Cogn Affect Behav Neurosci (2014) 14:266–277
DOI 10.3758/s13415-013-0213-3

Individual differences in the anterior insula are associated with the likelihood of financially helping versus harming others

Steven Greening · Loretta Norton · Karim Virani · Ambrose Ty · Derek Mitchell · Elizabeth Finger

Published online: 5 October 2013

Abstract The neural basis of individual differences in positive and negative social decisions and behaviors in healthy populations is yet undetermined. Recent work has focused on the potential role of the anterior insula in guiding social and nonsocial decision making, but the specific nature of its activation during such decision making remains unclear. To identify the neural regions mediating individual differences in helpful and harmful decisions and to assess the nature of insula activation during such decisions, in the present study we used a novel fMRI task featuring intentional and unintentional decisions to financially harm or help persons in need. Based on a whole-brain, unbiased approach, our findings indicate that individual differences in dorsal anterior insula, anterior cingulate cortex (ACC), and right temporo-parietal junction activation are associated with behavioral tendencies to financially harm or help another. Furthermore, activity in the dorsal anterior insula and ACC was greatest during unintended outcomes, whether these were gains or losses for a charity or for oneself, supporting models of the role of these regions in salience prediction error signaling. Together, the results suggest that individual differences in risk anticipation, as reflected in the dorsal anterior insula and dorsal ACC, guide social decisions to refrain from harming others.

Keywords Insula · Guilt · Moral emotions · Prosocial

Electronic supplementary material The online version of this article (doi:10.3758/s13415-013-0213-3) contains supplementary material, which is available to authorized users.

S. Greening · L. Norton · K. Virani · A. Ty · D. Mitchell · E. Finger
University of Western Ontario, London, Ontario, Canada

E. Finger (✉)
University Hospital, University of Western Ontario, 301
Winderemere Road, London, Ontario N6A 5A5, Canada
e-mail: Elizabeth.Finger@lhsc.on.ca

Our daily news demonstrates the heroic and horrific extremes of human prosocial and antisocial behavior. Although case reports of acquired antisocial behaviors have indicated roles for the orbitofrontal cortex, ventromedial prefrontal cortex, amygdala, and insula in guiding emotional responses and behavior (Bar-On, Tranel, Denburg, & Bechara, 2003; Eslinger & Damasio, 1985; Grafman et al., 1996), the neural correlates of individual differences in positive and negative social tendencies in general populations are not yet understood. Candidate neural regions that may mediate individual differences in these tendencies in healthy populations can be identified from fMRI studies modeling moral transgressions and guilt. The moral emotion of guilt is considered to be a critical mediator for prosocial decision making, and its absence is associated with antisocial behaviors (Hare, 1970; Harenski, Harenski, Shane, & Kiehl, 2010). In healthy adults, fMRI studies attempting to model the neural processing of moral transgressions have used vignettes featuring guilt- or embarrassment-inducing situations (Finger, Marsh, Kamel, Mitchell, & Blair, 2006; Takahashi et al., 2004) or mentalizing harm to another from visual stimuli (Decety & Porges, 2011). These methods have suggested that processing such social behaviors involves neural regions implicated in theory-of-mind processing, including temporo-parietal junction (TPJ) and medial prefrontal cortex (PFC; Decety & Porges, 2011; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Heekeren et al., 2005; Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Moll, de Oliveira-Souza, Bramati, & Grafman, 2002; Moll, de Oliveira-Souza, Eslinger, et al., 2002) and regions of frontal cortex including dorsomedial and ventrolateral PFC implicated in response change (Finger et al., 2006).

Of interest, several of the regions identified in the moral-transgression paradigms have subsequently been associated with individual differences in economic decision-making or charitable-donation paradigms. Recent work has suggested that individual differences in some forms of cooperative economic decision making may be mediated by the insula, TPJ, orbitofrontal cortex, and dorsomedial PFC. In a novel paradigm featuring contributions to charitable organizations, decisions not to donate to a charity were associated with increased signal in inferior orbitofrontal cortex (BA 47/11; Moll et al., 2006). During cooperative economic decision-making paradigms, increased activity in the anterior insula, dorsomedial PFC extending to ACC, and TPJ has been associated with guilt aversion and avoidance of inequitable decisions (Chang, Smith, Dufwenberg, & Sanfey, 2011; Waytz, Zaki, & Mitchell, 2012; Zaki & Mitchell, 2012), whereas generous decision making has been associated with dorsomedial PFC activity (Waytz et al., 2012). In this context, activity in the TPJ and dorsomedial PFC may reflect an aspect of theory of mind processing. Specifically, TPJ is recruited during various types of ToM tasks and may be a key node in the detection of agency and initial formation of attributions of mental states (Abu-Akel, 2003; Samson, Apperly, Chiavarino, & Humphreys, 2004). The nature of anterior insula activity in these contexts is less clear. Although the anterior insula has been shown to be active during situations that may produce guilt aversion or the anticipation of guilt (Chang et al., 2011; Shin et al., 2000), activation has not been consistently demonstrated in other script-based paradigms designed to elicit feelings of guilt (Finger et al., 2006; Takahashi et al., 2004). These discrepancies raise the question of the specific functional nature of anterior insula activation related to decisions and actions that may produce harm to others and feelings of guilt. The anterior insula has been associated with a wide range of affective cognitive functions, including empathy, guilt, disgust, moral judgment, interoception, risk aversion, and goal-directed cognition (Chang, Yarkoni, Khaw & Sanfey, 2013; Chapman & Anderson, 2012; Craig, 2009; Duncan & Owen, 2000; Rudorf, Preuschoff, & Weber, 2012; Singer et al., 2004). It has been proposed that the role of the anterior insula in conjunction with the ACC in social and nonsocial decision making is broader than the processing of guilt or empathy and reflects the guiding of decisions on the basis of risk calculation and avoidance of risk for the benefit of the individual (Bossaerts, 2010).

In this study, we aimed first to examine the neural basis of individual differences in helpful and harmful financial decisions in healthy populations, and second to clarify the roles of the insula and other neural regions involved in decisions around moral transgressions. We designed a novel fMRI task that required participants to make real-time decisions to assign monetary gains or losses to themselves or a specific charity, and thereby inflict a financial harm on someone in need. The use of financial harm was selected given the difficulty of ecologically modeling other types of interpersonal harm within the constraints of fMRI. To address the first aim, we assessed correlations between participants' trait ratings of empathic tendencies and guilt proneness, behavioral choices during the charity task, and BOLD signal during helpful or harmful decisions. The decision and outcome phases were modeled separately to identify the neural regions that confer individual differences in social decision making. To address the second aim, to further delineate the nature of neural activations during helpful and harmful decisions, particularly in the anterior insula, we used probabilistic outcomes to evaluate the effects of intention to harm or help another. Personal responsibility for a moral transgression is strongly correlated with the degree of guilt experienced (Kubany & Watson, 2003; McGraw, 1987). Furthermore, attributions of blame and personal responsibility for an outcome can be manipulated by introducing probabilistic outcomes (Coricelli & Rustichini, 2010; Lagnado & Channon, 2008; Shaver, 1985). By manipulating the participant's control over the outcome, we aimed to vary the degrees of personal responsibility and guilt generated by the two outcome scenarios. With this design, we were able to examine two possibilities related to the role of the anterior insula in this form of decision making. First, if anterior insula activation is involved in the representation of guilt aversion, we anticipated that insula activation would be increased for losses assigned to charities as compared to losses assigned to the self in the choice stage. If insula activation also reflects the experience of guilt, activation would be (1) increased during feedback for losses to the charity, and (2) reduced when personal responsibility for harm was reduced (i.e., for probabilistic feedback when a decision resulted in an unintentional loss for the charity as compared to a loss intentionally assigned to the charity). In contrast, if the anterior insula activation during moral transgressions more broadly represented risk aversion and harm, activations would occur during decisions related to harm to self *and* others, and in the feedback phase would not be mitigated by manipulations of intention/personal responsibility.

Method

Subjects

Eighteen healthy participants (nine male, nine female) were recruited via fliers placed on the campus of the University of Western Ontario. All of the subjects granted informed consent and had no history of neurological or psychiatric conditions, as determined by screening. All of the subjects had normal or corrected-to-normal vision and were right-handed. The participants' consent was obtained according to the Declaration of Helsinki (Lynöe, Sandlund, Dahlqvist, & Jacobsson, 1991), and the study was approved by the Health Sciences Research Ethics Board at the University of Western Ontario, London, Ontario, Canada. Whereas the full 18 participants completed the fMRI scan and task as described below, only 16 of the participants were available to complete trait measures of guilt and psychopathy. Thus, all individual differences analyses were performed on the subsample of 16 participants.

Experimental task

The "charity task" is designed to elicit social decisions followed by intended or unintended outcomes by presenting participants with trials in which they must decide whether (1) to reward either themselves or a charity with bonus dollars or (2) to penalize themselves or the charity by taking away bonus dollars. At the beginning of the study, participants were told that they would have the opportunity to earn bonus dollars for their participation in the study, and that they would be given the opportunity to give some of the bonus to a charity the lab is working with (see the supplemental materials for the full instructions). Participants were told that the experiment would be evaluating the effectiveness and appeal of different charities, and that they should make whatever responses came naturally to them. Participants were also informed that they would receive a proportion of the dollars that they chose not to allocate to the charity over the course of the task. Finally, the probabilistic nature of the feedback was explained. Specifically, participants were told that for two out of three trials, the result would be what they intended—that is, if on a gain trial, the designated recipient would receive the gain on two trials, but would receive a loss on one out of three trials. Similarly, if the trial was a loss trial, in the feedback stage two out of three times the outcome would be a loss for the selected recipient, whereas one out of three times the feedback would be a gain for the recipient. Immediately prior to being scanned, participants viewed a website and informational slide show about a fictional charity to aid earthquake victims. All pictures in the slide show were taken from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008). Participants then underwent 3-T fMRI while performing the charity task depicted in Fig. 1. Each trial involved a decision phase followed by an outcome phase. The decision phase began with a choice interval, during which participants were presented with a dollar amount that indicated that the trial was a gain or loss trial. They then chose to assign the gain ($2 or $5) or loss (−$2 or –$5) to the charity or themselves, with each of the four dollar amounts occurring the same number of times in a random order. Participants made the selection of their own name or the charity name by performing a left (index finger) or right (middle finger) buttonpress. Each choice was presented on the right or left side of the screen with equal probability (50 %), at random. The decision phase ended with a response interval that depicted the choice that the participant had made by placing a yellow square around either the charity or the participant's name. The duration of the decision phase was varied (duration of 2, 4, or 6 s) to allow for dissociation of the decision phase from the outcome phase.

During the outcome phase, we assessed the role of intention by using probabilistic outcomes. The outcome phase was a single interval that provided feedback to the participants about their decision (e.g., "You gave $5.00 to Earthquake Victims"), which lasted 2 s. On one-third of the trials, the valence of the monetary value was switched unexpectedly at

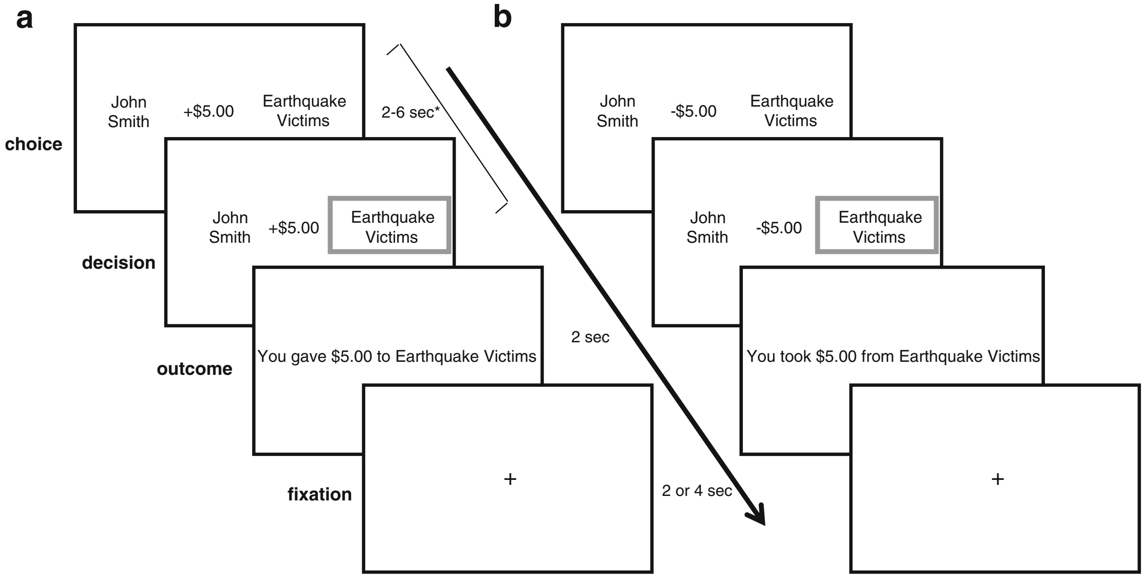

Fig. 1 Charity fMRI task: Sample trial screens for (**a**) a gain trial and (**b**) a loss trial resulting in the intended outcome. The boxes in the decision phase indicate the selection made by the participant via buttonpress. Participants completed six runs of the task, lasting 6 min and 16 s each, with 48 trials per run. *Durations of the choice and decision screens were randomly varied among 2, 4, and 6 s for each trial to allow for dissociation of the BOLD signal of this epoch from subsequent feedback. Similarly, the final fixation was presented for 2 or 4 s to permit dissociation of BOLD responses from the feedback with those from the subsequent choice trial

the outcome phase. This meant that although the participant may have chosen to give $5 to the charity, he or she could have unintentionally taken $5 from the charity (an unintended-outcome trial). Since differential processes were likely involved in the decision and outcome phases of each trial, we modeled these phases independently for the purposes of the multiple regression analysis of the fMRI data.

Each trial either ended with a fixation cross of variable duration (2 or 4 s) or ended following the outcome phase. The use of a variable stimulus interval allowed for the independent analysis of the decision and outcome intervals. The experimental task consisted of a total of 288 trials: 144 trials with likely gain, with one third of the trials (48 trials) resulting in unintentional loss of money; and 144 trials with likely loss, with one third of the trials resulting in an unintentional gain of money. Following completion of the study procedures, participants were debriefed and reconsented with a letter that revealed the fictitious nature of the charity and that no actual dollars were sent to the charity. Following the debriefing, all participants were informed that they would receive a $15 “bonus” and that this same amount was given to each task participant.

Trait measures

In a separate session following the MRI scan, participants completed the Test of Self-Conscious Affect (TOSCA; Tangney, Wagner, & Gramzow, 1989) and the Psychopathic Traits Personality Inventory–Revised (PPI; Lilienfeld & Widows, 2005). The TOSCA is a validated questionnaire featuring 15 scenarios (ten negative and five positive) that can be encountered in daily life and yields indices of shame, guilt, externalization, detachment/unconcern, alpha pride, and beta pride. Shame, guilt, and externalization are represented in all 15 scenarios, detachment in the ten negative scenarios, and alpha pride in the five positive scenarios.

The PPI (Lilienfeld & Widows, 2005) is a 154-item self-report personality measure designed to examine psychopathic traits in nonincarcerated populations. The overall score represents global psychopathy on the basis of seven subscales: impulsive nonconformity, blame externalization, Machiavellian egocentricity (30 items), carefree nonplanfulness, stress immunity, social potency, fearlessness, and cold-heartedness. Prior studies have shown that the PPI correlates with psychophysiological measures of empathy (Harenski, Kim, & Hamann, 2009), including fMRI BOLD signal (Han, Alders, Greening, Neufeld, & Mitchell, 2012).

MRI data acquisition

Participants were scanned while completing the task in a 3-T Siemens MRI scanner with a 32-channel head coil. Task-related blood oxygenation level dependent (BOLD) activity was measured throughout all six runs using a T2*-gradient echoplanar imaging sequence (repetition time = 2,000 ms; echo time = 30 ms; 240 × 240 mm field of view). Whole-brain coverage was obtained with 33 interleaved slices (3-mm thickness, 3 × 3 mm in-plane resolution) with anterior to posterior phase encoding. Each scan session ended with a high-resolution T1 anatomical image (T.R. = 2,300 ms; T.E. = 4.25 ms; F.o.V. = 256 mm by 256 mm; 192 slices) comprising 1-mm isovoxels.

fMRI analysis

Individual and group analyses were performed using the Analysis of Functional NeuroImages software (Cox, 1996). The first six volumes of each run were discarded to ensure that magnetization equilibrium was reached. To correct for subject motion, every volume of each run was registered to the last volume of Run 6, which immediately preceded the anatomical image acquisition. All data were spatially smoothed using a 4-mm isotropic Gaussian kernel. The time series of each voxel was scaled such that each time point within a voxel was represented as a percent change from the mean voxel intensity. Each voxel time series was then regressed against models of each of our conditions of interest, as well as against a model for baseline plus linear drift and quadratic trend. Thus, the resulting regression coefficients represented the percent signal change from mean voxel activity.

Separate regressor models were created for each condition of interest in each of the decision and outcome phases. The decision phase comprised eight models: high (−$5) loss to charity, low (−$2) loss to charity, high gain to charity, low gain to charity, high loss to self, low loss to self, high gain to self, and low loss to self. The outcome phase contained similarly named models with the additional level of intention (intended and unintended outcomes). This resulted in 16 regressors for the outcome phase.

Analyses of variance (ANOVAs) were then conducted in the choice and outcome phases as described below. In order to identify the brain regions encoding individual differences in brain–behavior correlations, we entered participants’ response frequencies as continuous covariates into two planned whole-brain contrasts at the group level of analysis of the BOLD signal during the decision phase (see the Results), using AFNI’s 3dttest++. This approach identifies clusters with a significant linear relationship between the continuous covariate and the BOLD signal (first during decisions to assign a harm to the charity, and second during decisions to assign a gain to the charity), which indicates that these two variables are correlated. This was followed by an exploratory whole-brain analysis. Clusters were thresholded at $p < .005$ and corrected for multiple comparisons to $p < .05$ using AFNI’s 3dClustSim, with 10,000 Monte Carlo simulations, to protect against Type I errors.

Results

Trait scales

The analysis of participants' scores on the trait measures revealed that the mean TOSCA guilt score was 61.1 (*SD* 9.2, range 38–72). On the measure of psychopathic traits, the mean PPI total score was 273.7 (*SD* 34.8, range 191–330). No other subscales of either measure were included in the analysis.

Behavioral results

We conducted a 2 (valence: gain vs. loss) × 2 (choice: self vs. charity) × 2 (magnitude: \$2 vs. \$5) ANOVA on the frequencies of decisions in the decision phase. This demonstrated a trend toward a Valence × Choice interaction [$F(1, 15) = 3.3$, $p = .09$ two-tailed], showing that participants were slightly more likely to select themselves for a loss trial (mean charity-loss choices = 61 trials [44 %] vs. self-loss choices = 79 trials [56 %]), and the charity for a gain trial (mean charity-gain choices = 87 [62 %] vs. self gain choices = 53 [38 %]) (Table 1). We found no significant main effects or interactions. We then conducted correlation analyses to test the hypothesis that trait guilt scores as indexed by the TOSCA and psychopathic traits as measured by the PPI would be correlated with decision frequencies during the task (Fig. 2). In line with our predictions, these demonstrated a significant positive correlation between TOSCA scores and charity-gain decisions (for high-magnitude trials) ($r = .62$, $p < .05$); that is, higher guilt traits were associated with more allocations of gain for charity. A significant negative correlation was observed between PPI scores and the decision to help the charity ($r = -.51$, $p < .05$; i.e., higher psychopathic traits were associated with fewer high-magnitude gains given to charity), and a trend toward a positive correlation between psychopathic traits and decisions to harm the charity (i.e., high-magnitude losses assigned to the charity; $r = .45$, $p = .08$).

Table 1 Mean decision frequencies during the charity fMRI task for gain and loss choices for charity and self

	Magnitude	Choice	
		Charity	Self
Gain	High (\$5)	46.9 (16.5)	23.2 (17.1)
	Low (\$2)	40.2 (16.7)	30.1 (16.2)
Loss	High (\$5)	30.1 (18.5)	40.3 (17.5)
	Low (\$2)	31.3 (18.7)	38.8 (18.0)

The data are mean numbers of decisions, with standard deviations in parentheses, for each trial type

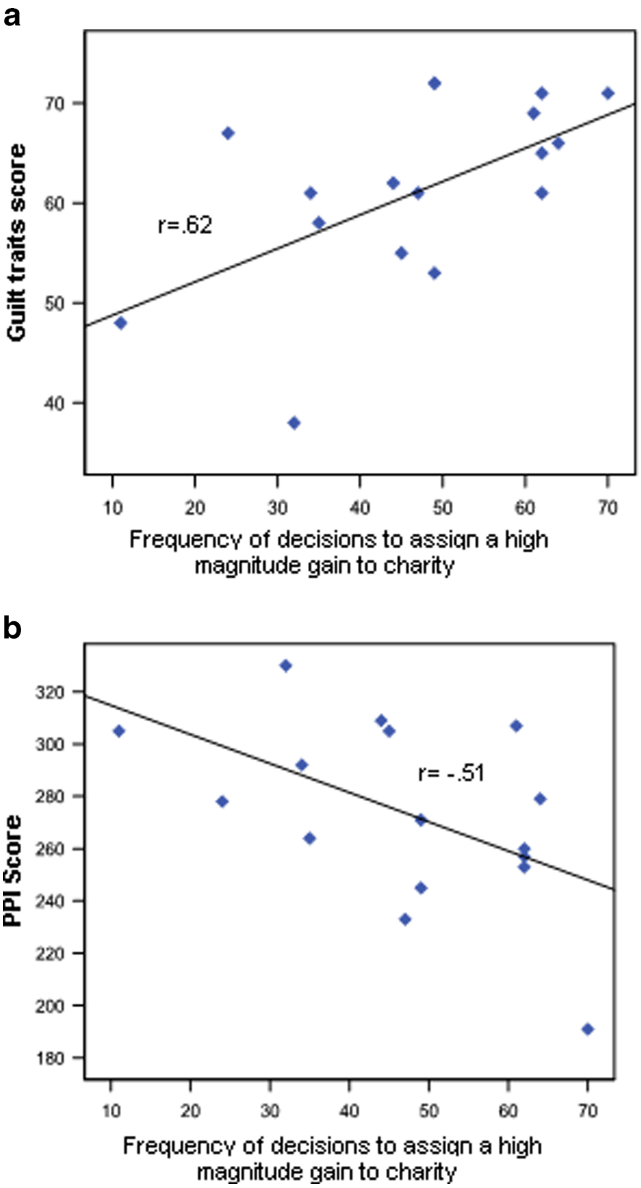

Fig. 2 Correlations between personality traits and behavioral choices. (*Top*) A significant positive correlation was observed between self-reported tendencies to experience the emotion of guilt, as rated by the Test of Self-Conscious Affect, and decisions to help the charity by assigning high-magnitude gains ($p < .05$). (*Bottom*) A significant negative correlation was observed between self-reported psychopathic traits on the Psychopathic Traits Personality Inventory–Revised and decisions to help the charity by assigning high-magnitude gains ($p < .05$)

Imaging individual differences during decision phase

To identify neural regions that might convey individual differences in social decision making, we examined the potential relationship between neural activation during decisions to harm or help the charity and the frequency of decisions to harm or help the charity.

This individual-differences analysis was conducted using a whole-brain, voxel-wise approach with a one-sample t test during the decision phase when a loss was assigned to the charity (decision phase BOLD response vs. baseline BOLD), in which the baseline, as modeled in the AFNI, consisted of the mean signal intensity of that voxel for each run. The number of decisions to harm the charity (to assign it a loss) was included as a continuous covariate.

This demonstrated significant negative correlations between the number of charity-loss decisions (to harm the charity) and BOLD signal in the right (33, 18, 6) and left (−42, 15, 12) anterior insula ($p < .001$) and anterior cingulate cortex (2, 29, 32) (Fig. 3, Table 2). Participants who demonstrated greater insula and ACC activation during charity-loss decisions assigned fewer losses to the charity throughout the task. To ensure that this correlation was not driven by habituation to guilt-inducing trials, we created a regressor consisting of the first five trials on which the participants assigned a loss to the charity, so that the event numbers included in the analysis were equal across all participants. We then plotted the percentages of signal change in the same insula functional regions of interest (ROIs) used in the initial correlation, by using a median split and t test to compare participants who made more charity-loss assignments with those that made fewer. We selected five trials, as this allowed for the inclusion of all participants (the maximum number of charity-loss decisions made by one participant was five). Since no a priori cutoffs designated an absolute number characterizing low versus high charity-loss decisions, a median split was then conducted to divide the group into low charity-loss decisions (mean charity-loss decisions = 35 (*SD* 21) and high charity-loss decisions (mean = 87, *SD* 21). This demonstrated that even on the first five trials alone, greater activation in the anterior insula was present in participants who were less likely to assign the charity a loss during the task [$t(14) = 2.2$, $p < .05$].

A parallel whole-brain analysis using the same procedures described above for charity-loss decisions was next conducted on the BOLD signal for neural responses during the decision phase for charity-gain decisions (decisions to help the charity by assigning a gain), using the number of charity-gain decisions as the continuous covariate. This demonstrated a significant positive correlation between BOLD signal in the right TPJ [specifically, inferior parietal lobule (63, –35, 33) and supramarginal gyrus (57, –53, 33)] and the frequency of decisions to help the charity ($p < .005$; Fig. 4 and Table 2).

ANOVA: decision phase

To examine the nature of activity in the insula, ACC, and TPJ identified in the individual-differences analysis and to identify additional neural regions that might mediate decisions to aid or harm another, we conducted an exploratory 2 (trial type: gain or loss) × 2 (decision: self or charity) × 2 (magnitude: low [$2] or high [$5]) ANOVA on the BOLD signal data from the choice and decision intervals (see Table 3 for the detailed results). Notably, increased activity was observed in dorsomedial (−2, 1, 51) and left dorsolateral PFC (−42, 5, 25), inferior parietal lobule (−61, –47, 21), and posterior cingulate cortex (19, –44, 23) when assigning a loss as compared with a gain, whereas the

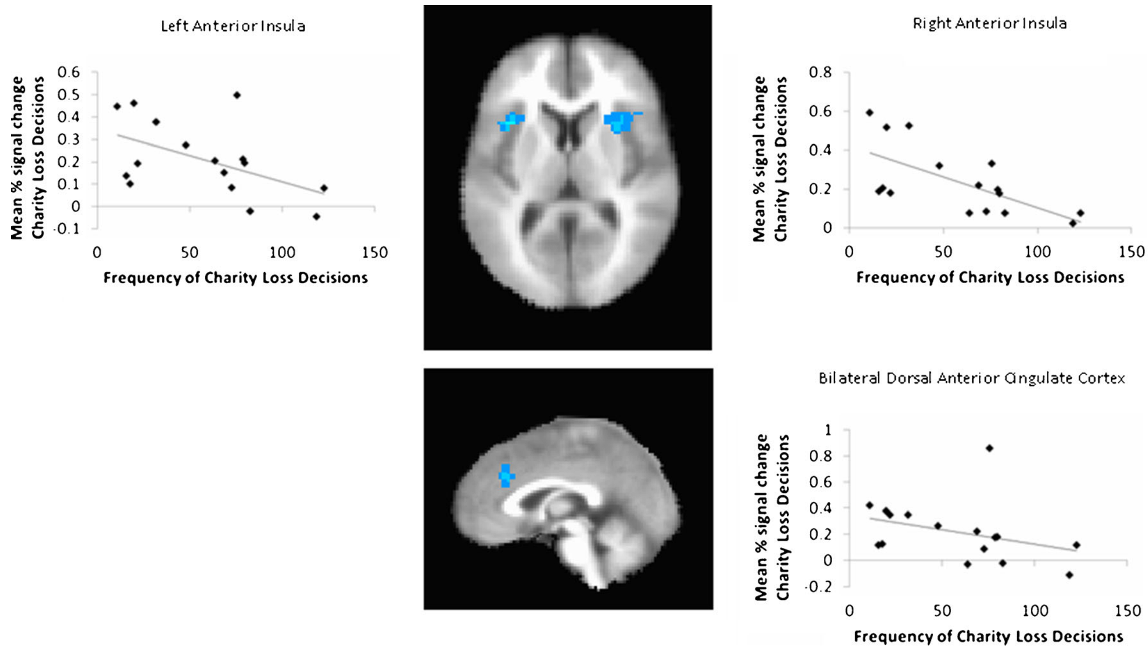

Fig. 3 Increased anterior insula activity during the decision phase is associated with fewer decisions to assign a loss to others. A whole-brain voxel-wise analysis of neural regions correlating with the frequency of decisions to assign a loss to charity (rather than to oneself) demonstrated significant negative correlations ($p < .005$), indicating that participants who showed greater activation in the anterior insula and dorsal ACC during decisions to assign a loss to charity were less likely to assign losses to charity throughout the task

Table 2 Significantly active clusters displaying a correlation between individual differences of frequency of decisions and BOLD response during the decision phase

Location	R/L	BA	*X*, *Y*, *Z*	Cluster size	*t* Value	Regression slope
BOLD versus behavioural frequency for charity-loss [High + Low] conditions						
Anterior insula/IFG	R	13/45	33, 18, 6	90	−3.82	Negative
IFG/Anterior insula	L	44/13	−42, 15, 12	63	−4.13	Negative
Dorsal anterior cingulate	L/R	32	2, 29, 32	35	−3.72	Negative
BOLD versus behavioural frequency for charity-gain [High + Low] conditions						
Inferior parietal lobe	R	40	63, –35, 33	32	4.12	Positive
Supramarginal gyrus	R	40	57, –53, 33	33	4.46	Positive
Cerebellum (cerebellar lingual)	L		−2, –44, –20	37	3.80	Positive
Midbrain	R/L		2, –23, –21	27	3.80	Positive

The Brodmann location (BA) is provided, along with coordinates for the center of mass in MNI space (*X*, *Y*, *Z*). Cluster size represents the number of contiguous voxels sharing a face, and the *t* value is the mean for all voxels in the cluster. Regression slope denotes the direction of the association between BOLD activity and the decision frequency in the respective conditions. All clusters were FWE corrected to $p < .05$ (reported cluster sizes are from an uncorrected threshold of $p < .005$)

opposite pattern was observed in the right thalamus. A partially overlapping network showed magnitude-related effects, with dorsolateral and dorsomedial PFC and the inferior parietal lobule showing increased activity during low ($2) versus high ($5) value trials. No significant interactions were observed at the uncorrected threshold of $p < .005$ (and $p < .05$ corrected).

ANOVA: outcome phase

To identify the neural regions differentially processing intended and unintended gains or losses for one's self or for another (the charity), we conducted a 2 (trial type: gain or loss) × 2 (decision: self or charity) × 2 (magnitude: low [$2] or high [$5]) × 2 (outcome: intended vs. unintended) ANOVA on the

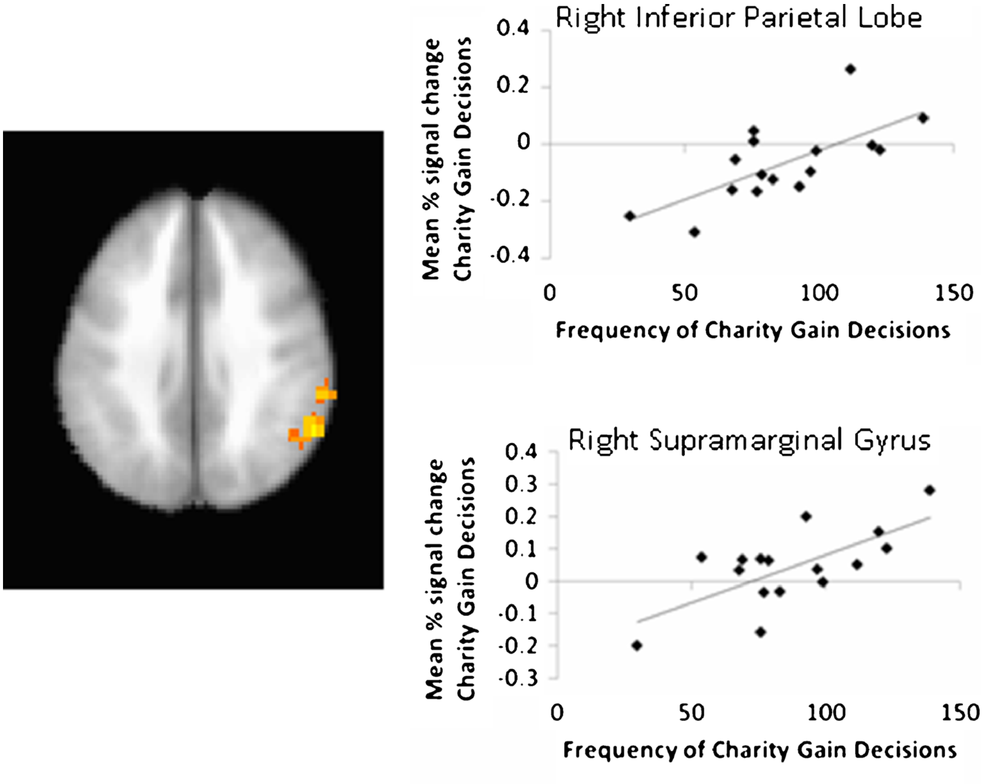

Fig. 4 Increased BOLD activity in temporo-parietal junction (TPJ) during the decision phase is associated with more decisions to assign a gain to others. A whole-brain voxel-wise analysis of neural regions correlating with the frequency of decisions to assign a gain to charity demonstrated significant positive correlations ($p < .005$), indicating that participants who showed greater activation in the right TPJ during decisions to assign a gain to charity were more likely to assign gains to charity throughout the task

Table 3 Significantly active clusters from the decision-phase ANOVA

Location	R/L	BA	*X*, *Y*, *Z*	Cluster size	*F* Value	Direction of effect
Main effect of magnitude [Low ($2) vs. High ($5)]						
DLPFC	R	8/9	41, 19, 45	225	14.85	Low > High
DLPFC	L	8/9/46	−46, 18, 37	151	13.60	Low > High
DLPFC	L	9/10	−40, 36, 29	51	14.51	Low > High
Inferior frontal gyrus	L	47/45	−50, 36, –1	26	14.12	Low > High
DMPFC	R/L	32/6	−5, 6, 50	56	14.09	Low > High
Angular gyrus	R	39	42, –65, 34	85	14.15	Low > High
Inferior parietal lobe	R	40	48, –46, 38	34	13.66	Low > High
Inferior parietal lobe	L	40	−45, –47, 41	26	12.36	Low > High
Superior parietal lobe	R	7	16, –67, 58	26	12.57	Low > High
Precuneus	L	7	−7, –53, 60	29	12.14	Low > High
Main effect of valence [Loss vs. Gain]						
Inferior frontal gyrus	L	9	−42, 5, 25	30	13.35	Loss > Gain
DMPFC	R/L	32/24	−2, 1, 51	63	12.42	Loss > Gain
Inferior parietal lobe	L	40	−61, –47, 21	37	13.50	Loss > Gain
Posterior cingulate cortex	R	31	19, –44, 23	26	13.56	Loss > Gain
Thalamus	R		8, –9, 4	27	14.52	Gain > Loss
Main effect of choice [Self vs. Charity]						
Cuneus	R	17	12, –93, 1	72	15.15	Charity > Self
Cuneus	L	18	−26, –97, –4	30	15.59	Charity > Self

The Brodmann location (BA) is provided, along with coordinates for the center of mass in MNI space (*X*, *Y*, *Z*). Cluster size represents the number of contiguous voxels sharing a face, and the *F* value is the mean for all voxels in the cluster. All clusters were FWE corrected to $p < .05$ (uncorrected threshold of $p < .005$)

BOLD signal data from the outcome interval (see Table 4 for the detailed results). This revealed several significant findings. Of note, significant main effects demonstrated increased anterior insula activity (right: 30, 20, 1; left: –31, 20, 6) during unintended relative to intended outcomes for both the charity and the self, and for high- relative to low-magnitude outcomes (Fig. 5). Second, a significant interaction of choice (self or charity) the valence (gain or loss) demonstrated increased activity in bilateral dorsolateral PFC (right: –45, 9, 40; left: 50, 11, 36) when assigning a loss to the charity relative to oneself, and increased activity when assigning a gain to the self relative to the charity (Table 3).

Discussion

The present study further elucidates the relationship between behavioral choice and neural signaling during decisions to cause social financial harms. Specifically, we highlight three main findings. First, we found that greater activity in the dorsal anterior insula and dorsal anterior cingulate cortex during a decision to financially harm another was correlated with a behavioral tendency to avoid such harmful decisions. Although this finding supports a role for the dorsal anterior insula and dorsal ACC in guiding behavioral choice related to social decisions that may evoke moral emotions such as guilt (Chang et al., 2011), the results support broader models of dorsal anterior insula functions during social decision making related to "risk aversion" and salience prediction error signaling (Bossaerts, 2010).

In the assessment of individual differences, greater activity in the anterior insula, along with ACC, was specifically associated with fewer decisions to financially harm the charity over the course of the task. The potential functional role of the insula in this regard may be clarified by examination of its pattern of activation during the outcome phase of the task. When participants were presented with the outcomes of their decisions to harm or help, the anterior insula responded in several task-specific ways. Increased activity in the anterior insula was observed during unintended as compared to intended outcomes. If anterior insula activation reflected feelings of guilt or empathy related to harm done to others, differential activation patterns during losses versus gains assigned to the charity should have been observed, and neural activity related to guilt should be greater for intended than for unintended harmful outcomes. However, our findings also indicate that dorsal anterior insula responded robustly to unintentional gains and losses assigned to both the charity and oneself. Increased activation in the anterior insula has previously been reported in risk prediction error signaling

Table 4 Significantly active clusters from the outcome-phase ANOVA

Location	R/L	BA	*X*, *Y*, *Z*	Size	*F* Value
Main effect of magnitude [All regions with high ($5) > Low ($2)]					
Anterior insula	R	13/47	42, 17, 20	439	15.33
Anterior insula	L	13/47	−38, 16, 0	60	14.67
DLPFC	L	9	−43, 14, 46	72	17.05
DLPFC	R	10	33, 41, 16	28	13.78
Superior frontal gyrus	R	9	15, 39, 35	31	16.29
Dorsal anterior cingulate	R	32/24	7, 31, 29	77	14.76
Dorsal anterior cingulate	L	32/24	−6, 33, 25	28	14.03
DMPFC	R/L	8	2, 22, 49	108	14.45
Angular gyrus	R	39	50, −65, 29	102	13.97
Caudate	R		10, 2, 15	81	16.56
Caudate	L		−13, −3, 20	47	15.34
Main effect of outcome					
[Unintended > Intended]					
Anterior insula	R	13	30, 20, 1	37	13.22
Anterior insula	L	13	−31, 20, 6	35	13.91
Middle temporal gyrus	R		58, −28, −8	26	13.93
[Intended > Unintended]					
Postcentral gyrus	L	5	−16, −46, 63	27	12.49
Main effect of valence [Gain > Loss]					
Rostral anterior cingulate	R/L	32	2, 42, 10	31	12.44
Interaction of valence × Outcome					
Frontal pole	R	10	33, 59, 17	34	15.68
Interaction of choice × Valence					
DLPFC	L	9/8	−45, 9, 40	117	14.11
DLPFC	R	9	50, 11, 36	52	13.27
Premotor area	R	6	35, −1, 58	54	12.82
Middle temporal gyrus	L	39	−56, −53, 12	60	13.69
Interaction of outcome × Magnitude					
Inferior occipital gyrus	R	18	31, −91, −8	32	13.30
Interaction of outcome × Choice × Valence					
Precentral gyrus	L	4	−43, −18, 47	62	16.47
Postcentral gyrus	R	3/2	46, −22, 52	15	14.16
Paracentral gyrus	R	5	2, −36, 58	47	13.46
Inferior parietal lobe	R	40	30, −42, 54	55	13.86
Middle temporal gyrus	R	39/37	54, −61, 4	38	12.98

The Brodmann location (BA) is provided, along with coordinates for the center of mass in MNI space (*X*, *Y*, *Z*). Cluster size represents the number of contiguous voxels sharing a face, and the *F* value is the mean for all voxels in the cluster. All clusters were FWE corrected to $p < .05$ (uncorrected threshold of $p < .005$)

(Preuschoff, Quartz, & Bossaerts, 2008), and recently during positive and negative prediction errors related to gustatory stimuli (Metereau & Dreher, 2013). Our findings are consistent with the proposed role of anterior insula in prediction error processing (Kuhnen & Knutson, 2005), and more specifically, with unsigned or salience prediction error models, in which the direction of signal change is the same for positive and negative prediction errors. This interpretation is further supported by a study of patients with insular cortex lesions on a decision-making task, demonstrating a role of the insular cortex in signaling the probability of aversive outcomes (Clark et al., 2008). A role in such outcome anticipation was also supported by the association of these regions with decision frequencies in the current task. Interestingly, coactivation of the anterior insula and ACC is common across a variety of tasks, in particular those related to emotion and decision making (Craig, 2009). Although the ACC is activated in a variety of conflict-monitoring and decision-making tasks, in the context of emotions and social behavior, it has been suggested more specifically that the ACC integrates homeostatic and emotional information from the insula to initiate behaviors and control autonomic responses (Chang, Yarkoni, Khaw, & Sanfey, 2013; Craig, 2009).

The assessment of individual differences revealed that activation in the TPJ (inferior parietal lobule) was positively correlated with the frequency of decisions to financially help the charity by assigning it a gain. Activity in dorsomedial PFC, along with regions of parietal cortex proximal to those identified here, has recently been associated with decisions of generosity or helping behaviors (Decety & Porges, 2011; Waytz et al., 2012). It has been put forth that the relationship between altruistic decisions and these regions is related to their role in theory-of-mind processing to understanding another's mental state (Waytz et al., 2012). In the present study, TPJ regions consistently activated by tasks involving theory-of-mind processing and mentalizing about others were the cortical regions positively correlated with generous decisions toward the charity. These results support a proposed link between interindividual differences related to perspective taking and generosity, and also extend prior findings focused on dorsomedial PFC and altruistic behaviors, to include another key node in the theory-of-mind network—the right TPJ.

A significant Valence × Choice interaction revealed greater activity in dorsolateral PFC during self-gain and charity-loss outcomes relative to self-loss and charity-gain outcomes. It is interesting to note that both of these scenarios represent the less altruistic outcomes. One possibility is that the observed dorsolateral PFC activity is associated with heighted attention related to the assessment of this potentially conflicting decision. This interpretation will require further validation but is based on the idea that these regions of dorsolateral PFC are activated in the context of suboptimal decision making or during decisions that generate more response conflict (Mitchell, 2011; Mitchell et al., 2009). During the decision phase, two additional regions of PFC, the inferior frontal gyrus and dorsomedial PFC, demonstrated increased activity during decisions to assign a loss to either one's self or the charity. These regions were identified previously during processing of moral transgressions or social transgressions specifically requiring response change or behavioral modification

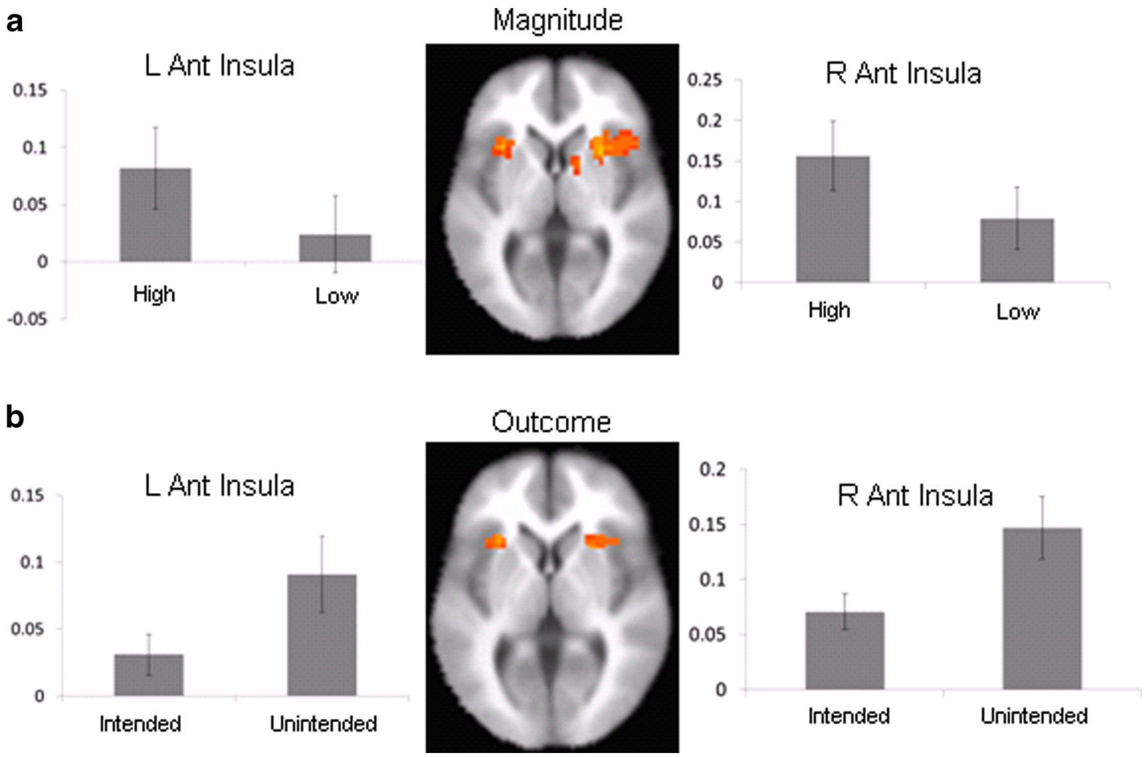

Fig. 5 Activation in dorsal anterior insula during feedback processing. Whole-brain voxel-wise analyses demonstrated (**a**) a significant main effect of magnitude, with increased BOLD signal in bilateral dorsal anterior insular cortex during high- versus low-magnitude outcomes; (**b**) a significant main effect of outcome, with increased BOLD signal in highly proximal regions of dorsal anterior insular cortex during unintended versus intended outcomes

(Finger et al., 2006). Inferior frontal gyrus and dorsomedial PFC have been implicated in representing and priming alternate response options in the setting of suboptimal responding (Budhani, Marsh, Pine, & Blair, 2007; Mitchell, 2011; Mitchell et al., 2009), such as in situations associated with negative feedback including loss of points during reversal learning errors (Budhani et al., 2007; Finger et al., 2008), a victim's distress (Finger et al., 2006), or anger (Blair & Cipolotti, 2000). The present findings add to prior work by demonstrating that these regions are also specifically activated during decisions to harm another, even before feedback is received. Thus, their activation may represent anticipation of the negative consequences of a decision, and the priming an alternate response in anticipation of negative feedback.

One potential limitation of the study design is that decisions to harm the charity were associated with a benefit for oneself, and vice versa. Thus, in each decision, multiple emotional and contextual factors beyond guilt or empathy, such as altruism, desire to please the examiner, value of the reward for self, and so forth, likely influenced participants' choices. In the present study, we did not model these potential emotions or influences separately. Although the task limits the dissociation of self and other consequences, the design was selected to best model this form of social decision making, in which personal "costs" also influence individual decisions in real environments—that is, to help another typically involves the donation of the agent's time, energy, or other resources. A second caveat for the interpretation of the three-way interaction in the feedback phase is that a small number of participants (from 0–3) had fewer than ten trial events for three of the self-as-recipient conditions (self intended gain, self unintended loss, and self unintended gain). Individual and mean trial numbers were robust for all two-way interactions and main effects, but the results from the three-way interaction, though not a focus of the results, may be underpowered.

Considering the results of the individual-differences analysis together with the task-related patterns of anterior insula activity, the question arises of how insula and ACC activity, putatively associated with risk prediction, may contribute to, or play a role in, individual differences in prosocial decision making. The finding that individuals with greater anterior insula and ACC activity during the decision phase made fewer decisions to harm the charity suggests that in some individuals such a decision is calculated as being more costly/aversive than taking the loss for oneself. Coupled with the finding that in the outcome phase, anterior insula and ACC activity were greatest for all unintended/unexpected outcomes, which is suggestive of a prediction error signal, the present results raise the possibility that the individual differences in insula and ACC activity during this type of decision making may be fundamentally mediated by individual differences in risk/cost calculations, rather than specifically by guilt representations.

Novel paradigms that can fully dissociate the emotion of guilt from risk calculations and prediction errors will be required to further test this hypothesis.

In summary, the present study indicates that activity in neural regions including the anterior insula, ACC, and TPJ reflects individual differences in helpful and harmful social decision making. Application of this functional neuroanatomy may inform future studies of neuropsychiatric disorders, such as psychopathic personality disorder or frontotemporal dementia, that feature frequent antisocial behaviors, abnormal risk assessment, and low capacity for prosocial emotions such as guilt and empathy.

References

Abu-Akel, A. (2003). A neurobiological mapping of theory of mind. *Brain Research Reviews, 43,* 29–40.

Bar-On, R., Tranel, D., Denburg, N. L., & Bechara, A. (2003). Exploring the neurological substrate of emotional and social intelligence. *Brain, 126,* 1790–1800.

Blair, R. J., & Cipolotti, L. (2000). Impaired social response reversal: A case of "acquired sociopathy. *Brain, 123,* 1122–1141.

Bossaerts, P. (2010). Risk and risk prediction error signals in anterior insula. *Brain Structure and Function, 214,* 645–653. doi:10.1007/s00429-010-0253-1

Budhani, S., Marsh, A. A., Pine, D. S., & Blair, R. J. (2007). Neural correlates of response reversal: Considering acquisition. *NeuroImage, 34,* 1754–1765. doi:10.1016/j.neuroimage.2006.08.060

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron, 70,* 560–572. doi:10.1016/j.neuron.2011.02.056

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex, 23,* 739–749. doi:10.1093/cercor/bhs065

Chapman, H. A., & Anderson, A. K. (2012). Understanding disgust. *Annals of the New York Academy of Sciences, 1251,* 62–76. doi:10.1111/j.1749-6632.2011.06369.x

Clark, L., Bechara, A., Damasio, H., Aitken, M. R., Sahakian, B. J., & Robbins, T. W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. *Brain, 131,* 1311–1322.

Coricelli, G., & Rustichini, A. (2010). Counterfactual thinking and emotions: Regret and envy learning. *Philosophical Transactions of the Royal Society B, 365,* 241–247. doi:10.1098/rstb.2009.0159

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10,* 59–70. doi:10.1038/nrn2555

Decety, J., & Porges, E. C. (2011). Imagining being the agent of actions that carry different moral consequences: An fMRI study. *Neuropsychologia, 49,* 2994–3001. doi:10.1016/j.neuropsychologia.2011.06.024

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23,* 475–483.

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient EVR. *Neurology, 35,* 1731–1741.

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,* 414–421. doi:10.1016/j.neuroimage.2006.06.011

Finger, E. C., Marsh, A. A., Mitchell, D. G., Reid, M. E., Sims, C., Budhani, S., & Blair, J. R. (2008). Abnormal ventromedial prefrontal cortex function in children with psychopathic traits during reversal learning. *Archives in General Psychiatry, 65,* 586–594. doi:10.1001/archpsyc.65.5.586

Grafman, J., Schwab, K., Warden, D., Pridgen, A., Brown, H. R., & Salazar, A. M. (1996). Frontal lobe injuries, violence, and aggression: A report of the Vietnam Head Injury Study. *Neurology, 46,* 1231–1238.

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,* 389–400.

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,* 2105–2108.

Han, T., Alders, G. L., Greening, S. G., Neufeld, R. W., & Mitchell, D. G. (2012). Do fearful eyes activate empathy-related brain regions in individuals with callous traits? *Social Cognitive and Affective Neuroscience, 7,* 958–968. doi:10.1093/scan/nsr068

Hare, R. D. (1970). *Psychopathy: Theory and research*. New York, NY: Wiley.

Harenski, C. L., Harenski, K. A., Shane, M. S., & Kiehl, K. A. (2010). Aberrant neural processing of moral violations in criminal psychopaths. *Journal of Abnormal Psychology, 119,* 863–874.

Harenski, C. L., Kim, S. H., & Hamann, S. (2009). Neuroticism and psychopathy predict brain activation during moral and nonmoral emotion regulation. *Cognitive, Affective, & Behavioral Neuroscience, 9,* 1–15. doi:10.3758/CABN.9.1.1

Heekeren, H. R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H. P., & Villringer, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage, 24,* 887–897.

Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *NeuroReport, 14,* 1215–1219.

Kubany, E. S., & Watson, S. B. (2003). Guilt: Elaboration of a multidimensional model. *Psychological Record, 53*(1), 4.

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron, 47,* 763–770.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition, 108,* 754–770. doi:10.1016/j.cognition.2008.06.009

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual*. Gainsville, FL: University of Florida, Center for Research in Psychophysiology (Technical Report No. A-8).

Lilienfeld, S. O., & Widows, M. R. (2005). *Professional manual for PPI-R: The Psychopathic Personality Inventory–Revised*. Lutz, FL: Psychological Assessment Resources.

Lynöe, N., Sandlund, M., Dahlqvist, G., & Jacobsson, L. (1991). Informed consent: Study of quality of information given to participants in a clinical trial. *British Medical Journal, 303,* 610–613.

McGraw, K. M. (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of Personality and Social Psychology, 53,* 247–256.

Metereau, E., & Dreher, J. C. (2013). Cerebral correlates of salient prediction error for different rewards and punishments. *Cerebral Cortex, 23,* 477–487. doi:10.1093/cercor/bhs037

Mitchell, D. G. (2011). The nexus between decision making and emotion regulation: A review of convergent neurocognitive substrates. *Behavioural Brain Research, 217,* 215–231.

Mitchell, D. G., Luo, Q., Avny, S. B., Kasprzycki, T., Gupta, K., Chen, G., & Blair, R. J. (2009). Adapting to dynamic stimulus–response values: Differential contributions of inferior frontal, dorsomedial, and dorsolateral regions of prefrontal cortex to decision making. *Journal of Neuroscience, 29,* 10827–10834. doi:10.1523/JNEUROSCI.0963-09.2009

Moll, J., de Oliveira-Souza, R., Bramati, I. E., & Grafman, J. (2002a). Functional networks in emotional moral and nonmoral social judgments. *NeuroImage, 16,* 696–703.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002b). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience, 22,* 2730–2736.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., & Grafman, J. (2006). Human fronto-mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences, 103,* 15623–15628. doi:10.1073/pnas.0604475103

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience, 28,* 2745–2752. doi:10.1523/JNEUROSCI.4286-07.2008

Rudorf, S., Preuschoff, K., & Weber, B. (2012). Neural correlates of anticipation risk reflect risk preferences. *Journal of Neuroscience, 32,* 16683–16692. doi:10.1523/JNEUROSCI.4235-11.2012

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience, 7,* 499–500.

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York, NY: Springer.

Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., & Rauch, S. L. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry, 48,* 43–50.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303,* 1157–1161. doi:10.1126/science.1093535

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: An fMRI study. *NeuroImage, 23,* 967–974.

Tangney, J. P., Wagner, P. E., & Gramzow, R. (1989). *The test of self-conscious affect*. Fairfax, VA: George Mason University.

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. *Journal of Neuroscience, 32,* 7646–7650. doi:10.1523/JNEUROSCI.6193-11.2012

Zaki, J., & Mitchell, J. P. (2012). Equitable decision making is associated with neural markers of intrinsic value. *Proceedings of the National Academy of Sciences, 108,* 19761–19766. doi:10.1073/pnas.1112324108