processes and affective responses during normative judgments (Buckholtz et al., 2008; Krueger and Hoffman, 2016; Schleim et al., 2011; Yamada et al., 2012; Yoder and Decety, 2014). Capestany and Harris (2014) demonstrated an association between left precuneus activity and disgust emotion during legal punishment decisions. We assume that the left precuneus may be related to aversive emotion processing in legal judgments involving mitigating circumstances. The reverse contrast (bad > good) during the reading phase revealed greater activities in the left dorsal anterior cingulate cortex and right insula in the bad-intentioned condition. The dACC and insula are core nodes in the salience network (SN), which detects aversive emotional experiences and guides legal judgment actions. We believe that scenarios with bad intentions and outcomes generate strong aversive emotions and, thus, become motivationally more relevant for participants than those with good intentions and bad outcomes. Altogether, these findings suggest that inner aversive emotional states are integrated with reading legal circumstances and may act as important heuristic devices that guide people's legal judgments (Buckholtz and Marois, 2012).

4.3. Relationship between moral and legal judgments

There is significant conceptual overlap between moral and legal judgments, both of which involve normative judgments. Thus far, no unique neural substrates have been found to be dedicated to either of the two normative judgments. In contrast, the more domain-general contributions of complex cognitive and social-affective systems have been discussed in the neuroscientific study of morality and law (Buckholtz et al., 2008; Young and Dungan, 2012). As indicated by previous studies and the present research, despite the crucial role and predictive value of morality in legal judgments, there are essential differences between these two normative judgments. Moral judgments primarily rely on inferences about good and bad intentions in evaluations of the moral blameworthiness of an agent. However, a legal decision distinguishes among purposeful, knowing, reckless and negligent actions based on legal norms (Shen et al., 2011). Even a layperson recognizes that a purposeful violation of the law should be subjected to criminal punishment regardless of whether the intent is good or bad. Making penalty judgments for crimes with good intentions requires the inhibition of prepotent responses. According to a recent theory of moral judgments, intuitively perceived harm ultimately drives moral condemnation of wrongful acts (Schein and Gray, 2017). Therefore, moral judgments may rely more on personal values and reflect greater individual differences than legal judgments that are more rule-based (Schleim et al., 2011). Yoder and Decety (2014) demonstrated that dispositional differences in justice sensitivity significantly predicted behavioral and neural responses of moral judgments. However, no individual differences in justice sensitivity or moral emotions were found to be related to legal judgments in this study (see supplementary results). In summary, we believe that the internal motivation for punishment and external constraints under legal rules are intrinsic features that distinguish legal from moral judgments.

4.4. Limitations and conclusions

Several limitations should be addressed in this study. First, similar to most research in this field, we examined how laypersons make legal decisions without any legal training. Previous findings have indicated that professional judges are as prone to cognitive biases and emotional experiences as jurors (Vidmar, 2011). However, legal training may make a difference when judges are given more detailed information and more time to deliberate. Compared to other academics, lawyers with normative expertise have been found to show less emotional involvement behaviorally and demonstrate reverse neural response patterns in the dACC while making legal judgments (Schleim et al., 2011). However, whether legal professionals resolve moral conflicts differently from nonlegal professionals at the behavioral and neural level remains unknown. Second, the recruitment of female-only college students in this study restrains the generalization of the findings to the whole population. Substantial evidence suggests that women differ from men in affective processing, which is a crucial component of normative judgments (Brody and Hall, 2000; Fischer et al., 2004). Significant gender differences in behavioral and neural responses have been observed in moral judgments research (Belacchi and Farina, 2018; Friesdorf et al., 2015; Fumagalli et al., 2010; Harenski et al., 2008). However, the possible gender effects on legal decision making have not been thoroughly discussed. Whether the neural basis of legal decision making in women is distinct from that in men remains to be explored in the future. In addition, most criminals in our stimuli materials were men. The disproportionate gender ratio in the scenarios may confound the results as women might perceive male criminals differently than female criminals. Additionally, the results should be further tested in community samples in future research. Third, we focused on the sentencing stage during which judges or jurors are granted with discretion. In real cases, the conviction, which occurs before sentencing, should rely more on the explicit application of legal rules. Whether and how moral conflicts differentially affect guilt and punishment judgments also requires future investigation.

Nevertheless, our study sheds some light on the cognitive and neural processes underlying legal judgments when apparent moral conflicts are involved. Moral acceptability of offenses with good intentions was significantly correlated with the severity of the punishment for the good-intentioned crimes. The brain area of the rTPJ plays a special role in translating mental states evaluations into punishment responses in making legal decisions. The higher recruitment of the rdlPFC and the enhanced connectivity between the rdlPFC and mPFC in the good-intentioned crimes scenarios compared with those in the bad-intentional scenarios supported the cognitive control hypothesis of the dlPFC region in criminal punishment judgments. Finally, the aversive experiences, rather than the sympathy emotion, were found to be correlated with the behavioral and neural responses to the criminal punishment judgments. Neural circuits, such as the precuneus, dACC and insula, were associated with affective processing in legal judgments.

Acknowledgment

This work was funded by the National Natural Science Foundation of China (31500907,31440046) and the Humanity and Social Science Youth Foundation of the Ministry of Education of the People's Republic of China (17YJC190011). We thank Prof. Wei xiaojun and Dr. He xuefeng and Judge Zhang Qinghua for their advice regarding the experimental materials. The authors do not have any conflicts of interest to declare.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.neuropsychologia.2019.01.020.

References

Alter, A.L., Kernochan, J., Darley, J.M., 2007. Transgression wrongfulness outweighs its harmfulness as a determinant of sentence severity. Law Hum. Behav. 31 (4), 319–335. https://doi.org/10.1007/s10979-006-9060-x.

Alvarez, J.A., Emory, E., 2006. Executive function and the frontal lobes: a meta-analytic review. Neuropsychol. Rev. 16 (1), 17–42. https://doi.org/10.1007/s11065-006-9002-x.

Belacchi, C., Farina, E., 2018. Utilitarian and emotion-related components of moral judgement: gender and age effects and the relationship with prosocial and hostile roles in bullying. Eur. J. Dev. Psychol. 15 (4), 438–451. https://doi.org/10.1080/17405629.2017.1301254.

Bellucci, G., Chernyak, S., Hoffman, M., Deshpande, G., Dal Monte, O., Knutson, K.M., Krueger, F., 2017. Effective connectivity of brain regions underlying third-party punishment: functional MRI and Granger causality evidence. Soc. Neurosci. 12 (2), 124–134. https://doi.org/10.1080/17470919.2016.1153518.

Bright, D.A., Goodman-Delahunty, J., 2006. Gruesome evidence and emotion: anger, blame, and jury decision-making. Law Hum. Behav. 30, 183–202. https://doi.org/10.1007/s10979-006-9027-y.

Brody, L.R., Hall, J.A., 2000. Gender, emotion, and expression. In: Lewis, M., Haviland-Jones, J.M. (Eds.), Handbook of Emotions: Part IV: Social/Personality Issues, 2nd ed. Guilford Press, New York, NY, pp. 325–414.

Buckholtz, J.W., Marois, R., 2012. The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. Nat. Neurosci. 15, 655–661.

Buckholtz, J.W., Asplund, C.L., Dux, P.E., Zald, D.H., Gore, J.C., Jones, O.D., Marois, R., 2008. The neural correlates of third-party punishment. Neuron 60 (5), 930–940. https://doi.org/10.1016/j.neuron.2008.10.016S0896-6273(08)00889-1. ([pii]).

Buckholtz, J.W., Martin, J.W., Treadway, M.T., Jan, K., Zald, D.H., Jones, O., Marois, R., 2015. From blame to punishment: disrupting prefrontal cortex activity reveals norm enforcement mechanisms. Neuron 87 (6), 1369–1380. https://doi.org/10.1016/j.neuron.2015.08.023.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S.B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct. Funct. 217 (4), 783–796. https://doi.org/10.1007/s00429-012-0380-y.

Capestany, B.H., Harris, L.T., 2014. Disgust and biological descriptions bias logical reasoning during legal decision-making. Soc. Neurosci. 9 (3), 265–277. https://doi.org/10.1080/17470919.2014.892531.

Carlsmith, Kevin M., Darley, John M., Robinson, Paul H., 2002. Why do we punish?: deterrence and just deserts as motives for punishment. J. Personal. Social. Psychol. 83 (2), 284–299. https://doi.org/10.1037/0022-3514.83.2.284.

Diamond, A., 2011. Biological and social influences on cognitive control processes dependent on prefrontal cortex. Prog. Brain Res. 189, 319–339. https://doi.org/10.1016/B978-0-444-53884-0.00032-4.

Dixon, M.L., 2015. Cognitive control, emotional value, and the lateral prefrontal cortex. Front. Psychol. 6, 758. https://doi.org/10.3389/fpsyg.2015.00758.

Duverne, S., Koechlin, E., 2017. Rewards and cognitive control in the human prefrontal cortex. Cereb. Cortex 27 (10), 5024–5039. https://doi.org/10.1093/cercor/bhx210.

Fischer, A.H., Rodriguez Mosquera, P.M., van Vianen, A.E., Manstead, A.S., 2004. Gender and culture differences in emotion. Emotion 4 (1), 87–94. https://doi.org/10.1037/1528-3542.4.1.87.

Friesdorf, R., Conway, P., Gawronski, B., 2015. Gender differences in responses to moral dilemmas: a process dissociation analysis. Personal. Soc. Psychol. Bull. 41 (5), 696–713. https://doi.org/10.1177/0146167215575731.

Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., Priori, A., 2010. Gender-related differences in moral judgments. Cogn. Process 11 (3), 219–226. https://doi.org/10.1007/s10339-009-0335-2.

Ginther, M.R., Bonnie, R.J., Hoffman, M.B., Shen, F.X., Simons, K.W., Jones, O.D., Marois, R., 2016. Parsing the behavioral and brain mechanisms of third-party punishment. J. Neurosci. 36 (36), 9420–9434. https://doi.org/10.1523/JNEUROSCI.4499-15.2016.

Goodenough, O.R., 2001. Mapping cortical areas associated with legal reasoning and moral intuition. Jurimetrics 41, 429–442.

Goodenough, O.R., Prehn, K., 2004. A neuroscientific approach to normative judgment in law and justice. Philos. Trans. R. Soc. Lond. B Biol. Sci. 359 (1451), 1709–1726. https://doi.org/10.1098/rstb.2004.1552.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. Neuron 44 (2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027. (doi: S0896627304006348 [pii]).

Haidt, J., 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol. Rev. 108 (4), 814–834.

Harenski, C.L., Antonenko, O., Shane, M.S., Kiehl, K.A., 2008. Gender differences in neural mechanisms underlying moral sensitivity. Soc. Cogn. Affect Neurosci. 3 (4), 313–321. https://doi.org/10.1093/scan/nsn026. (doi: nsn026 [pii].

Jeurissen, D., Sack, A.T., Roebroeck, A., Russ, B.E., Pascual-Leone, A., 2014. TMS affects moral judgment, showing the role of DLPFC and TPJ in cognitive and emotional processing. Front. Neurosci. 8, 18. https://doi.org/10.3389/fnins.2014.00018.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E., 2006. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. Science 314 (5800), 829–832. https://doi.org/10.1126/science.1129156.

Koster-Hale, J., Saxe, R., Dungan, J., Young, L.L., 2013. Decoding moral judgments from neural representations of intentions. Proc. Natl. Acad. Sci. USA 110 (14), 5648–5653. https://doi.org/10.1073/pnas.1207992110.

Krall, S.C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P.T., Eickhoff, S.B., Konrad, K., 2015. The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. Brain Struct. Funct. 220 (2), 587–604. https://doi.org/10.1007/s00429-014-0803-z.

Krueger, F., Hoffman, M., 2016. The emerging neuroscience of third-party punishment. Trends Neurosci. 39 (8), 499–501. https://doi.org/10.1016/j.tins.2016.06.004.

Krueger, F., Hoffman, M., Walter, H., Grafman, J., 2013. An fMRI investigation of the effects of belief in free will on third-party punishment. Soc. Cogn. Affect Neurosci. https://doi.org/10.1093/scan/nst092.

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 (4), 1277–1286. https://doi.org/10.1016/j.neuroimage.2012.03.068.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.

Molenberghs, P., Johnson, H., Henry, J.D., Mattingley, J.B., 2016. Understanding the minds of others: a neuroimaging meta-analysis. Neurosci. Biobehav. Rev. 65, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020.

Ott, D.V., Ullsperger, M., Jocham, G., Neumann, J., Klein, T.A., 2011. Continuous theta-burst stimulation (cTBS) over the lateral prefrontal cortex alters reinforcement learning bias. Neuroimage 57 (2), 617–623. https://doi.org/10.1016/j.neuroimage.2011.04.038.

Salerno, J.M., Bottoms, B.L., 2009. Emotional evidence and jurors' judgments: the promise of neuroscience for informing psychology and law. Behav. Sci. Law 27 (2), 273–296. https://doi.org/10.1002/bsl.861.

Schein, Chelsea, Gray, Kurt, 2017. The theory of dyadic morality: reinventing moral judgment by redefining harm. Personality and Social Psychology Review. https://doi.org/10.1177/1088868317698288.

Schleim, S., Spranger, T.M., Erk, S., Walter, H., 2011. From moral to legal judgment: the influence of normative context in lawyers and other academics. Soc. Cogn. Affect Neurosci. 6 (1), 48–57. https://doi.org/10.1093/scan/nsq010.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009.

Sellaro, R., Guroglu, B., Nitsche, M.A., van den Wildenberg, W.P., Massaro, V., Durieux, J., Colzato, L.S., 2015. Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction. Neuropsychologia 77, 400–408. https://doi.org/10.1016/j.neuropsychologia.2015.09.016.

Shen, F.X., Hoffman, M.B., Jones, O.D., Greene, J.D., Marois, R., 2011. Sorting guilty minds. N. Y. Univ. Law Rev. 86, 1306–1360 (Medline).

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., Fehr, E., 2007. The neural signature of social norm compliance. Neuron 56 (1), 185–196. https://doi.org/10.1016/j.neuron.2007.09.011.

Treadway, M.T., Buckholtz, J.W., Martin, J.W., Jan, K., Asplund, C.L., Ginther, M.R., Marois, R., 2014. Corticolimbic gating of emotion-driven punishment. Nat. Neurosci. 17 (9), 1270–1277. https://doi.org/10.1038/nn.3781.

Vidmar, N., 2011. The psychology of trial judging. Curr. Dir. Psychol. Sci. 20 (1).

Yamada, M., Camerer, C.F., Fujie, S., Kato, M., Matsuda, T., Takano, H., Takahashi, H., 2012. Neural circuits in the brain that are activated when mitigating criminal sentences. Nat. Commun. 3, 759. https://doi.org/10.1038/ncomms1757.

Yan, C.G., Wang, X.D., Zuo, X.N., Zang, Y.F., 2016. DPABI: data processing & analysis for (Resting-State) brain imaging. Neuroinformatics 14, 339–351. https://doi.org/10.1007/s12021-016-9299-4.

Yoder, K.J., Decety, J., 2014. The good, the bad, and the just: justice sensitivity predicts neural response during moral evaluation of actions performed by others. J. Neurosci. 34 (12), 4161–4166. https://doi.org/10.1523/JNEUROSCI.4648-13.2014.

Young, L., Dungan, J., 2012. Where in the brain is morality? Everywhere and maybe nowhere. Soc. Neurosci. 7 (1), 1–10. https://doi.org/10.1080/17470919.2011.569146.

Young, L., Saxe, R., 2009. Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. Neuropsychologia 47 (10), 2065–2072. https://doi.org/10.1016/j.neuropsychologia.2009.03.020.

Young, L., Saxe, R., 2011. When ignorance is no excuse: different roles for intent across moral domains. Cognition 120 (2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005. (doi: S0010-0277(11)00110-7 [pii]).

Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. Proc. Natl. Acad. Sci. USA 104 (20), 8235–8240. https://doi.org/10.1073/pnas.0701408104.

Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., 2010. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc. Natl. Acad. Sci. USA 107 (15), 6753–6758. https://doi.org/10.1073/pnas.0914826107.