

Neural correlates of conventional and harm/welfare-based moral decision-making

Stuart F. White^{1,2} · Hui Zhao^{1,3} · Kelly Kimiko Leong^{1,4} · Judith G. Smetana⁵ · Larry P. Nucci⁶ · R. James R. Blair^{1,2}

Published online: 26 September 2017
© Psychonomic Society, Inc. (outside the US) 2017

Abstract The degree to which social norms are processed by a unitary system or dissociable systems remains debated. Much research on children’s social-cognitive judgments has supported the distinction between “moral” (harm/welfare-based) and “conventional” norms. However, the extent to which these norms are processed by dissociable neural systems remains unclear. To address this issue, 23 healthy participants were scanned with functional magnetic resonance imaging (fMRI) while they rated the wrongness of harm/welfare-based and conventional transgressions and neutral vignettes. Activation significantly greater than the neutral vignette baseline was observed in regions implicated in decision-making regions including rostral/ventral medial frontal, anterior insula and dorsomedial frontal cortices when evaluating both harm/welfare-based and social-conventional transgressions. Greater activation when rating harm/welfare-based relative to social-conventional transgressions was seen through much of ACC and bilateral inferior frontal gyrus. Greater activation was observed in superior temporal gyrus, bilateral middle temporal gyrus, left PCC, and temporal-parietal junction when rating social-conventional transgressions relative to harm/welfare-based transgressions. These data suggest that decisions regarding the wrongness of actions, irrespective of

whether they involve care/harm-based or conventional transgressions, recruit regions generally implicated in affect-based decision-making. However, there is neural differentiation between harm/welfare-based and conventional transgressions. This may reflect the particular importance of processing the intent of transgressors of conventional norms and perhaps the greater emotional content or salience of harm/welfare-based transgressions.

Keywords Morality · Decision-making · fMRI · Theory of mind · Moral/conventional

There has been considerable debate as to whether moral prescriptions and social norms are processed by a single unitary system or whether differentiable systems are involved in different forms of norm processing. There have been suggestions that all norms are communicated by a similar form of cultural transmission (Shweder, Mahapatra, & Miller, 1987) and/or that they are all processed by a unitary moral faculty (Huebner, Dwyer, & Hauser, 2009; Kohlberg, 1976; Mikhail, 2007).

An alternative view is that different forms of norms are generated and processed by at least partially dissociable systems. One of the earliest versions of this latter position is articulated by social domain theory and related to the “moral/conventional distinction”: the distinction between norms whose violation results in unjust or unfair treatment or in harmful consequences to others (harm/welfare-based; e.g., one person hitting another) as opposed to those whose violation challenges contextually relative and arbitrary social conventions or norms that structure social interactions (social-conventional; e.g., one person talking to another during class; Nucci & Nucci, 1982a, b; Smetana, Jambon, & Ball, 2014; Turiel, 1983). Within this literature, it is argued that processing of these two forms of rule occurs within distinct conceptual domains. In the distinct domain view, justice-based issues (relating to distribution of resources; Blair, Marsh,

✉ Stuart F. White
stuart.white@boystown.org

¹ Department of Health and Human Services, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

² Boys Town National Research Hospital, Omaha, NE, USA

³ China University of Political Science and Law, Beijing, China

⁴ Tufts University School of Dental Medicine, Boston, MA, USA

⁵ University of Rochester, Rochester, NY, USA

⁶ University of California, Berkeley, Berkeley, CA, USA

Finger, Blair, & Luo, 2006) are also part of the moral domain. More recently, some have claimed that there are other types of norms (Blair et al., 2006; Haidt & Graham, 2007) that should be included within the “moral” category. These include disgust (also referred to as purity norms and primarily relating to sexual acts; Haidt & Graham, 2007) and in-group/loyalty (valuing patriotism; Graham et al., 2011).

Much of the fMRI literature has concentrated on either harm/welfare-based norms concerning actions that harm others (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Harenski, Antonenko, Shane, & Kiehl, 2010; Luo et al., 2006; Shenhav & Greene, 2010) or, more recently, disgust-based norms (e.g., Chakroff et al., 2016; Parkinson et al., 2011). This literature has particularly implicated brain regions, including the amygdala, ventromedial prefrontal cortex (vmPFC), rostral medial frontal cortex, and posterior cingulate cortex (PCC; for a recent meta-analytic review, see Boccia et al., 2016), that are implicated in reinforcement-based decision-making (cf. Blair, 2007). In particular, it has been claimed that the amygdala is critical for generating associations between actions that harm others and the aversive quality of the distress cues of those harmed, and that vmPFC is critical in representing the subjective value of the action (Blair, 2007). The amygdala shows increased activity for more severe harm/welfare-based transgressions (Luo et al., 2006), and participants’ judgments of transgression severity correlates with amygdala activity (Harenski et al., 2010). However, it should be noted that the amygdala also shows strong activation by disgust-based norms in some studies and may even show stronger activation to disgust-based transgressions than to harm/welfare-based transgressions (see Parkinson et al., 2011). It has been argued that the negative valence of disgust-based transgressions is at least partly acquired through observation of the disgusted emotional reactions to the acts of those who consider the acts to be transgressions (Blair, 2007; Blair et al., 2006). Certainly, actions considered to be wrong by participants because they judge them to involve “abnormal use of body parts” are associated with disgust reactions (Giner-Sorolla, Bosson, Caswell, & Hettinger, 2012). The amygdala and anterior insula cortex (aIC) are responsive to other individuals’ disgust expressions (Murphy, Nimmo-Smith, & Lawrence, 2003) and implicated in disgust-based (or at least taste aversion) learning (Cubero, Thiele, & Bernstein, 1999; Jeon et al., 2010). In line with the suggestion that vmPFC is critical in representing the subjective value of the action, recent work has reported a relationship between number of lives saved and vmPFC activity when considering trolley problems (Hutcherson, Montaser-Kouhsari, Woodward, & Rangel, 2015; Shenhav & Greene, 2010). Furthermore, vmPFC, like amygdala, shows strong activation by disgust-based norms and may even show stronger activation to disgust-based transgressions than to harm/welfare-based transgressions (see Chakroff et al., 2016; Parkinson et al., 2011).

A series of studies have also examined neural systems involved in justice-based judgments (e.g., Fliessbach et al.,

2012; Güroğlu, Will, & Crone, 2014; Hsu, Anen, & Quartz, 2008; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003; White, Brislin, Sinclair, & Blair, 2014b). Justice-based transgressions are associated with activations within dorsomedial frontal cortex (dmFC), aIC, and dorsolateral prefrontal cortex (Güroğlu et al., 2014; Rilling et al., 2008; Sanfey et al., 2003; White et al., 2014b). Interestingly, activations in vmPFC are not typically seen unless participants are correcting unfair resource allocations that have helped themselves relative to others. That is, their actions are taken with the expectation of bringing value to others (Güroğlu et al., 2014).

In contrast to the work on harm/welfare, disgust, and justice-based transgressions, relatively little work has examined the processing of social-conventional transgressions (Berthoz, Armony, Blair, & Dolan, 2002; Carr et al., 2015; Finger, Marsh, Kamel, Mitchell, & Blair, 2006) or contrasted responses to harm/welfare-based and social-conventional norms (although, see Finger et al., 2006; Lahat, Helwig, & Zelazo, 2013). This is despite the fact that the literature on moral/conventional distinctions is the earliest and still most robust example of the multiple social norm system approach (Ball, Smetana, & Sturge-Apple, 2016; Jambon & Smetana, 2017; Nucci & Herman, 1982; Nucci & Nucci, 1982a, b; Smetana, 1981, 1985; Smetana & Braeges, 1990; Smetana et al., 2012). Berthoz et al. (2002) found that processing social-conventional transgressions relative to normative behavior was associated with increased activity in regions implicated in theory of mind (ToM; i.e., dmFC, temporal parietal junction [TPJ], temporal pole). ToM involves the representation of the mental states of others (their thoughts, intentions and beliefs; Frith & Frith, 1999). Of course, intent information can be critical when processing harm/welfare-based transgressions (cf. Jambon & Smetana, 2014; Killen, Mulvey, Richardson, Jampol, & Woodward, 2011; Young, Cushman, Hauser, & Saxe, 2007). But this occurs particularly in contexts where it is ambiguous whether the actor intends to cause harm or does so accidentally (e.g., does a person intend to put poison in coffee, or does the person think the powder is sugar?; Young et al., 2007). Similar to Berthoz et al. (2002), Finger and colleagues found a region proximal to TPJ that showed greater activity to witnessed social-conventional transgressions relative to normative vignettes (Finger et al., 2006). However, activity within this region did not significantly differ for witnessed social-conventional transgressions and harm/welfare-based transgressions (Finger et al., 2006). Lahat et al. (2013) examined the distinction between harm/welfare-based and social-conventional transgressions in an event-related potential study. The authors found increased N2 amplitudes to harm/welfare-based violations when they were *not* against the rules relative to when they were against the rules. No differences between the rule/no-rule conditions were observed for social-conventional transgressions. The authors suggest that the N2 differences reflected conflict signaling in

the harm/welfare-based/no-rule condition (individuals do not typically process the permissibility of harm/welfare-based transgressions as rule-dependent; hitting another is bad whether a rule exists or not; Blair, 1995). In short, very little work has contrasted the neural responses associated with harm/welfare-based and social-conventional norms in the same study. Moreover, work conducted to date either does not allow source localization (Lahat et al., 2013) or involved items that were embarrassing (e.g., “You start to yell for help and run to a phone when you walk into a parking meter and fall down. People on the street watch”) rather than clearly social-conventional transgressions (Finger et al., 2006).

It should be noted that the regions implicated in moral judgments in previous work (see Boccia et al., 2016), including vmPFC, aIC, dmFC, PCC, and (to a lesser extent) the amygdala, are consistently implicated in valence-based decision-making generally (Clithero & Rangel, 2014; Kuhn & Knutson, 2005). To the extent that judgments about different forms of social norms are value judgments, it is plausible that these judgments might recruit these similar neural regions. Of course, it is also possible that not all forms of social norm judgment are examples of valence-based decision-making. Judgments about justice-based transgressions appear to reflect the organization of neural regions implicated in behavioral change (e.g., aIC and dmFC; Corradi-Dell’acqua, Civai, Rumiati, & Fink, 2012; Rilling et al., 2008; Sanfey et al., 2003; Strobel et al., 2011; Tabibnia, Satpute, & Lieberman, 2008; White et al., 2014b) and often do not involve activation of vmPFC (e.g., Sanfey et al., 2003; Strobel et al., 2011). Critically, when vmPFC is implicated in evaluating justice-based transgressions, it is in contexts where the value of a choice is being represented (e.g., free vs. costly wins; Corradi-Dell’acqua et al., 2012; or costly punishment of unfair offers; White et al., 2014b). It is unclear to what extent judgments of social-conventional transgressions involve valence-based decision-making. There have been suggestions that social-conventional transgressions are not associated with affect (Kagan & Lamb, 1987). Research on children’s emotion attributions for different types of transgressions suggests that children attribute neutral emotions to social-conventional transgressions and more negative emotions to harm/welfare-based transgressions, although this varies by individuals’ roles in the transgressions (Arsenio, 1988). Alternatively, judgments of transgressions may be associated with specific forms of affect (such as affect induced by distress cues or the disgust of others). Transgressions of social-conventional norms are certainly exposed to the anger of caregivers and teachers (see Blair et al., 2006).

Given this, the current study sought to directly examine and compare neural responses to harm/welfare-based and social-conventional transgressions. Healthy adults were asked to rate the acceptability (severity) of vignettes portraying either a harm/welfare-based or a social-conventional transgression, depicted

by a line drawing and accompanied by a descriptive sentence. The study had two goals. Our first goal was to determine the neural regions involved in both harm/welfare-based and social-conventional transgressions. Our aim here was to determine the extent of existence of any form of unitary moral faculty (cf. Huebner et al., 2009; Kohlberg, 1976; Mikhail, 2007) and specifically to determine whether this might reflect the recruitment of neural systems implicated in valence-based decision-making (vmPFC, aIC, dmFC, and PCC).

Our second goal was to determine regions showing differential responsiveness to harm/welfare-based versus social-conventional transgressions. Our aim here in particular was to determine the degree to which regions showing differential responsiveness were included within, or independent of, those regions showing responses to both harm/welfare-based and social-conventional transgressions. Based on previous findings, it was hypothesized that (a) amygdala, vmPFC, and AIC would show greater activation during the rating of harm/welfare-based transgressions relative to social-conventional transgressions (Luo et al., 2006; Shenhav & Greene, 2010), and (b) as determining the intent of the actors will be more challenging than in harm/welfare-based transgressions, greater activation would be seen in regions associated with ToM (medial prefrontal cortex, temporal-parietal junction, temporal pole, precuneus) during the rating of social-conventional transgressions relative to harm/welfare-based transgressions (Berthoz et al., 2002; Finger et al., 2006).

Method

Participants

Twenty-three healthy right-handed adults (17 female), ages 21 to 35 years (mean = 25.08 years, $SD = 2.94$), from the Washington, DC, metropolitan area, volunteered for this study and were compensated monetarily for their participation. A licensed physician screened subjects for good health. The medical examination excluded subjects with past history of any neurologic disease or psychiatric disorder based on the Structured Clinical Interview for the DSM-IV Axis I Disorders (SCID; First, Spitzer, Gibbon, & Janet, 2002). Subjects gave written informed consent to participate in the study, which was approved by the National Institutes of Health Combined Neurosciences Institutional Review Board.

The moral judgment task

The moral judgment task involved the presentation of 60 vignettes. Twenty harm/welfare-based (e.g., punching a classmate) and 20 social-conventional transgressions (e.g., a boy entering the girls’ bathroom) were depicted twice each. Twenty neutral actions, or actions that did not depict a

transgression, were also depicted twice each. The vignettes contained both a single sentence describing the action and a line drawing visually depicting the vignette (see Fig. 1). Thirty-two of the vignettes described a male's actions (e.g. "A boy paints a picture of his dog"), and 22 vignettes featured only males, 21 featured only females, and 17 featured males and females. Participants were presented with each vignette for 5,000 ms, during which time the subject was asked to rate via button press how wrong they believed the action to be on a 4-point Likert scale (1 = *not wrong*, 2 = *a little bit wrong*, 3 = *fairly wrong*, 4 = *very wrong*). Each vignette was followed by a randomly jittered fixation of between 500 and 3,500 ms. The subjects completed two runs of 7 minutes and 26 seconds each. During each run, participants rated all 60 vignettes once. Trials were randomized within each run for each participant.

fMRI data acquisition and preprocessing

Whole-brain blood-oxygen-level dependent (BOLD) fMRI data were acquired using a 3T GE Signa scanner. A total of 186 functional images per run were taken with a gradient echo planar imaging (EPI) sequence (repetition time = 2,400 ms; echo time = 27 ms; 64×64 matrix; 90° flip angle; 24-cm field of view). Whole-brain coverage was obtained with 43 axial slices (thickness, 2.5 mm; .5-mm spacing; in-plane resolution, 3.75×3.75 mm). A high-resolution anatomical scan (three-dimensional spoiled gradient recalled acquisition in a steady state; repetition time = 7 ms; echo time = 2.984 ms; 24-cm field of view; 12° flip angle; 128 axial slices; thickness, 1.2 mm; 256×192 matrix) in register with the EPI data set was obtained covering the whole brain.

Imaging data preprocessing

Data were analyzed within the framework of the general linear model using Analysis of Functional Neuroimages software (AFNI; Cox, 1996). Both individual and group-level analyses were conducted. The first five volumes in each scan series,

collected before equilibrium magnetization was reached, were discarded. Motion correction was performed by registering all volumes in the EPI data set to a volume collected close to acquisition of the high-resolution anatomical data set.

The EPI data sets for each subject were spatially smoothed (isotropic 6-mm kernel) to reduce variability among individuals and generate group maps. Next, the time-series data were normalized by dividing the signal intensity of a voxel at each time point by the mean signal intensity of that voxel for each run and multiplying the result by 100, producing regression coefficients representing percentage-signal change.

Following this, three regressors were generated: (a) harm/welfare-based transgressions, (b) social-conventional transgressions, and (c) neutral vignettes. A regressor of no interest was also included to model trials in which the participant did not respond. These regressors were created by convolving the train of stimulus events with a gamma-variate hemodynamic response function to account for the slow hemodynamic response. The participants' anatomical scans were individually registered to the Talairach and Tournoux atlas (Talairach & Tournoux, 1988). The individuals' functional EPI data were then registered to their Talairach anatomical scan within AFNI. Linear regression modeling was performed using the three regressors described above plus six head-motion regressors. This produced a β coefficient and associated t statistic for each voxel and regressor.

fMRI data analysis

In order to examine the common involvement of decision-making architecture during transgression evaluation, a conjunction analysis was conducted on regions showing activation during ratings of both harm/welfare-based and social-conventional transgressions. The conjunction analysis included all voxels significant at $p = .005$ in a harm/welfare-based transgression relative to fixation contrast and in a social-conventional transgression relative to fixation contrast. In order to examine BOLD activation involved with the processing

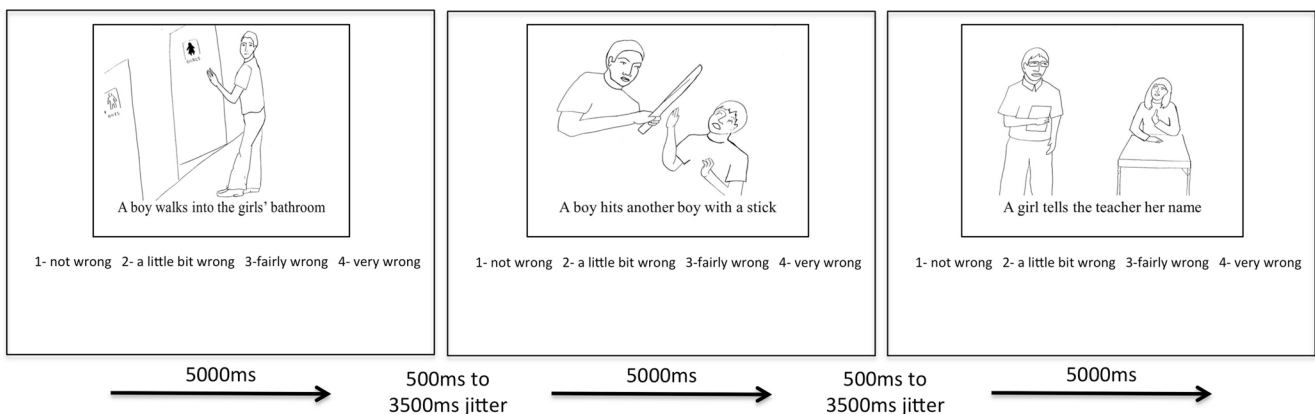


Fig. 1 The moral judgement task. Participants rated social-conventional transgressions (left panel), harm/welfare transgressions (middle panel), and neutral actions (right panel) as *not at all wrong*, *a little bit wrong*, *fairly wrong*, or *very wrong*

of transgressions, BOLD response to neutral vignettes was subtracted from both harm/welfare transgressions and social-conventional transgressions and entered into a one-way (transgression type: [harm/welfare-based transgression – neutral vignette], [social-conventional transgression – neutral vignette]) analysis of variance (ANOVA). The 3dClustSim program in AFNI, using the autocorrelation function (-acf) option, was used to establish an extent threshold correcting for multiple comparisons to $p = .05$ (initial threshold: $p = .005$, $k = 37$ voxels). All reported regions exceed this threshold except where noted. Post hoc analyses were performed to facilitate interpretations. For these analyses, average percentage signal change was measured across all voxels within each region of interest (ROI) generated from the functional masks, and data were analyzed using appropriate follow-up tests within SPSS.

Secondary analyses

In previous studies, harm/welfare-based transgressions were rated as more wrong than social-conventional transgressions (Nucci & Nucci, 1982a, b; Smetana et al., 2014; Turiel, 1983) and, by design, transgressions should be rated as more wrong than neutral scenarios. Thus, it was possible that any results from the main analysis would be driven by the level of wrongness of transgressions, as opposed to the content. If this were the case, modulating each individual transgression by level of wrongness would yield similar results to the main ANOVA.

In order to examine this possibility, two secondary analyses were conducted. In the first analysis, all transgressions were parametrically modulated by wrongness rating. In the second analysis, only social conventional items were modulated by level of wrongness to account for the fact that neutral scenarios and harm/welfare-based transgressions had only very limited variability in wrongness ratings between and across subjects. These analyses were conducted on 10 subjects drawn from the original analysis. Data from only 10 subjects were available due to a data storage failure during a move between institutions. Participants were right-handed adults (8 female), ages 22 to 26 years (mean = 24.32 years, $SD = 1.56$). The subsample of participants did not significantly differ from the original sample in terms of age ($t = .73$, $p = .47$) or in gender composition ($\chi^2 = .15$, $p = .70$).

Motion correction, alignment, and smoothing did not change in the secondary analyses. In the first secondary analysis, an indicator regressor and a parametrically modulated regressor was generated for each trial. BOLD response at each trial was modulated by the wrongness rating of each trial. In the second analysis, an indicator regressor and a parametrically modulated regressor was generated for each social-conventional transgression trial. BOLD response at each social-conventional trial was modulated by the wrongness rating of that social-conventional trial. All other trials were included in a regressor of no interest. Regressor convolution and

registration procedures did not change from the main analysis. Linear regression modeling was performed separately using the two sets of regressors described above plus six head-motion regressors. This produced a β coefficient and associated t statistic for each voxel and regressor, including indicator and parametrically modulated regressors for each analysis.

Results

Behavioral results

A one-way ANOVA was conducted on participant ratings of each vignette and the response latencies of each rating (see Table 1). Significant main effects of transgression type were observed in participant ratings, $F(2, 21) = 748.16$, $p < .01$. Participants rated harm/welfare-based transgressions as significantly more wrong than social-conventional transgressions ($t = 16.01$, $p < .01$) or neutral actions ($t = 52.13$, $p < .01$). Social-conventional transgressions were rated as significantly more wrong than neutral actions ($t = 18.60$, $p < .01$). Significant main effects of transgression type were also observed in response latencies, $F(2, 21) = 114.91$, $p < .001$. Response latencies were slower to social-conventional transgressions relative to harm/welfare-based transgressions ($t = 11.27$, $p < .01$) and neutral actions ($t = 23.07$, $p < .01$). Response latencies did not differ between harm/welfare-based transgressions and neutral actions ($t = .27$, $p = .79$).

fMRI results

Our first goal was to determine regions involved in processing both harm/welfare-based and social-conventional transgressions. This was assessed using a conjunction analysis and revealed regions showing common activation to both harm/welfare-based and social-conventional transgressions that included rostral/ventral medial frontal cortex, left aIC, and a large region including dmFC (see Table 2 and Fig. 2).

Our second goal was to determine regions showing differential responsiveness to harm-welfare/care-based versus social-conventional transgressions. A one-way (transgression type: [harm/welfare-based transgression – neutral vignette], [social-conventional transgression – neutral vignette]) ANOVA conducted on the BOLD response data revealed activations including bilateral inferior frontal gyrus, a large region encompassing rostral to dorsal anterior cingulate cortex that extended into left motor cortex, right TPJ, left precuneus/PCC, left superior temporal gyrus and bilateral temporal pole (see Table 3). In bilateral inferior frontal gyrus (iFG), greater BOLD response was observed to harm/welfare-based transgressions relative to social-conventional transgressions (Fig. 3). Due to its large size, the large region encompassing dmFC and left motor cortex was reexamined using an initial threshold of .0005. Local

Table 1 Ratings and response latencies for the moral judgment task

| | Rating | | Response latency | |
|------------------------------------|--------|-------|------------------|----------|
| | Mean | (SD) | Mean | (SD) |
| Social-conventional transgressions | 2.68 | (.44) | 2,657.28 ms | (387.70) |
| Harm/welfare-based transgressions | 3.75 | (.26) | 1,902.48 ms | (266.46) |
| Neutral actions | 1.04 | (.07) | 1,884.02 ms | (363.23) |

maxima were identified within rostral anterior cingulate cortex (ACC), dorsal ACC, and left middle insula/precentral gyrus. In each region, greater activation was observed to harm/welfare-based transgressions relative to social-conventional transgressions. In contrast, greater BOLD response was seen in social-conventional transgressions relative to harm/welfare-based transgression in TPJ, left precuneus/PCC, bilateral temporal pole, and superior temporal gyrus (see Fig. 4).

Secondary Results

The behavioral results in the secondary analysis sample were consistent with the analyses of the whole sample. Significant main effects of transgression type were observed in participant ratings, $F(2, 8) = 458.89, p < .01$. Participants rated harm/welfare-based transgressions as significantly more wrong than social-conventional transgressions ($t = 14.72, p < .01$) or neutral actions ($t = 44.18, p < .01$). Social-conventional transgressions were rated as significantly more wrong than neutral actions ($t = 12.83, p < .01$). Significant main effects of transgression type

were also observed in response latencies, $F(2, 8) = 62.34, p < .001$. Response latencies were slower to social-conventional transgressions relative to harm/welfare-based transgressions ($t = 7.74, p < .01$) and neutral actions ($t = 19.06, p < .01$). Response latencies did not differ between harm/welfare-based transgressions and neutral actions ($t = .30, p = .77$).

In the first follow-up analysis, all items modulated by wrongness ratings were examined relative to baseline. Activation modulated by wrongness rating that significantly differed from baseline was observed in regions including left temporal parietal junction (TPJ) and two regions of left middle temporal gyrus (see Table 4). Significant activation modulated by wrongness was also observed in posterior cingulate cortex/precuneus and right TPJ, albeit at levels below the extent threshold ($k = 21$ and 32 , respectively). All of these regions showed reduced activation as a function of wrongness rating.

In the second follow-up analysis, social-conventional items modulated by wrongness ratings only were examined relative to baseline. Significant activation was not observed in any regions showing differential response to either harm/welfare-based or social-conventional transgressions (see Table 5).

Table 2 Brain regions demonstrating common significant activation to harm/welfare-based transgressions and social-conventional transgressions

| Coordinates of peak activation ^b | | | | | | |
|---|------------|---------|-------|-------|-------|--------|
| Region ^a | Left/Right | BA | x | y | z | Voxels |
| Rostral/ventromedial prefrontal cortex | Right | 10 | 10.5 | 46.5 | 26.5 | 125 |
| Anterior insula cortex/inferior frontal gyrus | Left | 13 | -25.5 | 25.5 | 2.5 | 37 |
| Medial frontal cortex | | 32/9/24 | -34.5 | 16.5 | 23.5 | 846 |
| Dorsolateral prefrontal cortex | Left | 8 | -13.5 | 31.5 | 38.5 | 50 |
| Superior frontal gyrus | Right | 10 | 22.5 | 46.5 | 20.5 | 25 |
| Middle temporal gyrus | Left | 41 | -43.5 | -40.5 | 5.5 | 38 |
| Superior temporal gyrus/middle insula | Right | 21 | 34.5 | -4.5 | -9.5 | 559 |
| Inferior parietal cortex | Right | 40 | 46.5 | -46.5 | 26.5 | 147 |
| Inferior parietal cortex | Left | 39 | -40.5 | -58.5 | 26.5 | 131 |
| Parahippocampal gyrus | Left | 21 | -40.5 | -7.5 | -9.5 | 61 |
| Precuneus | Left | 7 | -19.5 | -64.5 | 26.5 | 153 |
| Cingulate gyrus | Right | 23 | 1.5 | 22.5 | 23.5 | 39 |
| Visual cortex | Left | | -34.5 | -31.5 | -21.5 | 1,161 |
| Middle occipital cortex | Right | 30 | 28.5 | -73.5 | 11.5 | 41 |
| Thalamus/brain stem | Right | | 1.5 | -22.5 | -15.5 | 188 |

^a According to the Talairach Daemon Atlas (<http://www.nitrc.org/projects/tal-daemon/>)

^b Based on the Tournoux and Talairach standard brain template; BA = Brodmann's area

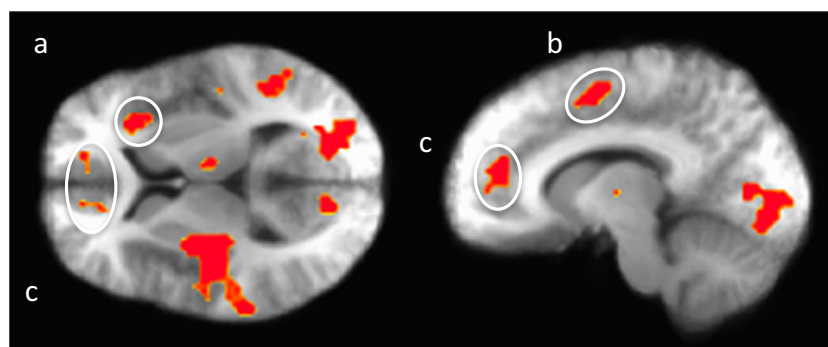


Fig. 2 Brain regions showing activation significantly greater than the neutral vignette baseline for both harm/welfare-based and social-conventional transgressions. Participants showed significant activation

in regions implicated in affect-based decision-making, including left anterior insula cortex/inferior frontal gyrus (a), dorsomedial prefrontal cortex (b), and ventromedial prefrontal cortex (c). (Color figure online)

Discussion

The current study had two goals. The first was to determine the neural regions involved in processing both harm/welfare-based and social-conventional transgressions. With respect to this goal, our study revealed that many of the regions involved in reinforcement-based decision-making, specifically rostral/ventral medial frontal cortex, dmFC, and left aIC/iFG, are responsive to both harm/welfare-based and social-conventional transgressions. Our second goal was to determine regions showing differential responsiveness to harm/welfare versus social-conventional transgressions. With respect to this goal, our study revealed significantly greater responses to harm/welfare-based transgressions relative to social-conventional transgressions in

regions within rostral and dorsal ACC, dmFC, and bilateral iFG, and significantly greater responses to social-conventional transgressions relative to harm/welfare-based transgressions in TPJ, left precuneus/PCC, bilateral temporal pole, and superior temporal gyrus.

Before considering the BOLD response data, it is worth briefly considering the behavioral data. Consistent with the previous literature (Nucci & Nucci, 1982a, b; Smetana et al., 2014; Turiel, 1983), participants generally regarded the harm/welfare-based and the social-conventional transgressions as wrong to do (while the neutral actions were acceptable). In addition, and consistent with previous work (Nucci & Nucci, 1982a, b; Smetana et al., 2014; Turiel, 1983), harm/welfare-based transgressions were judged significantly

Table 3 Brain regions demonstrating significant differences in BOLD response during ratings of harm/welfare-based transgressions and social-conventional transgressions

| Coordinates of peak activation ^b | | | | | | | | |
|---|------------|-------|----------|----------|----------|----------|----------|--------|
| Region ^a | Left/Right | BA | <i>x</i> | <i>y</i> | <i>z</i> | <i>F</i> | <i>p</i> | Voxels |
| Dorsomedial frontal and motor cortices | Left | | −31.5 | −1.5 | 14.5 | | | 3400 |
| Rostral anterior cingulate cortex* | Right | 32/10 | 1.5 | 37.5 | 11.5 | 30.86 | <.0001 | 45 |
| Dorsal anterior cingulate cortex* | Left | 31 | −7.5 | −16.5 | 44.5 | 44.83 | <.0001 | 234 |
| Middle insula/precentral gyrus* | Left | 2/13 | −31.5 | −1.5 | 14.5 | 100.0 | <.0001 | 1817 |
| Inferior frontal gyrus | Left | 46 | −37.5 | 31.5 | 14.5 | 21.84 | .0001 | 51 |
| Inferior frontal gyrus | Right | 46 | 46.5 | 31.5 | 11.5 | 24.64 | <.0001 | 40 |
| Temporal-parietal junction | Right | 39 | 37.5 | −73.5 | 32.5 | 29.04 | <.0001 | 152 |
| Temporal-parietal junction | Left | 39 | −52.5 | −61.5 | 26.5 | 22.05 | .0001 | 80 |
| Precuneus/posterior cingulate cortex | Left | 30 | −13.5 | −55.5 | 17.5 | 45.12 | <.0001 | 288 |
| Superior temporal gyrus | Right | 40/42 | 61.5 | −28.5 | 23.5 | 66.98 | <.0001 | 1081 |
| Temporal pole | Left | 22 | −46.5 | −10.5 | −6.5 | 20.69 | .0002 | 58 |
| Temporal pole | Right | 20/21 | 46.5 | −1.5 | −27.5 | 18.61 | .0003 | 42 |
| Middle temporal gyrus | Left | 21/22 | −49.5 | −34.5 | 2.5 | 20.16 | .0002 | 49 |
| Precentral gyrus | Right | 3 | 31.5 | −25.5 | 44.5 | 33.18 | <.0001 | 280 |
| Visual cortex | Right | 37 | 43.5 | −61.5 | 2.5 | 100.0 | <.0001 | 690 |
| Visual cortex | Left | 37 | −43.5 | −64.5 | −0.5 | 100.0 | <.0001 | 488 |

^a According to the Talairach Daemon Atlas (<http://www.nitrc.org/projects/tal-daemon/>)

^b Based on the Tournoux and Talairach standard brain template; BA = Brodmann's area

* Denotes local maxima at $p = .001$ within the larger region

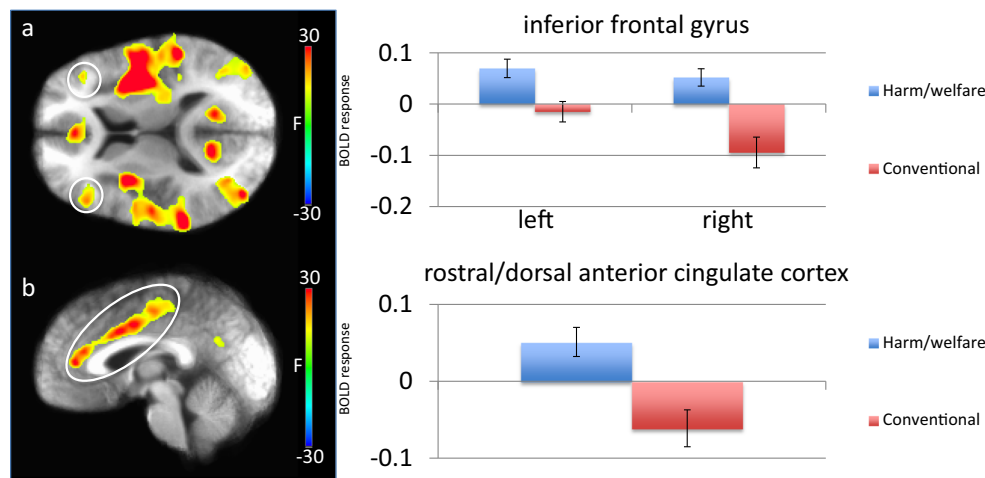


Fig. 3 Brain regions showing significantly greater activation when rating harm/welfare-based transgressions relative to social-conventional transgressions. Participants showed significantly greater activation when rating harm/welfare-based transgressions relative to social-conventional

transgressions within regions including bilateral inferior frontal gyrus (a) and within a large region including rostral and dorsal anterior cingulate and dorsomedial frontal cortices (b). (Color figure online)

more wrong than social-conventional transgressions. Also consistent with previous work (Lahat et al., 2013), response latencies were greater for judgments of social-conventional transgressions than for judgments of harm/welfare-based transgressions. It has been argued that slower response latencies for social-conventional transgressions may reflect a potential requirement for consideration of the context and societal rules when judging these transgressions (Lahat et al., 2013). Indeed, it has been argued on the basis of similar behavioral findings that harm/welfare-based judgments require fewer cognitive resources and may involve less deliberation, compared to judgments of social-conventional transgressions (Lahat et al., 2013). However, this argument must be qualified on the basis of the fMRI data considered below.

The first goal of this study was to determine the neural regions involved in processing both harm/welfare-based and social-conventional transgressions. Regions including vmPFC, dmFC, and left aIC/iFG are responsive to both harm/welfare-based and social-conventional transgressions. VmPFC has been widely implicated in moral reasoning (see Boccia et al., 2016). Within the reinforcement-based decision-making literature more generally, vmPFC is implicated in the representation of the subjective value of actions/objects available to be chosen (cf. Clithero & Rangel, 2014; O’Doherty, 2011). Given this, it has been argued that during moral reasoning, vmPFC represents the subjective value of the action to be judged (Blair, 2007). Consistent with this idea, vmPFC activation was positively associated with the number of lives saved when considering trolley problems (Shenhav & Greene, 2010). In addition, interesting recent work revealed that

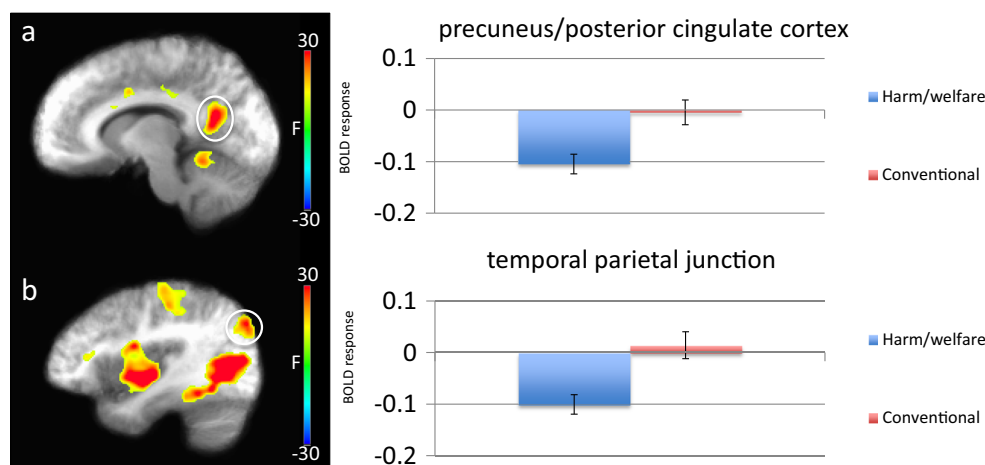


Fig. 4 Brain regions showing significantly greater activation when rating social-conventional transgressions relative to harm/welfare-based transgressions. Participants showed significantly greater activation when rating social-conventional transgressions relative harm/welfare-

based transgressions to within regions including precuneus/posterior cingulate cortex (a) and temporal-parietal junction (b). (Color figure online)

Table 4 Brain regions demonstrating significant levels of BOLD response modulated by level of punishment during ratings of all transgressions

| Coordinates of peak activation ^b | | | | | | | | |
|---|------------|-------|----------|----------|----------|----------|----------|--------|
| Region ^a | Left/Right | BA | <i>x</i> | <i>y</i> | <i>z</i> | <i>F</i> | <i>p</i> | Voxels |
| Regions showing decreased activation as a function of level of punishment | | | | | | | | |
| Temporal parietal junction | Left | 39/13 | −43.5 | −46.5 | 20.5 | 6.96 | <.0001 | 84 |
| Middle temporal gyrus | Left | 21 | −64.5 | −34.5 | −3.5 | 6.61 | <.0001 | 119 |
| Temporal pole | Left | 21 | −58.5 | −7.5 | −6.5 | 8.04 | <.0001 | 84 |
| Superior temporal gyrus/ middle insula/thalamus | Right | | 25.5 | −4.5 | 2.5 | 11.50 | <.0001 | 569 |
| Precentral gyrus | Right | 4 | 25.5 | −25.5 | 59.5 | 10.77 | <.0001 | 1,105 |
| Culmen | Left | | −16.5 | −46.5 | −15.5 | 9.37 | <.0001 | 281 |
| *Posterior cingulate cortex/precuneus | Right | 23 | 4.5 | −58.5 | 17.5 | 4.51 | .0015 | 21 |
| *Temporal parietal junction | Right | 39 | 43.5 | −67.5 | 26.5 | 5.21 | .0006 | 32 |
| Regions showing increased activation as a function of level of punishment | | | | | | | | |
| Middle/inferior temporal gyrus | Right | 37 | 43.5 | −55.5 | 2.5 | 7.54 | <.0001 | 165 |
| Precentral gyrus | Left | 3/4 | −34.5 | −28.5 | 47.5 | 10.92 | <.0001 | 389 |
| Culmen | Right | | 13.5 | −49.5 | −18.5 | 8.45 | <.0001 | 52 |

^a According to the Talairach Daemon Atlas (<http://www.nitrc.org/projects/tal-daemon/>)

^b Based on the Tournoux and Talairach standard brain template; BA = Brodmann's area

*Cluster below the extant threshold

a region of vmPFC proximal to that seen as responsive to both harm/welfare-based and social-conventional transgressions in the present study was shown to represent overall moral value (the study examined combination of emotional and utilitarian appraisals; Hutcherson et al., 2015). As such, we assume that this region was involved in the representation of the subjective value of the transgressions presented and that this occurred whether the transgressions were harm/welfare-based or social-conventional.

DmFC and aIC are also frequently implicated in reinforcement-based decision-making (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001; Kuhnen & Knutson, 2005; Liu et al., 2007). In the current study, there was significant activity

within these regions when participants responded to either harm/welfare-based or social-conventional transgressions relative to normative actions. These regions have been particularly implicated in avoidance behavior (Budhani, Marsh, Pine, & Blair, 2007; Casey et al., 2001; Kuhnen & Knutson, 2005; White et al., 2013; White et al., 2014a, 2016). These regions—particularly aIC and dmFC—show greater activity when suboptimal choices are about to be made and activity within these regions is modulated by expected value (Kuhnen & Knutson, 2005; White et al., 2013; White et al., 2014a; White et al., 2016). An alternative view regarding the function of these regions has suggested that they play a more attentional role and are implicated in initializing a task

Table 5 Brain regions demonstrating significant levels of BOLD response modulated by level of punishment during ratings of social-conventional transgressions

| Coordinates of peak activation ^b | | | | | | | | |
|---|------------|----|----------|----------|----------|----------|----------|--------|
| Region ^a | Left/Right | BA | <i>x</i> | <i>y</i> | <i>z</i> | <i>F</i> | <i>p</i> | Voxels |
| Regions showing decreased activation as a function of level of punishment | | | | | | | | |
| Postcentral gyrus | Left | 2 | −34.5 | −28.5 | 44.5 | 7.05 | <.0001 | 93 |
| Regions showing increased activation as a function of level of punishment | | | | | | | | |
| Postcentral gyrus | Right | 2 | 34.5 | −25.5 | 44.5 | 5.10 | .0007 | 53 |

^a According to the Talairach Daemon Atlas (<http://www.nitrc.org/projects/tal-daemon/>)

^b Based on the Tournoux and Talairach standard brain template; BA = Brodmann's area

set and maintaining it within a trial (Dosenbach et al., 2006; Power & Petersen, 2013; Sestieri, Corbetta, Spadone, Romani, & Shulman, 2014). It is argued that they play a role in performance reporting when adjustments are required by the task (Gratton et al., 2016). The current study was not designed to, and cannot distinguish, between these accounts. However, it is worth noting that there have been recent suggestions that the interactions of dmFC and aIC are differentiable according to the region of aIC involved (Droutman, Bechara, & Read, 2015). In particular, it has been suggested that interactions of dmFC with the more superior region of aIC seen in the current study typically underpin a more attentional role (a more inferior region of aIC is implicated in response control/avoidance; Droutman et al., 2015). Moral judgments are relatively complicated, as information both from the immediate stimulus, but also prior from semantic knowledge and emotional valence, may be recruited and processed before a judgment is made (cf. Ball et al., 2016; Turiel & Killen, 2010). Thus, it is conceivable that the activity seen here to both harm/welfare-based and social-conventional transgressions may represent some form of performance reporting given the requirement for adjustments in decision-making based on prior knowledge (cf. Gratton et al., 2016). Notably, however, these decision-making regions are active during affect-based decisions generally, not just during moral decision-making. As such, the data suggest less of a unitary moral faculty and rather a similarity between transgression wrongness judgments and other forms of affect-based decision-making.

In addition to determining neural regions involved in processing both harm/welfare-based and social-conventional transgressions, this study aimed to identify regions differentially responsive to harm/welfare-based transgressions as opposed to social-conventional transgressions. Notably, an extensive region of ACC, including both rostral and dorsal ACC as well as bilateral iFG, all showed greater responses when judging harm/welfare-based transgressions relative to social-conventional transgressions. These regions were proximal to, though not overlapping with, those regions responsive to both harm/welfare-based and social-conventional transgressions. In particular, the rostral portion of the activation within ACC was just posterior to the region of vmPFC seen as responsive to both harm/welfare-based transgressions and social-conventional transgressions. Similarly, the dorsal portion of the activation within ACC was inferior to the region of dmFC observed to both harm/welfare-based and social-conventional transgressions and the bilateral activations within IFG were anterior to the activations of aIC seen to both transgression types. This may reflect similar computational processes occurring for harm/welfare-based transgressions and social-conventional transgressions, potentially involving performance reporting/monitoring (cf. Droutman et al., 2015; Gratton et al., 2016), but that these occur more strongly for harm/welfare-based transgressions. Two caveats should be noted with respect to this conclusion, however. Performance reporting/monitoring has been related to

longer reaction times (Droutman et al., 2015; Gratton et al., 2016; Neta, Nelson, & Petersen, 2016). Yet reaction times for social-conventional transgressions were longer than those for harm/welfare-based transgressions. Moreover, the activation seen to harm/welfare-based transgressions relative to social-conventional transgressions recruited almost all of ACC (other than the subcallosal region). Thus, the current findings are likely indicative of a greater differentiation in functional processing between harm/welfare-based transgressions and social-conventional transgressions rather than simple differences in activation strength. Instead, it may reflect some of the emotion regulatory/processing roles of ACC (cf. Etkin, Egner, Peraza, Kandel, & Hirsch, 2006; Rudebeck, Buckley, Walton, & Rushworth, 2006). Certainly, other work has frequently observed amygdala responses to harm/welfare-based transgressions (e.g., Harenski et al., 2010; Luo et al., 2006)—though this was not observed here.

Several regions showed greater activation to social-conventional transgressions relative to harm/welfare-based transgression, specifically TPJ, left precuneus/PCC, bilateral temporal pole, and superior temporal gyrus. (While rostromedial frontal cortex did not show greater activity to social-conventional relative to harm/welfare-based transgressions, it did show strong activity to both forms of transgression). All these regions have been implicated in representing the mental states of others (ToM; Amodio & Frith, 2006; Happé & Frith, 2014; for a meta-analysis, see Molenberghs, Johnson, Henry, & Mattingley, 2016). Previous work has also reported that these regions are implicated in ToM respond to social-conventional transgressions (Berthoz et al., 2002; Finger et al., 2006). Of course, this does not preclude the relevance of mental states when processing harm/welfare-based transgressions (see Young et al., 2007). Intent information is critical for deciding whether an action that harms another *is* a transgression; accidental harm is not considered a transgression (see Young et al., 2007). However, the harm/welfare-based transgressions depicted in the current study were clearly intentional. The boy hitting the other boy intended to cause the other child harm. As such intent did not need to be calculated. Of course, the basic intent of the perpetrator of the social-conventional transgressions was similarly transparent (e.g., the boy intended to enter the girls' bathroom). But there are differences between these two forms of the transgressions in the social implications of these intents. An individual harming another is someone to avoid, because of the negative consequences of their actions for others' welfare. The aggressor is tagged with negative affect and is probably considered someone to avoid whatever their underlying intent (i.e., I might want to avoid both an individual who is hitting another to gain their resources and also an individual who is hitting another in retaliation for some slight). In contrast, an individual committing a social-conventional transgression is disrupting the social order or violating the standard hierarchical patterns of a

social group or social system. The intent of such an individual is critical to calculate—did they really mean to challenge the existing order or hierarchy, and might they affect *my* place in it? I might want to avoid the individual purposely challenging my status but be indifferent if his or her challenge has no implications for me. We assume that this importance of intent information for interpreting social-conventional transgressions underpins the greater activity in regions implicated in ToM when individuals processed social-conventional transgressions.

It is important to consider, however, that the harm/welfare-based transgressions chosen in this study elicited a relatively narrow range of wrongness responses (range: 2.85–4.00; mean = 3.75) relative to social-conventional transgressions (range: 1.75–3.69; mean = 2.68). This is consistent with past research that has shown that even minor moral transgressions are evaluated as more serious than major social-conventional offenses (Tisak & Turiel, 1988). This might lead to the concern that the results observed in this study are driven simply by the varying wrongness levels of the actions. If this were the case, we would expect that regions identified as differentiating harm/welfare-based and social-conventional transgressions would all show greater activity to harm/welfare-based transgressions. However, this was not seen. Furthermore, the secondary analyses of wrongness ratings correlated negatively with activation in regions associated with greater response to social-conventional transgressions only, including left TPJ, right superior temporal gyrus, and left temporal pole. Right PCC/precuneus and right TPJ showed the same pattern of results, but below the extent threshold ($k = 21$ and 32 , respectively). Notably, regions showing greater responsiveness to harm/welfare-based transgressions did not show modulation by level of judged wrongness of the item. In other words, activation in regions sensitive to harm/welfare-based violations cannot be considered to represent the general wrongness of the act. Finally, when only social-conventional transgressions modulated by wrongness ratings were considered, no significant modulated activation was observed in any regions associated with increased activation to during ratings of either harm/welfare-based or social-conventional transgressions. In short, we argue that the responding seen to social-conventional transgressions principally reflects greater processing of these items rather than greater processing of items judged less wrong (though the latter cannot be completely discounted).

It should be noted that the predicted greater amygdala and vmPFC activation when rating harm/welfare-based transgressions relative to social-conventional transgressions was not observed. Amygdala response to harm/welfare transgressions is often (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Harenski et al., 2010; Harenski & Hamann, 2006; Harenski, Harenski, Shane, & Kiehl, 2012; Luo et al., 2006; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002b; Moll et al., 2002a) though not always seen (Greene et al., 2001;

Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Shenhav & Greene, 2010; Young et al., 2007). Similarly, vmPFC findings are often (Harenski et al., 2010; Harenski et al., 2012; Heekeren et al., 2003; Luo et al., 2006; Moll et al., 2002b; Moll et al., 2002a; Shenhav & Greene, 2010; Young et al., 2007), but not always reported (Avram et al., 2013; Avram et al., 2014; Greene et al., 2001). With respect to the current study, we did see vmPFC involvement, albeit to both harm/welfare-based and social-conventional transgressions. As such, vmPFC may be highly dependent on the contrast made. Contrasting either harm/welfare-based or social-conventional transgressions against neutral actions was associated with vmPFC activity—even though the level of activity did not differ between these two transgression types.

Three caveats should be considered with respect to the current results. First, although the sample size ($N = 23$) was within the normative range for neuroimaging studies of healthy adults, it is relatively small, which could limit the generalizability of the findings. Second, as noted above, the harm/welfare-based transgressions chosen in this study elicited a relatively narrow range of wrongness responses relative to social-conventional transgressions. Future work might include very minor harm/welfare-based and more extreme social-conventional transgressions in the hope of generating more equivalent variability in wrongness responses. Third, it was possible to conduct the secondary analyses only on 10 participants. It is possible that the failure of the current study to obtain evidence that activation in regions sensitive to harm/welfare-based transgression represent the general wrongness of the act is due to Type II error. However it is important to note that modulation by wrongness level was observed in regions implicated in the processing of social-conventional transgressions.

In summary, the current results have implications for hypotheses regarding the degree to which “moral” transgressions are processed by a unitary system (or in a unitary way) or by dissociable systems in dissociable ways. Specifically, they suggest that there are a core set of regions that process social norms generally. These include vmPFC, dmFC, and left aIC/iFG (PCC and amygdala may also be implicated [cf. Boccia et al., 2016], though they were not seen in the current study). Notably, prior work has also shown their activation by disgust-based norms (Moll et al., 2005; Parkinson et al., 2011; Schaich Borg, Lieberman, & Kiehl, 2008; Schienle et al., 2002; Stark et al., 2007). All of these regions are implicated in valence-based decision-making (Clithero & Rangel, 2014; Kuhnen & Knutson, 2005). We believe that their activation here, and in previous work on disgust-based norms, reflects that a fundamental component of social norm judgment is the making of valence-based decisions, at least with respect to harm/welfare-based, social-conventional and disgust-based social norm judgments. But this is not to say that all social norms are processed in a unitary way. First, it is critical to

remember that the emotional learning experiences underpinning these different types of social norms differ. Data indicate the importance of appropriate emotional responding to distress cues for learning harm/welfare-based norms (Blair, Hwang, White, & Meffert, 2013), the importance of disgust reactions for learning disgust-based norms (Moll et al., 2005; Parkinson et al., 2011; Schaich Borg et al., 2008; Schienle et al., 2002; Stark et al., 2007), and it has been argued, at least, that responding to others' anger is critical for learning social-conventional norms (Blair, 2012). In addition, observational studies have shown that adults and children respond differently to naturally occurring harm/welfare-based and social-conventional rule violations (Nucci & Nucci, 1982a, b; Nucci & Turiel, 1978; Smetana, 1989), and these findings have been supported in experimental studies (Dahl, Sherlock, Campos, & Theunissen, 2014; Dahl & Tran, 2016). Second, the current data suggest the importance of representing intent information when processing social-conventional transgressions, perhaps because this information is particularly critical when considering potential future interactions with the transgressor. Third, our results suggest that harm/welfare-based transgressions also lead to extensive activation of cortical regions that are not seen in the response to social-conventional transgressions. In short, the findings from this study suggest a common process underpinning judgments of (at least harm/welfare-based, social-conventional, and disgust-based) social norms that reflects valence-based decision-making, but judgments of different norms are marked by differences in the forms of affect associated with their transgression and relative recruitment of specific computational processes.

Acknowledgements This work was supported by the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health under Grant No. 1-ZIA-MH002860-08 to R.J.R.B. Further support was provided by the National Institute of Mental Health, National Institutes of Health in grants to R.J.R.B. (1-K22-MH109558) and S.F.W. (1-K01-MH110643). Ethics approval for this study was granted by the NIH Combined Neuroscience Institutional Review Board under Protocol No. 93-M-0170.

References

- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews. Neuroscience*, 7(4), 268–277. doi:<https://doi.org/10.1038/nrn1884>
- Arsenio, W. F. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development*, 59(6), 1611–1622. doi:<https://doi.org/10.2307/1130675>
- Avram, M., Gutyrchik, E., Bao, Y., Pöppel, E., Reiser, M., & Blautzik, J. (2013). Neurofunctional correlates of esthetic and moral judgments. *Neuroscience Letters*, 534, 128–132. doi:<https://doi.org/10.1016/j.neulet.2012.11.053>
- Avram, M., Hennig-Fast, K., Bao, Y., Pöppel, E., Reiser, M., Blautzik, J., ... Gutyrchik, E. (2014). Neural correlates of moral judgments in first- and third-person perspectives: implications for neuroethics and beyond. *BMC Neuroscience*, 15, 39. doi:<https://doi.org/10.1186/1471-2202-15-39>
- Ball, C. L., Smetana, J. G., & Sturge-Apple, M. L. (2016). Following my head and my heart: Integrating preschoolers' empathy, theory of mind, and moral judgments. *Child Development*. doi:<https://doi.org/10.1111/cdev.12605>
- Berthoz, S., Armony, J. L., Blair, R. J. R., & Dolan, R. J. (2002). An fMRI study of intentional and unintentional (embarrassing) violations of social norms. *Brain*, 125(8), 1696–1708. doi:<https://doi.org/10.1093/brain/awf190>
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57(1), 1–29. 0010-0277(95)00676-P
- Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Sciences*, 11(9), 387–392. doi:<https://doi.org/10.1016/j.tics.2007.07.003>
- Blair, R. J. R. (2012). Considering anger from a cognitive neuroscience perspective. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(1), 65–74. doi:<https://doi.org/10.1002/wics.154>
- Blair, R. J. R., Hwang, S., White, S. F., & Meffert, H. (2013). Emotional learning mechanisms, differential moralities, reinforcement representations and the development of psychopathy. *Annals of the New York Academy of Sciences*, 1299, 36–41.
- Blair, R. J. R., Marsh, A. A., Finger, E., Blair, K. S., & Luo, J. (2006). Neuro-cognitive systems involved in morality. *Philosophical Explorations*, 9(1), 13–27. doi:<https://doi.org/10.1080/13869790500492359>
- Boccia, M., Dacquino, C., Piccardi, L., Cordellieri, P., Guariglia, C., Ferlazzo, F., ... Giannini, A. M. (2016). Neural foundation of human moral reasoning: An ALE meta-analysis about the role of personal perspective. *Brain Imaging and Behavior*. doi:<https://doi.org/10.1007/s11682-016-9505-x>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624–652.
- Budhani, S., Marsh, A. A., Pine, D. S., & Blair, R. J. (2007). Neural correlates of response reversal: Considering acquisition. *NeuroImage*, 34(4), 1754–1765. doi:<https://doi.org/10.1016/j.neuroimage.2006.08.060>
- Carr, A. R., Paholpak, P., Daianu, M., Fong, S. S., Mather, M., Jimenez, E. E., ... Mendez, M. F. (2015). An investigation of care-based vs. rule-based morality in frontotemporal dementia, Alzheimer's disease, and healthy controls. *Neuropsychologia*, 78, 73–79. doi:<https://doi.org/10.1016/j.neuropsychologia.2015.09.033>
- Casey, B. J., Forman, S. D., Franzen, P., Berkowitz, A., Braver, T. S., Nystrom, L. E., ... Noll, D. C. (2001). Sensitivity of prefrontal cortex to changes in target probability: A functional MRI study. *Human Brain Mapping*, 13(1), 26–33. doi:<https://doi.org/10.1002/hbm.1022>
- Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, 11(3), 476–484. doi:<https://doi.org/10.1093/scan/nsv131>
- Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social, Cognitive, and Affective Neuroscience*, 9(9), 1289–302. doi:<https://doi.org/10.1093/scan/nst106>
- Corradi-Dell'acqua, C., Civai, C., Rumiati, R. I., & Fink, G. R. (2012). Disentangling self- and fairness-related neural mechanisms involved in the ultimatum game: an fMRI study. *Social Cognitive and Affective Neuroscience*. doi:<https://doi.org/10.1093/scan/nss014>
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in*

- Biomedical Research*, 29(3), 162–173. doi:<https://doi.org/10.1006/cbmr.1996.0014>
- Cubero, I., Thiele, T. E., & Bernstein, I. L. (1999). Insular cortex lesions and taste aversion learning: Effects of conditioning method and timing of lesion. *Brain Research*, 839(2), 323–330.
- Dahl, A., Sherlock, B. R., Campos, J. J., & Theunissen, F. E. (2014). Mothers' tone of voice depends on the nature of infants' transgressions. *Emotion*, 14(4), 651–665. doi:<https://doi.org/10.1037/a0036608>
- Dahl, A., & Tran, A. Q. (2016). Vocal tones influence young children's responses to prohibitions. *Journal of Experimental Child Psychology*, 152, 71–91. doi:<https://doi.org/10.1016/j.jecp.2016.07.009>
- Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., ... Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, 50(5), 799–812. doi:<https://doi.org/10.1016/j.neuron.2006.04.031>
- Droutman, V., Bechara, A., & Read, S. J. (2015). Roles of the different sub-regions of the insular cortex in various phases of the decision-making process. *Frontiers in Behavioral Neuroscience*, 9, 309. doi:<https://doi.org/10.3389/fnbeh.2015.00309>
- Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R., & Hirsch, J. (2006). Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron*, 51(6), 871–882. doi:<https://doi.org/10.1016/j.neuron.2006.07.029>
- Finger, E. C., Marsh, A. A., Kamel, N., Mitchell, D. G., & Blair, J. R. (2006). Caught in the act: The impact of audience on the neural response to morally and socially inappropriate behavior. *NeuroImage*, 33(1), 414–421. doi:<https://doi.org/10.1016/j.neuroimage.2006.06.011>
- First, M. B., Spitzer, R. L., Gibbon, M., & Janet, B. W. (2002). Structured clinical interview for DSM-IV-TR axis I disorders (Research version, Patient ed.). New York, NY: Biometrics Research, New York State Psychiatric Institute.
- Fliessbach, K., Philippis, C. B., Trautner, P., Schnabel, M., Elger, C. E., Falk, A., & Weber, B. (2012). Neural responses to advantageous and disadvantageous inequity. *Frontiers in Human Neuroscience*, 6, 165. doi:<https://doi.org/10.3389/fnhum.2012.00165>
- Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science*, 286(5445), 1692–1695.
- Giner-Sorolla, R., Bosson, J. K., Caswell, T. A., & Hettinger, V. E. (2012). Emotions in sexual morality: Testing the separate elicitors of anger and disgust. *Cognition & Emotion*, 26(7), 1208–1222. doi:<https://doi.org/10.1080/02699931.2011.645278>
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–385. doi:<https://doi.org/10.1037/a0021847>
- Gratton, C., Neta, M., Sun, H., Ploran, E. J., Schlaggar, B. L., Wheeler, M. E., . . . Nelson, S. M. (2016). Distinct stages of moment-to-moment processing in the cinguloopercular and frontoparietal networks. *Cerebral Cortex*. doi:<https://doi.org/10.1093/cercor/bhw092>
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389–400. doi:<https://doi.org/10.1016/j.neuron.2004.09.027>
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. doi:<https://doi.org/10.1126/science.1062872>
- Güroğlu, B., Will, G.-J., & Crone, E. A. (2014). Neural correlates of advantageous and disadvantageous inequity in sharing decisions. *PLOS ONE*, 9(9), e107996. doi:<https://doi.org/10.1371/journal.pone.0107996>
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1), 19–19. doi:<https://doi.org/10.1007/s11211-007-0034-z>
- Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology & Psychiatry*, 55(6), 553–577. doi:<https://doi.org/10.1111/jcpp.12162>
- Harenski, C. L., Antonenko, O., Shane, M. S., & Kiehl, K. A. (2010). A functional imaging investigation of moral deliberation and moral intuition. *NeuroImage*, 49(3), 2707–2716. doi:<https://doi.org/10.1016/j.neuroimage.2009.10.062>
- Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, 30(1), 313–324. doi:<https://doi.org/10.1016/j.neuroimage.2005.09.034>
- Harenski, C. L., Harenski, K. A., Shane, M. S., & Kiehl, K. A. (2012). Neural development of mentalizing in moral judgment from adolescence to adulthood. *Developmental Cognitive Neuroscience*, 2(1), 162–173. doi:<https://doi.org/10.1016/j.dcn.2011.09.002>
- Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *NeuroReport*, 14(9), 1215–1219. doi:<https://doi.org/10.1097/01.wnr.0000081878.45938.a7>
- Hsu, M., Anen, C., & Quartz, S. R. (2008). The right and the good: Distributive justice and neural encoding of equity and efficiency. *Science*, 320(5879), 1092–1095. doi:<https://doi.org/10.1126/science.1153651>
- Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13(1), 1–6. doi:<https://doi.org/10.1016/j.tics.2008.09.006>
- Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(36), 12593–12605. doi:<https://doi.org/10.1523/JNEUROSCI.3402-14.2015>
- Jambon, M., & Smetana, J. G. (2014). Moral complexity in middle childhood: Children's evaluations of necessary harm. *Developmental Psychology*, 50(1), 22–33. doi:<https://doi.org/10.1037/a0032992>
- Jambon, M., & Smetana, J. G. (2017). Individual differences in prototypical moral and conventional judgments and children's proactive and reactive aggression. *Child Development*. doi:<https://doi.org/10.1111/cdev.12757>
- Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H. E., Lin, S. Y., . . . Shin, H. S. (2010). Observational fear learning involves affective pain system and Cav1.2 Ca²⁺ channels in ACC. *Nature Neuroscience*, 13(4), 482–488. doi:<https://doi.org/10.1038/nn.2504>
- Kagan, J., & Lamb, S. (1987). The emergence of morality in young children. Chicago, IL: University of Chicago Press.
- Killen, M., Mulvey, K. L., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, 119(2), 197–215. doi:<https://doi.org/10.1016/j.cognition.2011.01.006>
- Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral development and behaviour: Theory, research and social issues* (pp. 171–205). New York, NY: Holt, Rinehart & Winston.
- Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, 47(5), 763–770. doi:<https://doi.org/10.1016/j.neuron.2005.08.008>
- Lahat, A., Helwig, C. C., & Zelazo, P. D. (2013). An event-related potential study of adolescents' and young adults' judgments of moral and social conventional violations. *Child Development*, 84(3), 955–969. doi:<https://doi.org/10.1111/cdev.12001>
- Liu, X., Powell, D. K., Wang, H., Gold, B. T., Corbly, C. R., & Joseph, J. E. (2007). Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. *Journal of*

- Neuroscience*, 27(17), 4587–4597. doi:<https://doi.org/10.1523/JNEUROSCI.5227-06.2007>
- Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. (2006). The neural basis of implicit moral attitude—An IAT study using event-related fMRI. *NeuroImage*, 30(4), 1449–1457. doi:<https://doi.org/10.1016/j.neuroimage.2005.11.005>
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, 11(4), 143–152. doi:<https://doi.org/10.1016/j.tics.2006.12.007>
- Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, 65, 276–291. doi:<https://doi.org/10.1016/j.neubiorev.2016.03.020>
- Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002b). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, 16(3, Pt. 1), 696–703.
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002a). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22(7), 2730–2736.
- Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignacio, F. A., Bramati, I. E., Caparelli-Daquer, E. M., & Eslinger, P. J. (2005). The moral affiliations of disgust: A functional MRI study. *Cognitive and Behavioral Neurology*, 18(1), 68–78.
- Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive Affective and Behavioral Neuroscience*, 3(3), 207–233.
- Neta, M., Nelson, S. M., & Petersen, S. E. (2016). Dorsal anterior cingulate, medial superior frontal cortex, and anterior insula show performance reporting-related late task control signals. *Cerebral Cortex*. doi:<https://doi.org/10.1093/cercor/bhw053>
- Nucci, L. P., & Herman, S. (1982). Behavioral disordered children's conceptions of moral, conventional, and personal issues. *Journal of Abnormal Child Psychology*, 10(3), 411–425.
- Nucci, L. P., & Nucci, M. S. (1982a). Children's responses to moral and social conventional transgressions in free-play settings. *Child Development*, 53(5), 1337–1342. doi:<https://doi.org/10.2307/1129024>
- Nucci, L. P., & Nucci, M. S. (1982b). Children's social interactions in the context of moral and conventional transgressions. *Child Development*, 53(2), 403–412. doi:<https://doi.org/10.2307/1128983>
- Nucci, L. P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, 49(2), 400–407. doi:<https://doi.org/10.2307/1128704>
- O'Doherty, J. P. (2011). Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Annals of the New York Academy of Sciences*, 1239, 118–129. doi:<https://doi.org/10.1111/j.1749-6632.2011.06290.x>
- Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, 23(10), 3162–3180. doi:https://doi.org/10.1162/jocn_a_00017
- Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in Neurobiology*, 23(2), 223–228. doi:<https://doi.org/10.1016/j.conb.2012.12.009>
- Rilling, J. K., Goldsmith, D. R., Glenn, A. L., Jairam, M. R., Elfenbein, H. A., Dagenais, J. E., ... Pagnoni, G. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia*, 46(5), 1256–1266. doi:<https://doi.org/10.1016/j.neuropsychologia.2007.11.033>
- Rudebeck, P. H., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science*, 313(5791), 1310–1312. doi:<https://doi.org/10.1126/science.1128197>
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300(5626), 1755–1758. doi:<https://doi.org/10.1126/science.1082976>
- Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, 18(5), 803–817. doi:<https://doi.org/10.1162/jocn.2006.18.5.803>
- Schaich Borg, J., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience*, 20(9), 1529–1546. doi:<https://doi.org/10.1162/jocn.2008.20109>
- Schienle, A., Stark, R., Walter, B., Blecker, C., Ott, U., Kirsch, P., ... Vaitl, D. (2002). The insula is not specifically involved in disgust processing: An fMRI study. *NeuroReport*, 13(16), 2023–2026.
- Sestieri, C., Corbetta, M., Spadone, S., Romani, G. L., & Shulman, G. L. (2014). Domain-general signals in the cingulo-opercular network for visuospatial attention and episodic memory. *Journal of Cognitive Neuroscience*, 26(3), 551–568. doi:https://doi.org/10.1162/jocn_a_00504
- Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, 67(4), 667–677. doi:<https://doi.org/10.1016/j.neuron.2010.07.020>
- Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–83). Chicago, IL: University of Chicago Press.
- Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, 52(4), 1333–1336. doi:<https://doi.org/10.2307/1129527>
- Smetana, J. G. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology*, 21(1), 18–29. doi:<https://doi.org/10.1037/0012-1649.21.1.18>
- Smetana, J. G. (1989). Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology*, 25(4), 499–508. doi:<https://doi.org/10.1037/0012-1649.25.4.499>
- Smetana, J. G., & Braeges, J. L. (1990). The development of toddler's moral and conventional judgments. *Merrill-Palmer Quarterly*, 36(3), 329–346.
- Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen, J. G. Smetana, M. Killen, & J. G. Smetana (Eds.), *Handbook of moral development* (2nd ed., pp. 23–45). New York, NY: Psychology Press.
- Smetana, J. G., Rote, W. M., Jambon, M., Tasopoulos-Chan, M., Villalobos, M., & Comer, J. (2012). Developmental changes and individual differences in young children's moral judgments. *Child Development*, 83(2), 683–696.
- Stark, R., Zimmermann, M., Kagerer, S., Schienle, A., Walter, B., Weygandt, M., & Vaitl, D. (2007). Hemodynamic brain correlates of disgust and fear ratings. *NeuroImage*, 37(2), 663–673. doi:<https://doi.org/10.1016/j.neuroimage.2007.05.005>
- Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *NeuroImage*, 54(1), 671–80. doi:<https://doi.org/10.1016/j.neuroimage.2010.07.051>
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, 19(4), 339–347. doi:<https://doi.org/10.1111/j.1467-9280.2008.02091.x>

- Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. Stuttgart, Germany: Thieme.
- Tisak, M. S., & Turiel, E. (1988). Variation in seriousness of transgressions and children's moral and conventional concepts. *Developmental Psychology*, *24*(3), 352–357. doi:<https://doi.org/10.1037/0012-1649.24.3.352>
- Turiel, E. (1983). The development of social knowledge: Morality and social convention. Cambridge, UK: Cambridge University Press.
- Turiel, E., & Killen, M. (2010). Taking emotions seriously: The role of emotions in moral development. In W. F. Arsenio, E. A. Lemerise, W. F. Arsenio, & E. A. Lemerise (Eds.), *Emotions, aggression, and morality in children: Bridging development and psychopathology* (pp. 33–52). Washington, DC: American Psychological Association.
- White, S. F., Brislin, S. J., Sinclair, S., & Blair, J. R. (2014b). Punishing unfairness: Rewarding or the organization of a reactively aggressive response? *Human Brain Mapping*, *35*(5), 2137–47. doi:<https://doi.org/10.1002/hbm.22316>
- White, S. F., Fowler, K. A., Sinclair, S., Schechter, J. C., Majestic, C. M., Pine, D. S., & Blair, R. J. (2014a). Disrupted expected value signaling in youth with disruptive behavior disorders to environmental reinforcers. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(5), 579–588. doi:<https://doi.org/10.1016/j.jaac.2013.12.023>
- White, S. F., Pope, K., Sinclair, S., Fowler, K. A., Brislin, S. J., Williams, W. C., ... Blair, R. J. (2013). Disrupted expected value and prediction error signaling in youths with disruptive behavior disorders during a passive avoidance task. *American Journal of Psychiatry*, *170*(3), 315–323. doi:<https://doi.org/10.1176/appi.ajp.2012.12060840>
- White, S. F., Tyler, P. M., Erway, A. K., Botkin, M. L., Kolli, V., Meffert, H., ... Blair, J. R. (2016). Dysfunctional representation of expected value is associated with reinforcement-based decision-making deficits in adolescents with conduct problems. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. doi:<https://doi.org/10.1111/jcpp.12557>
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(20), 8235–8240. doi:<https://doi.org/10.1073/pnas.0701408104>