The Journal of Neuroscience, March 19, 2014 • 34(12):4161–4166 • 4161

Brief Communications

The Good, the Bad, and the Just: Justice Sensitivity Predicts Neural Response during Moral Evaluation of Actions Performed by Others

Keith J. Yoder[1] **and Jean Decety**[1,2]

[1]Department of Psychology and Center for Cognitive and Social Neuroscience, University of Chicago, Chicago, Illinois 60637, [2]Department of Psychiatry, University of Chicago Medicine, Chicago, Illinois 60637

Morality is a fundamental component of human cultures and has been defined as prescriptive norms regarding how people should treat one another, including concepts such as justice, fairness, and rights. Using fMRI, the current study examined the extent to which dispositions in justice sensitivity (i.e., how individuals react to experiences of injustice and unfairness) predict behavioral ratings of praise and blame and how they modulate the online neural response and functional connectivity when participants evaluate morally laden (good and bad) everyday actions. Justice sensitivity did not impact the neuro-hemodynamic response in the action-observation network but instead influenced higher-order computational nodes in the right temporoparietal junction (rTPJ), right dorsolateral and dorsomedial prefrontal cortex (rdlPFC, dmPFC) that process mental states understanding and maintain goal representations. Activity in these regions predicted praise and blame ratings. Further, the hemodynamic response in rTPJ showed a differentiation between good and bad actions 2 s before the response in rdlPFC. Evaluation of good actions was specifically associated with enhanced activity in dorsal striatum and increased the functional coupling between the rTPJ and the anterior cingulate cortex. Together, this study provides important knowledge in how individual differences in justice sensitivity impact neural computations that support psychological processes involved in moral judgment and mental-state reasoning.

Key words: dorsal striatum; dorsolateral prefrontal cortex; justice sensitivity; moral cognition; right TPJ

Introduction

Given the importance of moral evaluations in human cultures, neuroscience investigations are critical to elucidate which computational processes underpin moral cognition. fMRI and lesion studies indicate that moral judgments arise from the integration of cognitive and affective systems, which involve the posterior superior temporal sulcus (pSTS/TPJ), amygdala, insula, ventromedial prefrontal cortex (vmPFC), dorsolateral prefrontal cortex (dlPFC), and medial prefrontal cortex (mPFC) (Greene et al., 2004; Young and Koenigs, 2007; Buckholtz and Marois, 2012). Moreover, it has become clear that these systems are not specific to morality; rather, they support more domain-general processing, such as affective arousal, attention, intention understanding, and decision-making (Decety et al., 2012; Young and Dungan, 2012). Importantly, some of these regions overlap with a salience network anchored by orbital frontoinsula (FI) and dorsal anterior cingulate cortex (dACC) associated with orienting toward and facilitating the processing of personally and motivationally salient social information (Harsay et al., 2012).

Progress has been made in determining what psychological factors impact moral decision-making at the behavioral level (Young and Saxe, 2009a; Gleichgerrcht et al., 2011). Notably, some studies suggest that personality differences in justice sensitivity may be no less important than situational factors (Edele et al., 2013). Justice sensitivity reflects the individual's concern for justice and is an important predictor of justice-related emotion and behavior (Baumert et al., 2013). There has been no work on how such individual differences are represented neurally. Examining how justice sensitivity interacts with neural circuits involved in coding moral valence, action observation, or goal/intention representation has the potential to inform the larger debate over the relative roles of affect and cognition in moral judgment.

To characterize how individual differences in justice sensitivity impact neural responses when witnessing morally laden behavior, participants were shown nonverbal visual stimuli depicting a range of interactions between two individuals, from interpersonal harm (bad intent and outcome) to interpersonal assistance (good intent and outcome). Situations judged to be morally bad were expected to be more salient and so were hypothesized to show greater recruitment of areas implicated in socioemotional salience and third-party punishments, including FI, dACC, and inferior parietal cortex. Conversely, situations depicting assistance were predicted to be associated with reward processing (Decety and Porges, 2011) and to require additional

Received Nov. 1, 2013; revised Jan. 23, 2014; accepted Feb. 10, 2014.

Author contributions: J.D. designed research; J.D. and K.J.Y. performed research; J.D. and K.J.Y. analyzed data; J.D. and K.J.Y. wrote the paper.

This study was supported by National Institute of Mental Health R01 Grant 1R01MH087525-01A2 and grants from the John Templeton Foundation (The Science of Philanthropy Initiative and Wisdom Research) to J.D.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Jean Decety, Department of Psychology, University of Chicago, 5801 S. Ellis Avenue, Chicago, IL 60637. E-mail: decety@uchicago.edu.

DOI:10.1523/JNEUROSCI.4648-13.2014

Copyright © 2014 the authors 0270-6474/14/344161-06$15.00/0

cognitive processing for analyzing the goals and benefits depicted in the scenes, and were thus expected to recruit dorsomedial and dorsolateral prefrontal regions involved in goal representation (Miller and Cohen, 2001), and moral decision-making (Buckholtz and Marois, 2012). Moreover, given the primary importance of the rTPJ in analyzing intentionality and mental states (Young and Saxe, 2008), this region was selected as a seed for effective connectivity analysis and expected to yield differential neural coupling for good and bad actions. If justice sensitivity is based on affective arousal, higher scores should enhance recruitment and connectivity within the salience network. Conversely, if justice sensitivity taps into cognitive representations, it should modulate neural response in the executive control network.

Materials and Methods

Subjects. Forty healthy adults (21 female, mean age 21.00 ± 2.58 years) participated in the study and were paid for their participation. The study was approved by the institutional review board of the University of Chicago.

Dispositional measures. Before scanning, participants completed an online questionnaire that included age, gender, education level, political orientation, and religiosity. Participants also completed the Questionnaire of Cognitive and Affective Empathy (QCAE) (Reniers et al., 2011) and the Justice Sensitivity Inventory (Schmitt et al., 2010). The Justice Sensitivity Inventory is a self-report psychometric measure that assesses four perspectives of justice sensitivity and produces four scores between 1 and 6, which index an individual's disposition to react to unfair situations. Because beneficiary, observer, and perpetrator sensitivity are closely related, these scores are often collapsed to create a measure of other-oriented sensitivity (Edele et al., 2013). Although related, self-orientation and other-orientation represent reliably distinct constructs that can exert independent and opposing influences on behavior (Gollwitzer et al., 2009). Self-orientation tends be associated with higher neuroticism and lower agreeableness, whereas other-oriented justice sensitivity is related to high agreeableness, conscientiousness, and empathy (Schmitt et al., 2010).

Stimuli and procedure. A set of dynamic visual stimuli was created to depict ecologically valid dyadic everyday social interactions that resulted in either personal harm or personal assistance. For each interaction, one actor was instructed to carry out an intentional action directed toward another individual (e.g., pulling hair or helping up off of the floor). Importantly, the faces of protagonists were not visible; thus, there was no emotional reaction visible to participants. Three still frames were extracted from each clip and presented in succession to create apparent motion (1000, 200, and 1000 ms, respectively). These stimuli were assessed by 90 subjects who did not take part in the fMRI study (42 males and 48 females; age 25.59 ± 9.57 years) who were asked with a 7 point Likert scale to rate the outcome of the actions and motives of the acting agents. Based on the outcome and motive ratings, 90 scenes were then selected to represent a broad range of actions (mean motive = 3.93 ± 1.66; mean outcome = 3.97 ± 1.62). These ratings were then converted to a categorical variable (Bad or Good) for comparison with the participants' responses in the scanner.

In the scanner, participants classified each scene three times in a mixed block/event-related design. Before the first run, a brief training run familiarized participants with the task structure. At the start of each block, participants were shown a cue word to indicate whether they should "focus on the motive or intention of the person performing the action" ("Motive") or "focus on the outcome of the action" ("Outcome"). After the cue, six scenarios of mixed moral valence were shown. Each scenario was followed by a 1 s fixation cross, then a response screen where participants pressed one of three buttons to indicate whether the scene was good, bad, or neither. The response slide remained on screen for 3 s and was followed by a jittered fixation cross (2 ± 0.7 s). As a control, participants also completed blocks in which they indicated whether the interaction occurred inside or outside ("Location"). Block order and scene order within blocks were pseudo-randomized. After the fMRI scanning, participants viewed each scene again and indicated on a visual analog scale how much they would blame or praise the actor.

Table 1. Brain regions showing significant hemodynamic increase to moral valence of actions performed by others[a]

Contrast	Brain region	MNI coordinates *x*	*y*	*z*	Cluster size	T
Bad > good						
	R supramarginal gyrus	58	−32	24	1390	7.54
	L insula	−52	6	4	122	4.15
	L supramarginal gyrus	−54	−32	22	1233	8.67
	L midcingulate	−12	−28	42	145	4.87
	R midcingulate	14	−28	40	10	3.37
	L superior parietal lobule	−22	−56	68	66	4.41
	R superior frontal gyrus	14	−4	74	45	4.15
	L hippocampus	−20	−12	−10	7	3.46
	R hippocampus	38	−10	−18	5	3.38
Good > bad						
	R middle temporal	58	−2	−24	47	3.92
	L middle temporal	−56	−2	−24	146	5.38
	L inferior temporal	−56	−44	−18	380	5.40
	R inferior temporal	56	−54	−16	83	3.83
	L inferior frontal	−46	40	−10	605	7.02
	R inferior frontal	34	32	−14	81	3.82
	R dlPFC	46	48	12	799	4.68
	L dlPFC	−40	10	48	1257	7.00
	L precuneus	−2	−60	30	230	4.30
	L insula	−34	−20	18	141	4.51
	R insula	36	−6	22	30	3.69
	L caudate	−18	−2	24	35	3.49
	R caudate	18	2	22	29	3.70
	L superior medial frontal	−4	28	40	1102	5.22
	R superior frontal	24	22	58	152	3.75

[a] Voxel-wise FDR: $q < 0.001$. R, Right; L, left.

fMRI scanning. Participants were scanned with a 3T Philips Achieva Quasar scanner. High-resolution structural T_1-weighted images were acquired using a 3D MP-RAGE sequence (TR/TE = 8.1 ms/3.7 ms; 1 mm^3 voxels, matrix = 224 × 224). For functional images, 4-mm-thick transverse slices oriented along the AC-PC line were collected with a 0.5 mm skip gap using a single-shot EPI sequence (TR/TE = 2000 ms/25 ms; flip angle = 77°; 3.5 mm × 3.5 mm × 4 mm voxels; FOV = 224 mm × 224 mm; matrix = 64 × 64).

MRI image processing. MRI images were processed with SPM8 (Wellcome Department of Imaging Neuroscience, London) in MATLAB (MathWorks). Structural scans were coregistered to the SPM8 T1 template, and a skull-stripped image was created from the segmented gray matter, white matter, and CSF images. These segmented images were combined to create a subject-specific brain template. EPI images were realigned and filtered (128 s cutoff), then coregistered to these brain templates, normalized to MNI space, and smoothed (8 mm FWHM). Images showing $>$0.5 mm/TR were interpolated and then deweighted from first-level design matrices with the ArtRepair toolbox.

Data were entered into a general linear model, with movement parameters as nuisance regressors. Participant's in scanner responses (Bad, Good) within each block (Outcome, Motive, or Location) were modeled separately with duration 2.2 s beginning at the onset of the scenario. For Location blocks, the moral classification from Outcome and Motive blocks was used. Groupwise contrasts were corrected for multiple comparisons using a false discovery rate (FDR) of $q < 0.05$. To identify regions modulated by moral valence, hemodynamic activity during scenes classified as bad (MotiveBad and OutcomeBad) was subtracted from scenes classified as good (MotiveGood and OutcomeGood). Conversely, task effects were modeled by collapsing across valences (e.g., MotiveGood and MotiveBad). To identify regions modulated by dispositional differences, justice sensitivity and empathy scores were entered as covariates and thresholded at uncorrected $p < 0.001$ with a cluster extent of at least 25 contiguous voxels. Coordinates for two ROIs (rTPJ: $x = 62$, $y = -54$, $z = 16$; dmPFC: $x = 0$, $y = 54$, $z = 36$) were taken from a recent meta-analysis of fMRI studies of morality (Bzdok et al., 2012). An additional a priori ROI was taken from Buckholtz et al., 2008 (rdlPFC: $x = 39$,

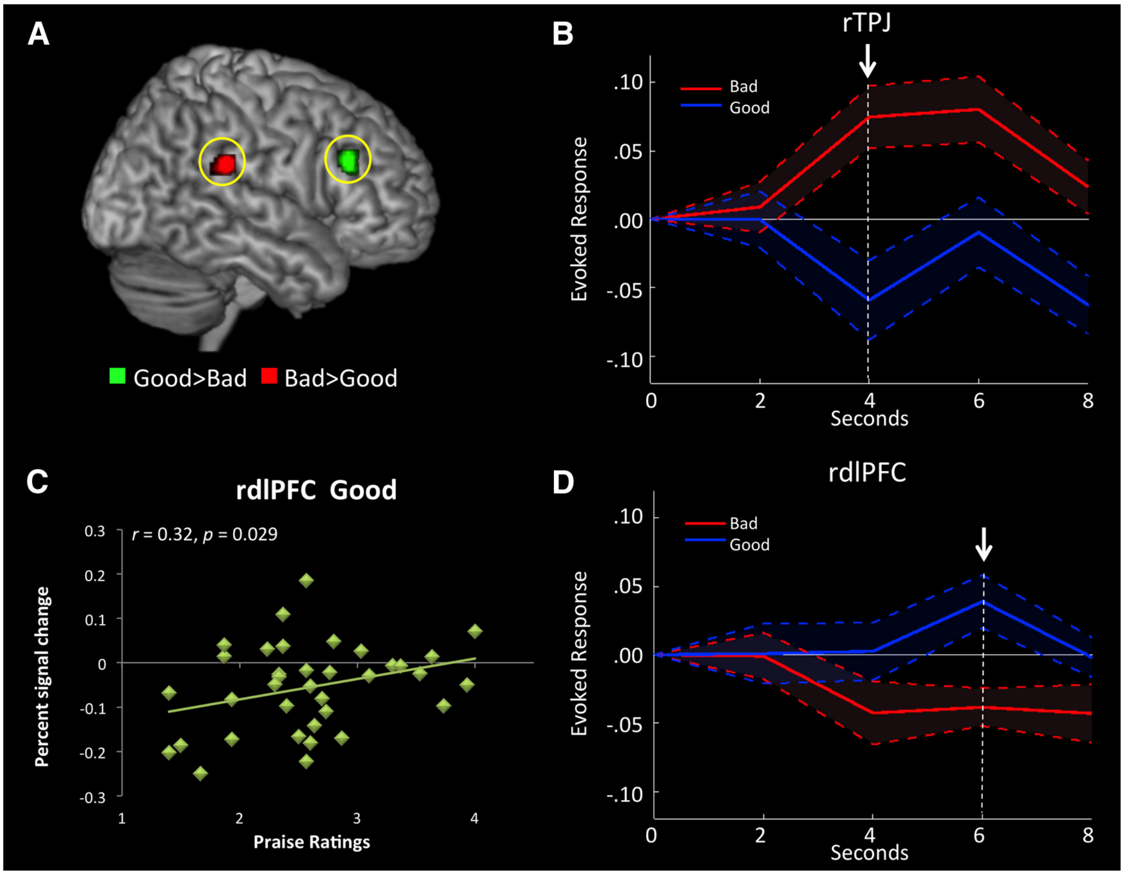

Figure 1. BOLD activity and time courses in rTPJ and rdlPFC. ***A***, Clusters in rTPJ (peak: $x = 60$, $y = -32$, $z = 24$) and rdlPFC (peak: $x = 42$, $y = 30$, $z = 26$) from the whole-brain contrast thresholded at $p < 0.0001$ for viewing. ***B***, ***D***, Percentage signal change from rTPJ and rdlPFC from whole-brain (FDR: $q < 0.05$) is plotted for 8 s after stimulus onset, evaluated as good (blue) and bad (red) actions. Arrows indicate the first bin with significant increase. The rTPJ showed signal change earlier (4 s, $T = 0.325$, $p = 0.002$) and was significantly more responsive to bad actions, whereas the response in rdlPFC was later (6 s, $T = 2.082$, $p = 0.022$) and was greater for good actions. Bounds are SEM. ***C***, BOLD activity in rdlPFC ROI ($x = 40$, $y = 37$, $z = 26$) for good actions significantly predicted praise ratings.

$y = 37$, $z = 26$). For each ROI, mean activity within a 6-mm-radius sphere was correlated against justice sensitivity scores, as well as ratings of praise and blame. Finally, effective connectivity was assessed using psychophysiological interaction in SPM8. The psychophysiological interaction seed was created by placing a 10 mm sphere around the rTPJ coordinates; and, given the a priori hypothesis on the role of this region in moral judgment (Young and Saxe, 2009b; Decety and Cacioppo, 2012), results were thresholded at $p < 0.001$ (uncorrected) with a spatial extent threshold of $k = 25$. Percentage signal change and time courses (peristimulus time histograms) from clusters were extracted using the rfxplot toolbox (Gläscher, 2009).

Results

Two subjects whose responses on the dispositional measures were >2 SDs away from the group mean were removed as outliers, producing a final sample of 38. In scanner, responses and postscan ratings indicated that participants classified harmful actions as morally bad and deserving of blame, whereas scenes depicting assistance were classified as morally good and deserving of praise (all $p < 0.001$), which matched those of the 90 validators (Cohen's $\kappa = 0.953$).

Dispositional measures

Participants who scored high on self ($r = 0.39$, $p = 0.009$) and other-oriented ($r = 0.45$, $p = 0.003$) justice sensitivity assigned significantly more blame when evaluating harm. Self-oriented justice sensitivity disposition also predicted more praise for interpersonal assistance ($r = 0.31$, $p = 0.033$). Total QCAE scores were significantly correlated with other-oriented ($r = 0.40$, $p = 0.002$) and trended with self-oriented ($r = 0.32$, $p = 0.054$) justice sensitivity. No QCAE measures were significantly related to any neural or behavioral data.

Moral valence

Regions with greater activity during scenarios classified as morally good included the middle and inferior temporal gyri, inferior frontal gyrus, dlPFC, insula, dACC, and precuneus (Table 1). The dorsal striatum was significantly more active during good scenes than bad ones. The reverse contrast showed increased signal during scenes classified as morally bad in the hippocampus and midcingulate cortex, superior frontal gyrus, rTPJ, supramarginal gyrus, left inferior and superior parietal lobules, and insula. Interestingly, the time courses and shapes of the response in rTPJ and rdlPFC (Fig. 1) demonstrate an earlier differentiation between good and bad actions in the former region (by 4 s) than in the latter (by 6 s). Neural response was greater for bad actions in the rTPJ, whereas the reverse was true for rdlPFC (Fig. 1). Mean percentage signal change in rdlPFC, rTPJ, and dmPFC was significantly correlated to blame/praise ratings (Figs. 1 and 2).

Outcome and motive

Focusing on Motives, compared with Outcomes, showed greater activity in regions implicated in mentalizing, including bilateral

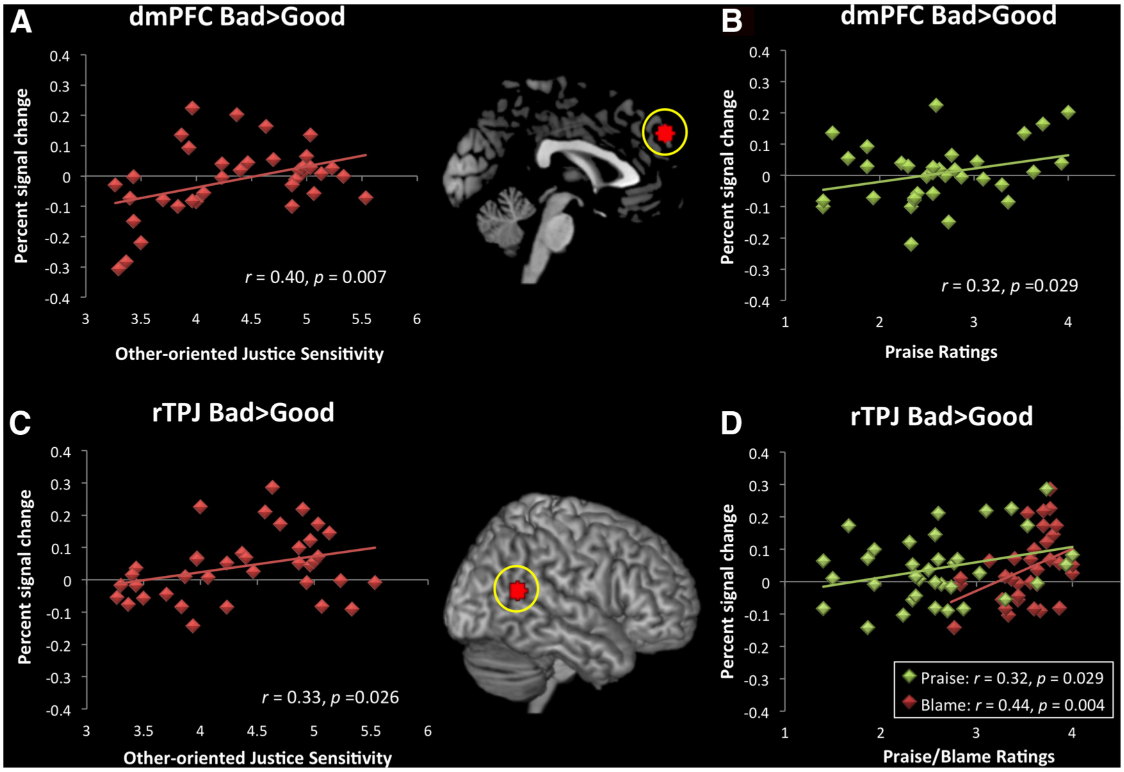

Figure 2. Justice sensitivity and postscan ratings predict moral sensitivity in ROIs dmPFC (top; $x = 0$, $y = 54$, $z = 36$) and rTPJ (bottom; $x = 62$, $y = -54$, $z = 16$). Other-oriented justice sensitivity was positively correlated with activity in dmPFC (***A***) and rTPJ (***C***) for Bad > Good. ***B***, Greater activity for Bad > Good in dmPFC was related to higher praise ratings (green). ***D***, Greater activity in rTPJ correlated with increased praise (green) and blame (red) ratings.

pSTS/TPJ ($x = 66$, $y = -34$, $z = 8$; $x = -50$, $y = -26$, $z = 30$), mPFC ($x = -14$, $y = 46$, $z = 20$), and PCC ($x = -16$, $y = -56$, $z = 42$) (Fig. 3). Motive blocks elicited greater BOLD signal in right fusiform ($x = 36$, $y = -42$, $y = -24$), middle temporal gyrus ($x = 56$, $y = -30$, $z = -8$), as well as right FI ($x = 34$, $y = 30$, $z = -8$) and left insula ($x = -42$, $y = -4$, $z = 4$). No clusters were found to show greater BOLD signal for Outcome > Motive.

Effective connectivity

Functional connectivity analyses (Fig. 3*C*,*D*) seeded in the rTPJ during the evaluation of good actions (vs bad actions) showed a significant increase in coupling with rdlPFC ($x = 48$, $y = 18$, $z = 26$), left dlPFC ($x = -30$, $y = 16$, $z = 44$), ACC/mPFC ($x = 0$, $y = 46$, $z = 10$), right MTG ($x = 50$, $y = -32$, $z = -6$), inferior temporal gyrus ($x = -42$, $y = -36$, $z = -22$), and PCC/precuneus ($x = -6$, $y = -50$, $z = 12$). No region showed significantly greater connectivity with rTPJ in the Bad–Good contrast.

Dispositional influences on neural activity

Other-oriented justice sensitivity was significantly positively correlated with signal change in the Bad versus Good contrast in dorsal SFG ($x = 16$, $y = 22$, $z = 60$), dmPFC, and rTPJ (Fig. 2). Other-oriented sensitivity did not significantly predict BOLD differences between outcome versus motive contrasts. Participants with higher scores on other-oriented justice sensitivity showed greater coupling between rTPJ and right parahippocampal gyrus ($x = 26$, $y = -24$, $z = -20$), right MTG ($x = 50$, $y = -64$, $z = 2$), bilateral SFG ($x = 26$, $y = 2$, $z = 60$; $x = -28$, $y = 8$, $z = 56$), and rdlPFC ($x = 20$, $y = 44$, $z = 30$).

Discussion

The current study demonstrates that individual dispositions of justice sensitivity not only predict subjective ratings of praise and blame but also modulate the online neural response and functional connectivity when individuals evaluate morally laden actions performed by others. Moreover, whether individuals focused on the motive or the outcome of the action, justice sensitivity did not impact the neural response in the action observation network (Grafton, 2009) but instead influenced higher-order computations that maintain goals and guide moral decision-making. This finding supports the notion that humans perceive social interactions from their conspecifics as fundamentally linked with mental states and readily make moral evaluations about others' actions (Malle and Guglielmo, 2012).

Individuals with high other-oriented justice sensitivity scores assigned more blame for harmful interactions and exhibited enhanced recruitment of rTPJ and dmPFC for bad actions. Further, this activity predicted greater praise and blame ratings (Fig. 2). This fits neatly with previous research, which demonstrated that rTPJ first encodes beliefs before integrating the moral valence of these beliefs with dmPFC (Young and Saxe, 2008). Sensitivity for other-directed injustice also predicted functional connectivity seeded in rTPJ for good actions. Whereas past work has linked other-oriented justice sensitivity with altruistic giving (Edele et al., 2013), our results provide some of the first evidence for the role of justice sensitivity in enhancing neural processing of moral information in specific components of the network involved in moral judgment. Interestingly, empathy scores did not predict moral evaluations nor BOLD changes in this study.

Among those regions showing enhanced coupling with rTPJ was the rdlPFC (Fig. 3). Previously, the rdlPFC was reported to vary parametrically with the level of assigned punishment (Buckholtz et al., 2008). Here the response in rdlPFC predicted praise ratings, and this region was more active during good scenarios than bad. This supports the role of rdlPFC in maintaining

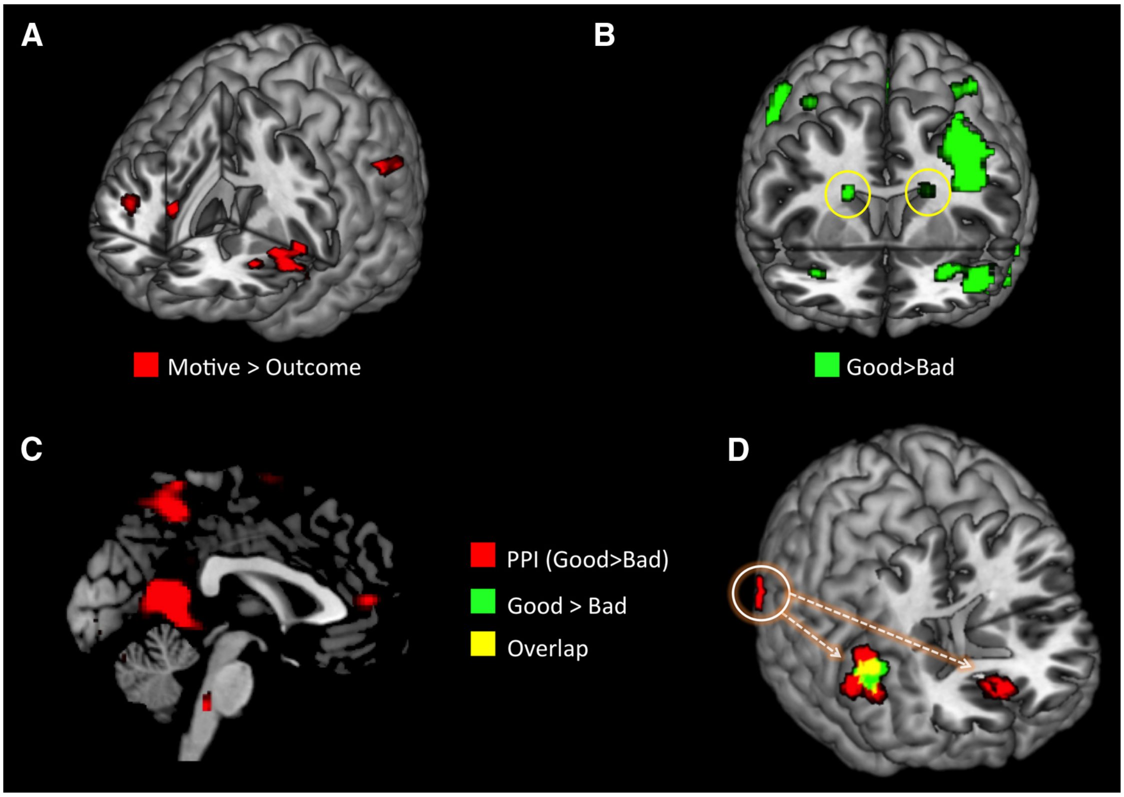

Figure 3. ***A***, Regions with greater activity for judgments made during Motive blocks ($p < 0.001$). ***B***, Whole-brain responses for good actions versus bad actions (FDR: $q < 0.05$). Striatum is circled. ***C***, Increased functional connectivity seeded in rTPJ ($x = 62, y = -54, z = 16$; circled in ***D***) is shown on a midline sagittal slice. ***D***, Increased coupling with ACC/mPFC ($x = 0, y = 46, z = 10$) and a cluster in rdlPFC (peak: $x = 54, y = 20, z = 16$) that overlaps with that found in the whole-brain Good > Bad contrast.

representations of actions and their outcomes in service of determining culpability (Buckholtz and Marois, 2012). Increased sensitivity to good actions was complemented by the rTPJ response to bad actions. Notably, time courses of the BOLD response in rTPJ occurred 2 s before the response in rdlPFC (Fig. 1). This is in keeping with previous research using fMRI (Buckholtz et al., 2008) and high-density EEG/ERP (Decety and Cacioppo, 2012) that found that the rTPJ show activity (~60 ms after stimulus onset) before rdlPFC when perceiving intentional harmful actions. Thus, evaluating interpersonal assistance (good actions) requires recruitment of prefrontal systems (e.g., dlPFC, dmPFC) in service of maintaining representations of others' intentions and outcomes of their actions, whereas evaluations of harm can be produced earlier and faster from the posterior STS/TPJ, a region particularly suited for parsing streams of input across modalities and extracting meaning from them (Pelphrey and Carter, 2008).

Interestingly, evaluating good actions elicited greater activity in dorsal striatum (Fig. 3), a region that receives inputs from dlPFC and the dopaminergic midbrain, and which has intrinsic connectivity with other prefrontal areas involved in cognition, such as dmPFC (Seeley et al., 2007). The dorsal striatum is also known to mediate important aspects of decision-making, particularly those related to encoding specific action–outcome associations, and the selection of actions on the basis of their currently expected reward value (Balleine et al., 2007). Recent work has also shown response in the dorsal striatum when individuals imagined helping others (Decety and Porges, 2011). Thus, striatum activity in our study may reflect enhanced processing of the reward value of good actions, relative to bad ones. This fits well with the suggestion that interactions between dorsal striatum and dlPFC are particularly important for guiding behavior according to long-term rewards, consistent with the ability of dlPFC to maintain stable goals representation over time (Miller and Cohen, 2001; Buckholtz and Marois, 2012).

An alternative explanation is that bad actions are more salient than good actions. This would explain increased activity for bad actions in some nodes of the salience network, including bilateral IPL, right SPL, and left dorsal anterior insula. However, good actions were associated with increased activity in bilateral anterior insula, right supramarginal gyrus, and a large dorsomedial prefrontal cluster that extended into dorsal ACC and SMA (another node in the salience network). Thus, bad actions do not seem to be uniformly processed as more salient. Rather, different nodes of the salience network show preferential sensitivity for different moral valences.

Of course, the salience network does not act in isolation. Recent work suggests that, during moral reasoning, one role of the salience network, anchored by the insula and ACC, is to coordinate activity between the cognitive control network and default mode network (Chiong et al., 2013). Effective connectivity analyses provide insights into this interplay. The rTPJ was chosen as a seed because this region is known to have an important role in understanding the mental states of others (Perner et al., 2006; Young et al., 2007; Young and Saxe, 2009a; Decety and Cacioppo, 2012), as well as serving as an attentional hub (Decety and Lamm, 2007). Evaluating good actions, compared with bad actions, showed increased connectivity between rTPJ and ACC/mPFC, suggesting one potential locus for the salience network to exert influence on moral decision-making. Moreover, rTPJ showed increased coupling with PCC/precuneus and bilateral dlPFC (Fig. 3). This pattern of connectivity mirrors a proposed network for integrating information about harm and intentionality to appropriately titrate punishment (Buckholtz and Marois, 2012).

Finally, in line with many previous studies (Young et al., 2007; Young and Saxe, 2009b; Decety et al., 2012), when participants focused on the motive of the agent, they demonstrated enhanced recruitment of mentalizing regions (bilateral TPJ, PCC, and mPFC).

Overall, our study provides important insights into how dispositional justice sensitivity influences evaluations of morally laden events not limited to interpersonal harm, but also good actions. Justice sensitivity modulates activity across several domain-general systems, particularly in regions of the prefrontal cortex involved in goal representations in service of moral decision-making, and importantly does not influence the action observation network. These findings extend the role of the dlPFC, rTPJ, and dorsal striatum in evaluating other people's behavior, and have important implications for the roles of rTPJ/pSTS in moral cognition.

References

Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. J Neurosci 27:8161–8165. CrossRef Medline

Baumert A, Rothmund T, Thomas N, Gollwitzer M, Schmitt M (2013) Justice as a moral motive: belief in a just world and justice sensitivity as potential indicator of the justice motive. In: Handbook of moral motivation (Heinrichs K, Oser F, Lovat T, eds), pp 159–180. Rotterdam, The Netherlands: Sense.

Buckholtz JW, Marois R (2012) The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. Nat Neurosci 15:655–661. CrossRef Medline

Buckholtz JW, Asplund CL, Dux PE, Zald DH, Gore JC, Jones OD, Marois R (2008) The neural correlates of third-party punishment. Neuron 60: 930–940. CrossRef Medline

Bzdok D, Schilbach L, Vogeley K, Schneider K, Laird AR, Langner R, Eickhoff SB (2012) Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct Funct 217:783–796. CrossRef Medline

Chiong W, Wilson SM, D'Esposito M, Kayser AS, Grossman SN, Poorzand P, Seeley WW, Miller BL, Rankin KP (2013) The salience network casually influences default mode network activity during moral reasoning. Brain 136:1929–1941. CrossRef Medline

Decety J, Cacioppo S (2012) The speed of morality: a high-density electrical neuroimaging study. J Neurophysiol 108:3068–3072. CrossRef Medline

Decety J, Lamm C (2007) The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist 13:580–593. CrossRef Medline

Decety J, Porges EC (2011) Imagining being the agent of actions that carry different moral consequences: an fMRI study. Neuropsychologia 49: 2994–3001. CrossRef Medline

Decety J, Michalska KJ, Kinzler KD (2012) The contribution of emotion and cognition to moral sensitivity: a neurodevelopmental study. Cereb Cortex 22:209–220. CrossRef Medline

Edele A, Dziobek I, Keller M (2013) Explaining altruistic sharing in the dictator game: the role of affective empathy, cognitive empathy, and justice sensitivity. Learn Individ Differ 24:96–102. CrossRef

Gläscher J (2009) Visualization of group inference data in functional neuroimaging. Neuroinformatics 7:73–82. CrossRef Medline

Gleichgerrcht E, Torralva T, Roca M, Pose M, Manes F (2011) The role of social cognition in moral judgment in frontotemporal dementia. Soc Neurosci 6:113–122. CrossRef Medline

Gollwitzer M, Rothmund T, Pfeiffer A, Ensenbach C (2009) Why and when justice sensitivity leads to pro- and antisocial behavior. J Res Pers 43:999–1005. CrossRef

Grafton ST (2009) Embodied cognition and the simulation of action to understand others. Ann N Y Acad Sci 1156:97–117. CrossRef Medline

Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. Neuron 44: 389–400. CrossRef Medline

Harsay HA, Spaan M, Wijnen JG, Ridderinkhof KR (2012) Error awareness and salience processing in the oddball task: shared neural mechanisms. Front Hum Neurosci 6:246. CrossRef Medline

Malle BF, Guglielmo S (2012) Are intentionality judgments fundamentally moral? In: Emotions, imagination, and moral reasoning (Langdon R, Mackenzie C, eds), pp 275–293. Hove: Psychology Press.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202. CrossRef Medline

Pelphrey KA, Carter EJ (2008) Charting the typical and atypical development of the social brain. Dev Psychopathol 20:1081–1102. CrossRef Medline

Perner J, Aichhorn M, Kronbichler M, Staffen W, Ladurner G (2006) Thinking of mental and other representations: the roles of left and right temporo-parietal junction. Soc Neurosci 1:245–258. CrossRef Medline

Reniers RL, Corcoran R, Drake R, Shryane NM, Völlm BA (2011) The QCAE: a Questionnaire of Cognitive and Affective Empathy. J Pers Assess 93:84–95. CrossRef Medline

Schmitt M, Baumert A, Gollwitzer M, Maes J (2010) The Justice Sensitivity Inventory: factorial validity, location in the personality facet space, demographic pattern, and normative data. Soc Justice Res 23:211–238. CrossRef

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 27:2349–2356. CrossRef Medline

Young L, Dungan J (2012) Where in the brain is morality? Everywhere and maybe nowhere. Soc Neurosci 7:1–10. CrossRef Medline

Young L, Koenigs M (2007) Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology. Br Med Bull 84:69–79. CrossRef Medline

Young L, Saxe R (2008) The neural basis of belief encoding and integration in moral judgment. Neuroimage 40:1912–1920. CrossRef Medline

Young L, Saxe R (2009a) Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. Neuropsychologia 47: 2065–2072. CrossRef Medline

Young L, Saxe R (2009b) An fMRI investigation of spontaneous mental state inference for moral judgment. J Cogn Neurosci 21:1396–1405. CrossRef Medline

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 U S A 104:8235–8240. CrossRef Medline