OXFORD

Social Cognitive and Affective Neuroscience, 2016, 253–262

doi: 10.1093/scan/nsv105
Advance Access Publication Date: 13 August 2015
Original Article

Neural signatures of third-party punishment: evidence from penetrating traumatic brain injury

Leila Glass,[1] Lara Moody,[2,3] Jordan Grafman,[4] and Frank Krueger[5,6]

[1]Department of Psychology, SDSU/UCSD Joint Doctoral Program in Clinical Psychology, San Diego, CA, USA, [2]Virginia Tech Carilion Research Institute, Roanoke, VA, USA, [3]Department of Psychology, Virginia Tech, Blacksburg, VA, USA, [4]Brain Injury Research Program, Rehabilitation Institute of Chicago, Chicago, IL, USA, [5]Molecular Neuroscience Department and [6]Department of Psychology, George Mason University, Fairfax, VA, USA

Correspondence should be addressed to Frank Krueger, Molecular Neuroscience Department/Department of Psychology, George Mason University, 4400 University Drive, Mail Stop 2A1, Fairfax, VA 22030, USA. E-mail: FKrueger@gmu.edu.

Abstract

The ability to survive within a cooperative society depends on impartial third-party punishment (TPP) of social norm violations. Two cognitive mechanisms have been postulated as necessary for the successful completion of TPP: evaluation of legal responsibility and selection of a suitable punishment given the magnitude of the crime. Converging neuroimaging research suggests two supporting domain-general networks; a mentalizing network for evaluation of legal responsibility and a central-executive network for determination of punishment. A whole-brain voxel-based lesion-symptom mapping approach was used in conjunction with a rank-order TPP task to identify brain regions necessary for TPP in a large sample of patients with penetrating traumatic brain injury. Patients who demonstrated atypical TPP had specific lesions in core regions of the mentalizing (dorsomedial prefrontal cortex [PFC], ventromedial PFC) and central-executive (bilateral dorsolateral PFC, right intraparietal sulcus) networks. Altruism and executive functioning (concept formation skills) were significant predictors of TPP: altruism was uniquely associated with TPP in patients with lesions in right dorsolateral PFC and executive functioning was uniquely associated with TPP in individuals with lesions in left PFC. Our findings contribute to the extant literature to support underlying neural networks associated with TPP, with specific brain-behavior causal relationships confirming recent functional neuroimaging research.

Key words: neurolaw; social cognition; altruistic punishment; morality; prefrontal cortex; traumatic brain injury

Introduction

A variety of selection pressures exist resulting in the survival of communities with specific sets of capacities. Impartial third-party punishment (TPP) has been hypothesized to be a capacity related to successful societal cooperation and collective functioning (Fehr and Fischbacher, 2004a). Legal TPP, the response to violations of rules that are communally agreed upon, is a fundamental part of a modern social system that enforces social norms (Fehr and Fischbacher, 2004a). The concept of impartial TPP—in which individuals evaluate and punish a violator of norms even when they are not directly affected—enforces cooperation with social standards and ultimately leads to a stable contemporary society (Fehr and Fischbacher, 2004a,b; Strobel *et al.*, 2011). Large-scale modern societies function on the expectation that criminal offenses will be punished by impartial state-empowered enforcers (e.g. judges and jurors) within an established criminal justice system.

Separate cognitive mechanisms have been proposed to underlie successful impartial TPP: (i) evaluation of legal responsibility and (ii) selection of punishment (Buckholtz and Marois, 2012). First, to be held legally responsible under criminal law, in most cases perpetrators must have acted with a guilty intent (*mens rea*) and must have committed a prohibited act (*actus reus*) inflicting actual harm on the victim (Shen *et al.*, 2011). Once the

Received: 27 February 2015; **Revised:** 9 July 2015; **Accepted:** 8 August 2015

© The Author (2015). Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

evaluation is conducted, one must be able to select a suitable punishment given the magnitude of the crime. These critical components of enacting successful TPP are thought to recruit a broad network of underlying cognitive mechanisms and their related neural correlates.

It is hypothesized that theory of mind, the ability to understand that others have beliefs, intentions and feelings that are different from one's own can facilitate the evaluation of responsibility (Premack and Woodruff, 1978). This ability allows for inferences regarding another's mental state, which are used to explain or predict social behavior (Frith and Frith, 1999). Theory of mind can be conceptualized as occurring on both affective and cognitive dimensions (Baron-Cohen *et al.*, 1997; Shamay-Tsoory and Aharon-Peretz, 2007; Kalbe *et al.*, 2010; Leopold *et al.* 2012). In the context of TPP, cognitive theory of mind is thought to occur when evaluating the intention of the perpetrator. In contrast, affective theory of mind can be interpreted as being related to the evaluation of harm to the victim. Once the evaluation of the responsibility is executed successfully, the second step of TPP is hypothesized to be the ability to determine an appropriate punishment that 'fits the crime' (Buckholtz and Marois, 2012). It is expected that determining and selecting the correct punishment from an array of contextually appropriate options requires intact executive functioning (e.g. intact comparison, concept formation, problem solving, inhibitory control, sorting, magnitude judgment) (Knoch *et al.*, 2006).

Neuroimaging studies suggest that successful TPP builds on two domain-general neural networks. The capacity for TPP is not a result of a specialized cognitive module that evolved specifically for this social function, but rather that other existing fundamental cognitive mechanisms (e.g. executive function, theory of mind, response selection) have been efficiently reallocated for this higher-order cognitive process (Buckholtz *et al.*, 2008; Schleim *et al.*, 2011; Buckholtz and Marois, 2012; Yamada *et al.*, 2012; Krueger *et al.*, 2014). On the one hand, the mentalizing brain network is thought to consist of core regions such as the dorsomedial prefrontal cortex (dmPFC) and temporo-parietal junction (TPJ), engaging in cognitive theory of mind and the ventromedial PFC (vmPFC) and amygdala engaging in affective theory of mind. On the other hand, the central-executive network is thought to consist of core regions such the intra-parietal sulcus (IPS) for context-specific representation of the response space used to construct a scale of punishment and dorsolateral PFC (dlPFC) for the selection of the appropriate magnitude of punishment (Buckholtz and Marois, 2012).

The majority of studies investigating the neural underpinnings of TPP have used functional imaging techniques—excellent methodology for identifying brain regions of interest in specific processes—though a less powerful tool for making causative conclusions about brain regions necessary for task completion (Poldrack, 2011). Neuroscientific techniques such as transcranial magnetic stimulation (TMS) and transcranial direct-current stimulation (tDCS) have also been employed to unpack the causal relationship between behavior and brain for TPP (van't Wout *et al.*, 2005; Knoch *et al.*, 2006; Knoch *et al.*, 2008). As with all techniques, there are limitations for neurostimulation including determining the precise location of stimulation, limited reach beyond cortical structures and debated interpretation. The need for converging evidence supporting neural networks across methodologies should not be understated and therefore lesion studies are needed to bolster relationships seen using other techniques (Chatterjee, 2005; Barbey *et al.*, 2012; Barbey *et al.*, 2014).

We combined a rank-order TPP task (Robinson and Kurzban, 2007) with a whole-brain voxel-based lesion-symptom mapping (VLSM) approach to identify brain regions necessary for TPP in a large sample of patients with penetrating traumatic brain injury (pTBI). Participants were given separate scenarios in which a perpetrator, 'John', engages in hypothetical criminal offenses. They were asked to rank-order the scenarios (which included a range of criminal offenses that varied in severity) to reflect the relative degree of punishment that John deserved. On the basis of the proposed neuropsychological model of impartial TPP (Buckholtz and Marois, 2012), we hypothesized that participants with pTBI, compared with non-injured normal controls, would demonstrate atypical TPP related to lesions in core regions of the theorized underlying TPP networks.

Materials and methods

Subjects

All participants were recruited as part of the W.F. Caveness Vietnam Head Injury Study (VHIS). The VHIS registry is a longitudinal study that includes a large sample of American male veterans who suffered from pTBI and non-injured control veterans who experienced combat but did not suffer brain damage while serving in the Vietnam War. The VHIS registry consists of four phases described in detail elsewhere (Raymont *et al.*, 2011). Phase I was the recruitment period for the registry, Phase II occurred between 1981 and 1984 at the Walter Reed Army Medical Center and involved administration of a neuropsychological battery, Phase III occurred approximately 20 years later between 2003 and 2006 and consisted of both neuropsychological testing and computed tomography (CT) acquisition at the National Navy Medical Center in Bethesda, MD, and Phase IV (2008-2012) was completed as a follow-up assessment, consisting of a week-long testing battery at the National Institute of Neurological Disorders and Stroke (NINDS), Bethesda, MD. For this study, a subset of the total sample population completed the TPP task during Phase IV, including both brain-injured veterans (pTBI=114) and a non-injured normal control group (NC=32). To ensure that veterans were eligible to participate in Phase IV, a phone interview prior to arrival and a neurological exam at the test site were conducted to screen all participants for psychological and neurological exclusion symptoms. All participants gave their written informed consent and the Institutional Review Board (IRB) at NINDS approved all study procedures.

CT acquisition and lesion identification

Axial CT scans without contrast were acquired at the Bethesda Naval Hospital on a GE Medial Systems Light Speed Plus CT scanner in helical mode during Phase III. Structural neuroimaging data were reconstructed with an in-plane voxel size of 0.4 $\times$0.4 mm, an overlapping slice thickness of 2.5 mm, and a 1 mm slice interval. Lesion location and volume were documented from the CT images by using the Analysis of Brain Lesion software (Makale *et al.*, 2002; Solomon *et al.*, 2007) implemented in MEDx v.3.44 (Medical Numeric) with enhancements to include the automated anatomical labeling (AAL) atlas (Tzourio-Mazoyer *et al.*, 2002).

Similar to the methodology used in many other lesion analysis studies (Heberlein *et al.*, 2004), lesion tracing was performed manually on each slice in native space by a neuropsychiatrist with clinical experience in reading CT scans (V.R.) and was subsequently reviewed by the principal investigator of the VHIS (J.G.). Both reviewers were blind to the results of the clinical evaluation

and neuropsychological testing, and reliable consensus was reached across reviewers regarding the boundaries of each lesion, reducing potential for bias. The CT image of each individual's brain was normalized to a CT template brain image in Montreal Neurological Institute (MNI) space using an automated image registration (AIR) algorithm with 12-parameter affine fit (Woods *et al*., 1993). Both the subject's brain and the MNI template's brain were skull-stripped to maximize the efficacy of the AIR registration from native space to MNI space and voxels inside the traced lesion were not included in the spatial normalization procedure (Tzourio-Mazoyer *et al*., 2002).

Neurobehavioral and neuropsychological testing

All participants underwent a week-long testing battery administered by trained experimenters that included both experimental tasks and standardized neuropsychological control measures. For the experimental task of interest, we used the rank-order TPP task that has been used in previous studies to assess normative judgment (Robinson and Kurzban, 2007). Participants were given a set of 24 vignettes of criminal offenses on separate index cards (Supplemental Table SI). Each card had a scenario consisting of two or three sentences, which described an event in which an offender 'John' engages in a hypothetical criminal offense (Robinson and Kurzban, 2007; Krueger *et al*., 2013, 2014). The average reading comprehension for all scenarios according to the Flesch-Kincaid grade level assessment was 8.6 with a Flesch reading ease of 71 (0, very difficult to 100, very easy), indicating a fairly easy readability level, especially for our sample where the average education level was 14.81 with a minimum education of 11 years across groups (Flesch, 1948; Flesch and Gould, 1949).

The offenses in the scenarios included a variety of situations that focused on a range of criminal offenses such as theft, property destruction, assault, burglary, robbery, rape, negligence, kidnapping, torture and murder. The scenarios increased in severity of crime on an ordinal scale from 1, which involved self-defense, to 24, which involved ransom, rape, torture and strangling of an 8-year-old girl. Each card had a brief header summarizing the scenario. An example of a scenario on the more severe end: 'Stabbing—John is offended by a woman's mocking remark and decides to hurt her badly. At work the next day, when no one else is around, he picks up a letter opener from his desk and stabs her. She later dies from the wound.'

To create the TPP situation, participants were given complete discretion to punish John for his actions. They were instructed to rank-order the cards (from 1 to 24) to reflect the relative degree of punishment that John deserved for each card, independently. Participants were instructed to take as much time as they needed, to not allow any ties and to base their answers solely on what they believed to be just, regardless of any laws. This behavior provided the measure of TPP as they decided which scenarios, based on evaluation of intent and harm, deserved more or less punishment. The actual a priori order, which was taken from a previous study that developed and provided the modal rank for the criminal scenarios (i.e. normative sample), was never revealed to the participants (Robinson and Kurzban, 2007) (see Supplementary Table S1). To determine the divergence from the expected a priori rank-order, bivariate Spearman correlation coefficients (i.e. rank-order-coefficient, ROC) were calculated between the a priori rank-order and the rank-order of scenarios of each participant from both groups (pTBI, NC) (Dimitrov *et al*., 1996).

As previous functional neuroimaging studies have generally used a different paradigm to investigate TPP—asking their participants to estimate how much punishment the offender deserved for each of the criminal offenses on a Likert-scale (0, no punishment to 100, extreme punishment)—we performed a pilot study to compare paradigms to assess whether there was equivalency between the rank-order task and rating task. This pilot study was approved by the IRB at George Mason University, Fairfax, VA. In this study, healthy volunteers (13 females, 13 males; years of age [mean±s.d.]: 25.7±4.3, years of education: 17.4±2.3) participated for financial compensation after giving their written informed consent. Using counterbalancing, half of the participants were randomly assigned to perform the rank-order task first, and then the rating task, whereas the other half first the rating task and then the rank-order task. The rank-order task used the same methodology as the current lesion study. To determine the divergence from the expected punishment rank-order performance, the same bivariate Spearman correlation coefficient (i.e. ROC) was calculated between the a priori rank-order and the rank-order of scenarios of each individual participant. Our results demonstrated that performances in both tasks (the rating paradigm and the rank-order paradigm) were highly correlated with each other (r=0.97, P<0.001), indicating that they both measure the degree of punishment that the perpetrator deserved for each offense. In addition, a recent study found that these punishment assessments through crime vignettes are ecologically valid as the Likert-scale ratings, provided (e.g. 1, 25, 50, 75) align in magnitude to real-world legal criminal judgments that participants equate to being justified by their respective numerical ratings by examining their internal scale of punishment (Krueger *et al*., 2014). Therefore, the rank-order task assesses the degree of punishment that the offender deserved in a comparable fashion to the Likert-scale methodology used in previous functional neuroimaging studies, suggesting equivalence between the tasks in terms of assessing TPP.

Participants in the lesion study were assessed on a variety of neuropsychological control measures, including naming ability (Boston Naming Test, BNT) (Kaplan *et al*., 2001), basic verbal comprehension (Token Test, TT) (McNeil and Prescott, 1978), executive functioning including concept formation skills and ordering ability (Delis-Kaplan Executive Functioning System, D-KEFS) (Delis *et al*., 2001), memory (Wechsler Memory Scale-III abbreviated) (Wechsler, 1997), pre-injury general intelligence (Armed Forces Qualification Test, AFQT-7A) (Bayroff and Anderson, 1963) and post-injury general intelligence (Wechsler Abbreviated Scale of Intelligence, WASI) (Wechsler, 1999). The AFQT was administered upon entry into the military; it has been extensively standardized by the US military and correlates highly with the WASI intelligence quotient (Grafman *et al*., 1988). Neurobehavioral and psychological symptoms were assessed using validated and standardized self-report measures including assessments of severity of depression (Beck Depression Inventory, BDI-II) (Beck *et al*., 1996), empathy (Interpersonal Reactivity Index, IRI) (Davis, 1983), trait anger (State-Trait Anger Inventory, STAXI) (Spielberger, 1999) and altruism (Self-Report Altruism Scale, SRAS) (Rushton *et al*., 1981). The Neurobehavioral Rating Scale (NBRS) was administered as a clinician-based measure of functioning and psychopathology (Sultzer *et al*., 1992).

Data analysis

All behavioral analyses were carried out using SPSS (Statistical Package for the Social Sciences, version 20.0, 2011) with alpha

set to $P < 0.05$. First, the experimental measure of TPP (ROC) was compared between pTBI and NC groups using a one-tailed independent-samples *t*-test. Note that a one-tailed interval was employed because of the a priori directional predictions for the lesion group compared with the control group regarding atypical TPP. Demographical and neuropsychological control measures were compared between pTBI and NC groups using two-tailed independent-samples *t*-tests or chi-square tests. Effect sizes as Cohen's *d* were calculated with 0.2, 0.5, and 0.8 described as small-, medium- and large-sized effects, respectively. Next, within the pTBI group, the influence of each neuropsychological control measure on TPP was calculated by applying a stepwise multiple linear regression analysis, including ROC as the dependent variable and naming ability, basic verbal comprehension, memory, concept formation skill, ordering ability, pre- and post-injury general intelligence, severity of depression, empathy, trait anger, altruism and psychopathology as predictor variables. Importantly, the analysis allowed an estimation of the relative contribution of each predictor to ROC at the same time as controlling for potential confounding factors that may also influence TPP.

Next, we computed the distribution of lesions of the pTBI population by overlapping patients' spatially normalized lesion images at each voxel onto a single-subject standardized brain to ensure that previously proposed core TPP brain areas were covered (Buckholtz and Marois, 2012). Furthermore, to identify brain regions necessary for TPP, we performed a whole-brain VLSM analysis (one-tailed *t*-test, *q*(FDR)=0.05, minimum cluster size of 50 voxels) using ROC as the dependent variable and the lesion status (gray matter) of each voxel as the independent variable. The analysis was restricted to a minimum overlap of three brain-injured subjects at a given voxel for the VLSM analysis to ensure adequate coverage. Gray matter structures were obtained by applying the AAL atlas (Tzourio-Mazoyer *et al.*, 2002; Barbey *et al.*, 2014). Finally, we calculated one-tailed partial correlations (applying Bonferroni correction) between our dependent variable (i.e. ROC) and significant predictors (i.e. neuropsychological control measures) from the regression analysis to determine the specificity of the lesion-behavior relationship for identified brain regions from the VLSM analysis.

Results

Our initial comparison included the pTBI and NC groups to assess the impact of brain injury on potential differences from expected TPP. The pTBI group differed significantly from the NC group (Cohen's d=0.45), with a lower ROC indicating higher deviation from the correct rank-order (Table 1). Groups were matched on demographic variables (age, education, handedness) and on most of the neuropsychological control measures (e.g. pre- and post-injury general intelligence, verbal comprehension, naming ability, ordering ability, altruism, pathology), with the exception of measures of general memory, concept formation and severity of depression, which were significantly lower in the pTBI than in the NC group; however, all were in the normal, non-impaired and non-clinical range.

Next, we ran a stepwise multiple linear regression analysis to assess the relative contributions of these variables. Two significant models emerged that allowed for an estimation of the relative contribution of each predictor to TPP performance. The final model accounting for 39% of the variance in the TPP performance ($F_{(2,100)}$=32.94 P <0.001; adjusted R^2=0.39) included two significant predictors. The higher patients' performance on concept formation skills (β=0.45; $P < 0.001$; adjusted R^2=0.22) and the higher the disposition of altruism (β=0.42; $P < 0.001$; adjusted $R^2 = 0.17$), the more consistent or similar the patient group's TPP performance was to the control group's TPP performance. No significant associations were found for ordering ability, naming ability, basic verbal comprehension, memory, pre- and post-injury general intelligence, severity of depression,

Table 1. Descriptive (mean±s.e.m.) and inferential statistics for demographic, experimental and neuropsychological control measures for the penetrating traumatic brain injury group (pTBI=114) and the normal control group (NC=32)

Measures	pTBI	NC	Statistics
Demographic variables			
Age (years)	63.36±0.27	63.41±0.67	t=−0.07, P=0.941
Education (years)	14.73±0.19	15.09±0.38	t=−0.87, P=0.384
Handedness (R:A:L)	92:2:20	25:2:5	χ^2=1.92, P=0.381
Experimental variable			
Third-Party Punishment (ROC)	0.88±0.01	0.94±0.02	t=−2.51, P<0.007
Control variables			
Pre-Injury Intelligence (AFQT-7, percentile)	67.20±2.18	7.73±3.72	t=−1.28, P=0.202
Post-Injury Intelligence (WASI-III, full scaled score)	114.47±2.32	113.29±2.15	t=0.39, P=0.693
General Memory (WMS-III-A, scaled score)	99.15±1.37	105.77±2.62	t=−2.23, P<0.027
Concept Formation (D-KEFS ST, scaled score)	10.57±0.29	12.44±0.54	t=−2.97, P<0.003
Ordering Ability (D-KEFS TM, letter sequencing errors)	0.01±0.09	0.00±0.00	t=−0.83, P=0.409
Naming Ability (BNT, raw score)	54.11±0.55	55.78±0.68	t=−1.50, P=0.135
Verbal Comprehension (TT, raw score)	98.10±0.21	98.38±0.35	t=−0.62, P=0.534
Depression (BDI-II, total score)	7.16±0.69	10.97±1.53	t=−2.48, P<0.014
Anger Trait (STAXI, scaled score)	44.72±0.91	48.69±2.02	t=−1.96, P=0.052
Empathy (IRI, total score)	83.28±0.96	79.91±1.99	t=1.61, P=0.110
Altruism (SRAS, total score)	63.90±1.05	59.94±1.84	t=1.79, P=0.074
Pathology (NBRS, total score)	35.48±0.93	36.00±1.57	t=−0.27, P=0.789

Handedness: R, right-handed; A, ambidextrous; L, left-handed; ROC, rank-order-coefficient; AFQT, Armed Forces Qualification Test; WASI-III, Wechsler Abbreviated Scale of Intelligence Third Edition; WMS-III-A, Wechsler Memory Scale Third Edition Abbreviated; D-KEFS ST; Delis–Kaplan Executive Function System Sorting Test; D-KEFS TM; Delis–Kaplan Executive Function System Trail Making Test; BNT, Boston Naming Test; TT, Token Test; BDI-II, Beck Depression Inventory Second Edition; STAXI, State-Trait Anger Inventory, IRI, Interpersonal Reactivity Index; SRAS, Self-Report Altruism Scale; NBRS; Neurobehavioral Rating Scale.

empathy, trait anger and psychopathology as predictor variables (Table 2).

Subsequently, we overlapped patients' spatially normalized lesion images onto a single subject standardized brain (Figure 1). Their lesions covered all a priori identified brain regions associated with TPP based on the neurocognitive hypothesis proposed by Buckholtz and Marois (2012) with an observed maximum power of coverage in the prefrontal areas, with the exception of the bilateral amygdala (Supplementary Table SII).

Then, we ran the whole-brain VLSM analysis (q(FDR) corrected value of 2.76, minimum cluster size of 50 voxels) and identified four significant clusters: (1) Left dmPFC-lPFC, including dmPFC, left dorsolateral PFC (L dlPFC), and left ventrolateral PFC (L vlPFC), (2) right vmPFC (R vmPFC), (3) R dlPFC-SMA, including R dlPFC and supplementary motor area (R SMA), and (4) R IPS (Figure 2, Table 3).

Finally, to determine the specificity of the lesion-behavior relationship of the identified clusters from the VLSM analysis, for participants in each cluster, we calculated partial correlations between our dependent variable (i.e. ROC) and each significant predictor (i.e. concept formation, altruism) from the regression analysis controlling simultaneously for the other predictors. We found that altruism and TPP performance was significantly correlated for participants with lesions in the R dlPFC-SMA cluster, whereas concept formation skills and TPP performance was significantly correlated in participants with lesions in the L dmPFC-lPFC cluster (Table 4).

Table 2. Multiple linear regression results for nonsignificant predictors

Predictor variable	Beta	*P*
Pre-Injury Intelligence (AFQT-7, percentile)	−0.01	0.83
Post-Injury Intelligence (WASI-III, full scaled score)	0.12	0.92
General Memory (WMS-III-A, scaled score)	0.10	0.78
Ordering Ability (D-KEFS TM, letter sequencing errors)	0.01	0.98
Naming Ability (BNT, raw score)	0.11	0.81
Verbal Comprehension (TT, raw score)	−0.05	0.82
Depression (BDI-II, total score)	0.11	0.98
Anger Trait (STAXI, scaled score)	−0.05	0.97
Empathy (IRI, total score)	−0.14	0.93
Pathology (NBRS, total score)	0.01	0.84

AFQT, Armed Forces Qualification Test; WASI-III, Wechsler Abbreviated Scale of Intelligence Third Edition; WMS-III-A, Wechsler Memory Scale Third Edition Abbreviated; D-KEFS TM; Delis–Kaplan Executive Function System Trail Making Test; BNT, Boston Naming Test; TT, Token Test; BDI-II, Beck Depression Inventory Second Edition; STAXI, State-Trait Anger Inventory, IRI, Interpersonal Reactivity Index; SRAS, Self-Report Altruism Scale; NBRS; Neurobehavioral Rating Scale.

Discussion

We aimed to contribute evidence from brain injury subjects to better understand and support the proposed neural correlates and networks associated with TPP behavior. Our findings corroborate existing hypotheses regarding the underlying neural networks associated with TPP and add specific brain-behavior causal inferences to better understand the repurposing of domain-general cognitive functions for this higher-order process (Buckholtz and Marois, 2012). The pTBI group demonstrated divergence from expected TPP performance, despite having no differences from controls on intelligence, the ability to order, verbal comprehension, empathy and other control measures. In our assessment of the predictors of atypical TPP in the pTBI group, we found that altruism and concept formation both emerged as accounting for approximately the same variance (17% and 22%, respectively). Altruism was uniquely associated with TPP in individuals with lesions in R dlPFC extending into the SMA and concept formation was uniquely associated with TPP in individuals with lesions in the L PFC. While these predictors account for approximately 39% of the variance in TPP, it is clear that there are many other domain general skills also contributing to successful performance. Impartial TPP within the criminal justice system is thought to rely on two components: the evaluation of legal responsibility based on intent of the perpetrator and harm inflicted on the victim and the determination of a suitable and commensurate punishment (Shen *et al.*, 2011), requiring the mentalizing and central-executive neural networks, respectively. As hypothesized, we found that participants with atypical TPP had lesions in core regions of both the mentalizing (dmPFC, vmPFC) and central-executive (R IPS, bilateral dlPFC) networks.

The mentalizing network consists of both cognitive and affective theory of mind as dimensions to understand the thoughts and emotions of others (Shamay-Tsoory and Aharon-Peretz, 2007; Stein *et al.*, 2007; Leopold *et al.*, 2012). These dimensions relate to two distinct neural pathways, the 'goal pathway' (dmPFC, TPJ) for cognitive theory of mind, enabling inferences about the intentions of the perpetrator, and the 'outcome pathway' (vmPFC, amygdala) for affective theory of mind, enabling inferences about the inflicted harm on the victim (Krueger *et al.*, 2009a). In terms of the cognitive theory of mind network, our lesion results showed that the dmPFC was crucial for successful completion of TPP. Numerous studies have demonstrated that the dmPFC is associated with a variety of social inferences, including dispositional trait attributions about others (Harris *et al.*, 2005; Mitchell *et al.*, 2006), self-reference (Heatherton *et al.*, 2006; Moran *et al.*, 2006) and processing of social schemata (Krueger *et al.*, 2007, 2009a). A previous voxel-based morphometry study demonstrated that the dmPFC is strongly associated

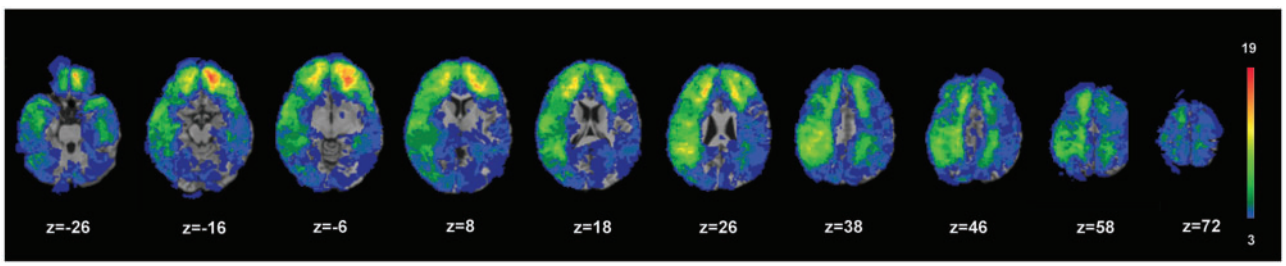

Fig. 1. Lesion overlay map. Axial slices in MNI space illustrating the number of overlapping lesions at each voxel across the whole pTBI population. The color bar indicates the number of overlapping lesions at each voxel. Red indicates a greater number of subjects and blue indicates a fewer number. The maximum overlap of 19 patients occurred in the prefrontal areas. Note that, in all slices, the right hemisphere is on the reader's left.

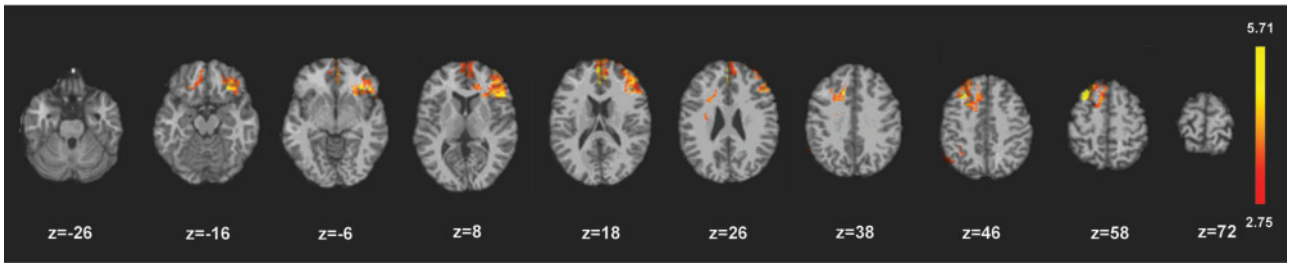

Fig. 2. VLSM analysis. Statistical map for whole-brain VLSM analysis using TPP performance as the dependent variable and the lesion status (gray matter) of each voxel as the independent variable (one-tailed *t*-test, *q*(FDR)=0.05, minimum cluster size of 50 voxels). Colorization indicates a significant relationship between lesion in voxel and TPP performance. Color range displays z-scores, from red (minimum z-score displayed on the right side) to yellow (maximum z-score). Axial slices display z-coordinates in MNI space. Note that in all slices, the right hemisphere is on the reader's left.

Table 3. VLSM results[a]

N	H	Anatomical label	All structures	Talairach			Peak z-score	Number of voxels
				x	y	z		
1	L	dmPFC-lPFC	Frontal middle Frontal superior Frontal inferior triangularis Frontal inferior operculum Frontal superior medial	−36	32	−6	5.71	3,622
2	R	vmPFC	Frontal superior orbital Frontal middle orbital	16	34	−20	3.66	175
3	R	dlPFC-SMA	Frontal superior Frontal middle Supplementary motor area	28	30	46	4.91	1,840
4	R	IPS	Parietal inferior Angular	48	−54	50	3.17	69

N, cluster; H, hemisphere; L, left; R, right; dmPFC, dorsomedial prefrontal cortex; lPFC, lateral prefrontal cortex; vmPFC; ventromedial prefrontal cortex; IPS, intraparietal sulcus.

[a]Brain regions necessary to perform third-party punishment rank-order task identified by whole-brain VLSM analysis (one-tailed *t*-test, *q*(FDR)=0.05, minimum cluster size of 50 voxels). Gray matter structures were obtained by applying the automated anatomical labeling (AAL) atlas.

Table 4. Lesion-behavior relationship[a]

N	H	Anatomical label	*r*	
			Altruism	Concept formation
1	L	dmPFC-lPFC (*n*=36)	0.33	0.62**
2	R	vmPFC (*n*=14)	0.32	0.34
3	R	dlPFC-SMA (*n*=23)	0.52***	0.34
4	R	IPS (*n*=7)	0.84	0.43

[a]Associations between TPP performance and neuropsychological control measures (i.e. altruism, concept formation skills) as identified from linear regression analysis for participants with lesions in identified brain clusters.

N, cluster; H, hemisphere; L, left; R, right; n, number of patients affected by lesions in each ROI; dmPFC, dorsomedial prefrontal cortex; lPFC, lateral prefrontal cortex; vlPFC, ventrolateral prefrontal cortex; dlPFC; dorsolateral prefrontal cortex; vmPFC; ventromedial prefrontal cortex; SMA, supplementary motor area; IPS, intraparietal sulcus.

**$P < 0.001$; ***$P < 0.01$.

with individuals' propensity for impartiality (i.e. the equal treatment of all human beings) during TPP (Baumgartner *et al.*, 2013): the larger the gray matter volume and thickness of the dmPFC, the more individuals show an impartial punishment pattern. While the TPJ is also thought to be strongly associated with social inferences of transitory states such as intentions of others (Saxe, 2006; Van Overwalle, 2009; Young *et al.*, 2010); we did not find a significant relationship between the TPJ and TPP. A previous functional neuroimaging study demonstrated that the TPJ is most strongly engaged by offenses where there was diminished responsibility to mitigate and reduced blameworthiness, such as having a mental illness or duress (Buckholtz and Marois, 2012). As our methodology did not split the scenarios up categorically according to responsibility or diminished responsibility scenarios, our task may not draw as strongly on social inferences of transitory states (e.g. intentions) to resolve ambiguity.

Regarding the affective theory of mind network, our lesion analysis revealed that the vmPFC was related to successful implementation of TPP, supporting previous functional neuroimaging studies where the vmPFC was implicated in moral judgments and normative decision making (de Quervain *et al.*, 2004; Moll *et al.*, 2005; Young *et al.*, 2010; Buckholtz and Marois, 2012). Lesion studies bolstered this presumed relationship, finding that damage to the vmPFC is related to impaired faux pas recognition (Leopold *et al.*, 2012), diminished emotional intelligence (Krueger *et al.*, 2009b) and irrational punishment during economic decision-making (Koenigs and Tranel, 2007). Moreover, these findings are echoed in populations with psychopathology and sociopathic tendencies (Fehr and Fischbacher, 2004b; Mendez, 2009). While the amygdala is also consistently associated with affective arousal and detecting harm (Buckholtz *et al.*, 2008; Buckholtz and Marois, 2012),

unfortunately due to the non-random dispersion of lesion locations in this study we did not achieve adequate lesion coverage to detect relations in that area.

Once the third-party decision maker has determined legal responsibility through the engagement of the mentalizing network, an appropriate punishment is determined through the central-executive network. This process consists of a context-specific representation of response space used to construct a scale of punishment (IPS) and selection of the appropriate magnitude of punishment and commensurate 'fairness' (dlPFC) (Knoch *et al.*, 2006; Buckholtz and Marois, 2012). In our study, lesions to the right IPS were related with diminished TPP. The IPS is often associated with numerical processing (Dehaene *et al.*, 2003; Krueger *et al.*, 2011), and may serve as a critical part of the central executive network as participants were asked to rank-order the criminal scenarios to reflect the relative degree of legal punishment that the perpetrator deserved. The IPS is thought to consist of a continuous mental number line that does not appear to be 'number-specific' but rather is involved in comparisons and estimations of both symbolic and non-symbolic quantities (Pia *et al.*, 2009). The IPS is also more active when conducting a comparative operation, for example comparing two numbers or two punishment situations, compared to simply reading them (Dehaene *et al.*, 2003). Therefore, the IPS may be creating a context-specific representation of the space in which to create a punishment scale to compare options, which then interacts with the dlPFC to perform the integration of the information and selection of punishment (Buckholtz and Marois, 2012).

Our lesion analysis identified the bilateral dlPFC to be associated with TPP (Kaller *et al.*, 2011, 2013). The R dlPFC has been consistently associated with altruistic punishment, valuation judgments and fairness (Greene *et al.*, 2004; Moll *et al.*, 2005; Guo *et al.*, 2013), whereas the L dlPFC has been found to be more related to executive function and impulse control (Ochsner *et al.*, 2002; Figner *et al.*, 2010; Barbey *et al.*, 2012). Altruism, a person's characteristic for sacrificing something for the well-being of others with no expectation of compensation nor direct or indirect benefits, was uniquely associated with TPP for veterans with lesions in the R dlPFC extending into the SMA. The altruism measure (Rushton Self Report Altruism Scale) is a series of 20 statements where one answers on a five-point Likert scale (never, once, more than once, often, very often) regarding different altruistic endeavors (e.g. 'I have donated blood', 'I have delayed an elevator and held the door open for a stranger') (Rushton *et al.*, 1981). In line with previous findings (Strobel *et al.*, 2011), self-reported altruism accounted for significant variance in the punishment paradigm indicating a systematic relationship wherein altruism co-varies with TPP.

In contrast, lesions in the L PFC (including dlPFC, vlPFC, and dmPFC) were related to concept formation skills, especially in terms of problem-solving skills and labeling of attributes in verbal and visual domains, the ability to inhibit previous responses, cognitive flexibility and the ability to elucidate and explain abstract concepts (Delis *et al.*, 2001; Latzman and Markon, 2010). Concept formation skills were assessed using an executive functioning measure that required individuals to divide six cards into two groups of three based on unique verbal/perceptual features, whereas the rank ordering TPP task required ordering based on magnitude. These disparate executive function tasks demonstrate both shared underlying cognitive mechanisms of basic comparison, sorting, planning and organizing, and distinct components such as magnitude judgment, ordering, value assessment and emotional processing.

Moreover, TMS studies support the dissociation of R and L dlPFC. It was found that disruption of the R, but not L, dlPFC, reduces the willingness of an individual to reject an intentionally unfair offer, even if they judge them as very unfair (Knoch *et al.*, 2006). Similarly, the disruption of the R, but not the L, dlPFC resulted in riskier decision-making. This further supports the dissociation for the R dlPFC for fairness-related behaviors, replicating previous findings that indicate the critical nature of the R dlPFC for strategic decision making (Van't Wout *et al.*, 2005) and altruistic punishment (Feng *et al.*, 2014).

In addition to our hypothesized neural correlates, we found two additional areas associated with TPP: vlPFC and SMA. We propose that although these areas are important in completing the TPP task, they are not inherently related to TPP. Previous research has shown that the L vlPFC is important for the cognitive control of memory and the ability to access past conceptual representation (Damasio, 1990; Badre and Wagner, 2007), whereas the SMA contributes to the control of movement important for implementing internally generated actions (Halsband *et al.*, 1994; Gerloff *et al.*, 1997).

Our results should be considered within certain limitations. First, as the lesions occurred in combat, inherently there is large variation of lesion size and location, which leads to potential limitations in averaging across a brain-injured group. Moreover, brain injuries were not randomly distributed; therefore non-independence of damage cannot be excluded and limited our coverage in important areas, such as the amygdala. Second, our population was entirely male, which limits the generalizability across genders, which is important to note given the potential functional differences in brain activity between sexes for decision-making tasks (Bolla *et al.*, 2004). Third, we used a rank-order task to assess TPP, which assesses the behavior of assigning relative punishment through rank-ordering (i.e. deciding that a crime deserves greater or lesser punishment compared with other crimes). While this is separate from actual sentencing or directly assigning punishment, it is a measure of TPP behavior as one must conceptualize the crime, evaluate the intent and harm, and then 'assign' relative punishment based on other scenarios relying on relative normative judgment. To help validate this measure we compared it to a TPP rating task of directly assigning punishment as was used in previous functional neuroimaging studies. Our pilot study in healthy adults demonstrated that both TPP paradigms (rating and rank-order) correlated very highly, indicating that they both measure the degree of punishment that the perpetrator deserved for each offense. However, future studies should replicate our findings with a rating task in brain-injured people to ensure equivalency in both clinical and non-clinical populations. Furthermore, since both TPP tasks rely on self-report measures, i.e. participants made third-party decisions regarding fictional characters of criminal scenarios, where their decisions involved no costs for themselves, future lesion studies should clarify if the reported findings are similar to third-party decision-making in the context of economic exchange utilization (e.g. economic game paradigms with monetary compensation). Despite these limitations, this study offered several strengths including the advantage of the opportunity to examine a large number of brain-injured veterans with discrete lesions and a control population with similar combat experience. Moreover, our population offered the availability of pre-injury general intelligence and the existence of stable and relatively uniform low-velocity focal lesions covering a large area of the brain that provides unique information about the effects of pTBI on cognitive function.

In summary, a growing literature has shed light on the underlying neural correlates of normative decision-making; however, the neural signatures associated with TPP have not been fully elucidated despite the societal importance. Using a lesion-based approach, we were able to provide convergent evidence for the neurocognitive hypothesis of TPP, proposing two domain-general networks (Buckholtz and Marois, 2012): the mentalizing network (dmPFC, vmPFC) to evaluate the intention of the individual committing the act and the actual harm inflicted onto the victim of the act and the central-executive network (dlPFC, R IPS) to determine a suitable punishment given the magnitude of the crime. Within clinical populations, understanding the networks associated with TPP and punishment of norm violations may also reveal greater knowledge regarding the underlying neuropathology of behavioral and cognitive deficits associated with determining right from wrong and theory of mind, such as antisocial personality disorder or psychopathy, as they tend to have overlapping implicated brain regions (Glenn and Raine, 2008; Koenigs *et al.*, 2011).

In conclusion, our findings have clinically relevant implications in these populations as early-onset PFC damage can result in syndromes that resemble psychopathy, despite having normative cognitive abilities (Koenigs *et al.*, 2011). These findings may also extend into the broader area of normative judgment and decision making and may help to inform the legal criminal system in terms of understanding psychopathy and influence rehabilitation and reduced recidivism.

Acknowledgements

We thank the NIH Clinical Center for the provision of their facilities and for their supportive services. We thank Sandy Bonifant, Michael Tierney, Lyanne Yozawitz, Olga Dal Monte, Carolee Noury, Vivien YJ Tsen, Anne Leopold and Selene Schintu, who worked tirelessly to test subjects and organize the study. As always, the authors are grateful to all of the Vietnam veterans and caregivers who participated in this study. Their unending commitment to improving the health care of veterans is the reason this study could be completed.

Funding

This study was conducted and supported by the U.S. National Institute of Neurological Disorders and Stroke Intramural Research Program and took place at the National Institutes of Health Clinical Research Center. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the US Government. For further information about the Vietnam Head Injury Study, contact J. G. at jgrafman@northwestern.edu. Additional support was provided by National Institute on Alcohol Abuse and Alcoholism (NIAAA) grant number F31 AA022261.

Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

References

Badre, D., Wagner, A.D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, **45**(13), 2883–901.

Barbey, A.K., Colom, R., Grafman, J. (2014). Neural mechanisms of discourse comprehension: a human lesion study. *Brain*, **137**(Pt 1), 277–87.

Barbey, A.K., Colom, R., Solomon, J., Krueger, F., Forbes, C., Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain*, **135**(Pt 4), 1154–64.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *J Child Psychol Psychiatry*, **38**(7), 813–22.

Baumgartner, T., Schiller, B., Hill, C., Knoch, D. (2013). Impartiality in humans is predicted by brain structure of dorsomedial prefrontal cortex. *Neuroimage*, **81**, 317–24.

Bayroff, A.G., Anderson, A.A. (1963). Development of the armed forces qualification test 7 and 8. United States Army Personnel Research Office.

Beck, A., Steer, R.A., Brown, G. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Bolla, K.I., Eldreth, D.A., Matochik, J.A., Cadet, J.L. (2004). Sex-related differences in a gambling task and its neurological correlates. *Cerebral Cortex*, **14**(11), 1226–32.

Buckholtz, J.W., Asplund, C.L., Dux, P.E., *et al.* (2008). The neural correlates of third-party punishment. *Neuron*, **60**(5), 930–40.

Buckholtz, J.W., Marois, R. (2012). The roots of modern justice: cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, **15**(5), 655–61.

Chatterjee, A. (2005). A madness to the methods in cognitive neuroscience? *Journal of Cognitive Neuroscience*, **17**(6), 847–9.

Damasio, A.R. (1990). Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends in Neurosciences*, **13**(3), 95–8.

Davis, M. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, **44**(1), 1130126.

de Quervain, D.J., Fischbacher, U., Treyer, V., *et al.* (2004). The neural basis of altruistic punishment. *Science*, **305**(5688), 1254–8.

Dehaene, S., Piazza, M., Pinel, P., Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, **20**(3), 487–506.

Delis, D.C., Kaplan, E., Kramer, J.H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: The Psychological Corporation.

Dimitrov, M., Grafman, J., Hollnagel, C. (1996). The effects of frontal lobe damage on everyday problem solving. *Cortex*, **32**(2), 357–66.

Fehr, E., Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Science*, **8**(4), 185–90.

Fehr, E., Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior*, **25**(2), 63–87.

Feng, C., Luo, Y.J., Krueger, F. (2014). Neural signatures of fairness-related normative decision making in the ultimate game: a coordinate-based meta-analysis. *Human Brain Mapping*, **36**(2), 591–602.

Figner, B., Knoch, D., Johnson, E.J., *et al.* (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, **13**(5), 538–9.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, **32**(3), 221–33.

Flesch, R., Gould, A.J. (1949). *The Art of Readable Writing*. New York: Harper.

Frith, C.D., Frith, U. (1999). Interacting minds—a biological basis. *Science*, **286**(5445), 1692–5.

Gerloff, C., Corwell, B., Chen, R., Hallett, M., Cohen, L.G. (1997). Stimulation over the human supplementary motor area interferes with the organization of future elements in complex motor sequences. *Brain*, **120** (Pt 9), 1587–602.

Glenn, A.L., Raine, A. (2008). The neurobiology of psychopathy. *Psychiatric Clinics of North America*, **31**(3), 463–75, vii.

Grafman, J., Jonas, B.S., Martin, A., *et al.* (1988). Intellectual function following penetrating head injury in Vietnam veterans. *Brain*, **111** (Pt 1), 169–84.

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.

Guo, X., Zheng, L., Zhu, L., *et al.* (2013). Increased neural responses to unfairness in a loss context. *Neuroimage*, **77**, 246–53.

Halsband, U., Matsuzaka, Y., Tanji, J. (1994). Neuronal activity in the primate supplementary, pre-supplementary and premotor cortex during externally and internally instructed sequential movements. *Neuroscience Research*, **20**(2), 149–55.

Harris, L.T., Todorov, A., Fiske, S.T. (2005). Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of mind. *Neuroimage*, **28**(4), 763–9.

Heatherton, T.F., Wyland, C.L., Macrae, C.N., Demos, K.E., Denny, B.T., Kelley, W.M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience*, **1**(1), 18–25.

Heberlein, A.S., Adolphs, R., Tranel, D., Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. *Journal of Cognitive Neuroscience*, **16**(7), 1143–58.

Kalbe, E., Schlegel, M., Sack, A.T., *et al.* (2010). Dissociating cognitive from affective theory of mind: a TMS study. *Cortex*, **46**(6), 769–80.

Kaller, C.P., Heinze, K., Frenkel, A., *et al.* (2013). Differential impact of continuous theta-burst stimulation over left and right DLPFC on planning. *Human Brain Mapping*, **34**(1), 36–51.

Kaller, C.P., Rahm, B., Spreer, J., Weiller, C., Unterrainer, J.M. (2011). Dissociable contributions of left and right dorsolateral prefrontal cortex in planning. *Cerebral Cortex*, **21**(2), 307–17.

Kaplan, E., Goodglass, H., Weintraub, S. (2001). *Boston Naming Test-2 (BNT-2)*. Austin, TX: Pro-Ed.

Knoch, D., Nitsche, M.A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cerebral Cortex*, **18**(9), 1987–90.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, **314**(5800), 829–32.

Koenigs, M., Baskin-Sommers, A., Zeier, J., Newman, J.P. (2011). Investigating the neural correlates of psychopathy: a critical review. *Molecular Psychiatry*, **16**(8), 792–9.

Koenigs, M., Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *Journal of Neuroscience*, **27**(4), 951–6.

Krueger, F., Barbey, A.K., Grafman, J. (2009a). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, **13**(3), 103–9.

Krueger, F., Barbey, A.K., McCabe, K., *et al.* (2009b). The neural bases of key competencies of emotional intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(52), 22486–91.

Krueger, F., Hoffman, M., Walter, H., Grafman, J. (2014). An fMRI investigation of the effects of belief in free will on third-party punishment. *Social Cognitive and Affective Neuroscience*, **9**(8), 1143–9.

Krueger, F., Landgraf, S., van der Meer, E., Deshpande, G., Hu, X. (2011). Effective connectivity of the multiplication network: a functional MRI and multivariate Granger Causality Mapping study. *Human Brain Mapping*, **32**(9), 1419–31.

Krueger, F., Moll, J., Zahn, R., Heinecke, A., Grafman, J. (2007). Event frequency modulates the processing of daily life activities in human medial prefrontal cortex. *Cerebral Cortex*, **17**(10), 2346–53.

Krueger, F., Parasuraman, R., Moody, L., *et al.* (2013). Oxytocin selectively increases perceptions of harm for victims but not the desire to punish offenders of criminal offenses. *Social Cognitive and Affective Neuroscience*, **8**(5), 494–8.

Latzman, R.D., Markon, K.E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment*, **17**(2), 172–84.

Leopold, A., Krueger, F., dal Monte, O., *et al.* (2012). Damage to the left ventromedial prefrontal cortex impacts affective theory of mind. *Social Cognitive and Affective Neuroscience*, **7**(8), 871–80.

Makale, M., Solomon, J., Patronas, N.J., Danek, A., Butman, J.A., Grafman, J. (2002). Quantification of brain lesions using interactive automated software. *Behavior Research Methods Instruments and Computers*, **34**(1), 6–18.

McNeil, M.R., Prescott, T.E. (1978). *Revised Token Test*. Austin, TX: Pro-Ed.

Mendez, M.F. (2009). The neurobiology of moral behavior: review and neuropsychiatric implications. *CNS Spectrums*, **14**(11), 608–20.

Mitchell, J.P., Macrae, C.N., Banaji, M.R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, **50**(4), 655–63.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J. (2005). Opinion: the neural basis of human moral cognition. *Nature Reviews Neuroscience*, **6**(10), 799–809.

Moran, J.M., Macrae, C.N., Heatherton, T.F., Wyland, C.L., Kelley, W.M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *Journal of Cognitive Neuroscience*, **18**(9), 1586–94.

Ochsner, K.N., Bunge, S.A., Gross, J.J., Gabrieli, J.D. (2002). Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *Journal of Cognitive Neuroscience*, **14**(8), 1215–29.

Pia, L., Corazzini, L.L., Folegatti, A., Gindri, P., Cauda, F. (2009). Mental number line disruption in a right-neglect patient after a left-hemisphere stroke. *Brain and Cognition*, **69**(1), 81–8.

Poldrack, R.A. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, **72**(5), 692–7.

Premack, D., Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, **1**(4), 515–26.

Raymont, V., Salazar, A.M., Krueger, F., Grafman, J. (2011). "Studying injured minds"—the Vietnam head injury study and 40 years of brain injury research. *Fronters in Neurology*, **2**, 15.

Robinson, P., Kurzban, R. (2007). Concordance and conflict in institutions of justice. *Minnesota Law Review*, **91**, 1829–907.

Rushton, J.P., Chrisjohn, R.D., Fekken, C.G. (1981). The altruistic personality and self-report altruism scale. *Personality and Individual Differences*, **2**(4), 293–302.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, **16**(2), 235–39.

Schleim, S., Spranger, T.M., Erk, S., Walter, H. (2011). From moral to legal judgment: the influence of normative context in lawyers and other academics. *Social Cognitive and Affective Neuroscience*, **6**(1), 48–57.

Shamay-Tsoory, S.G., Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, **45**(13), 3054–67.

Shen, F.X., Hoffman, M.B., Jones, O.D., Greene, J.D., Marois, R. (2011). Sorting guilty minds. *New York University Law Review*, **86**(5), 1306–60.

Solomon, J., Raymont, V., Braun, A., Butman, J.A., Grafman, J. (2007). User-friendly software for the analysis of brain lesions (ABLe). *Computer Methods and Programs in Biomedicine*, **86**(3), 245–54.

Spielberger, C.D. (1999) *The State-Trait Anger Expression Inventory-2 (STAXI-2): Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Stein, M.B., Simmons, A.N., Feinstein, J.S., Paulus, M.P. (2007). Increased amygdala and insula activation during emotion processing in anxiety-prone subjects. *The American Journal of Psychiatry*, **164**(2), 318–27.

Strobel, A., Zimmermann, J., Schmitz, A., *et al*. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. *Neuroimage*, **54**(1), 671–80.

Sultzer, D.L., Levin, H.S., Mahler, M.E., High, W.M., Cummings, J.L. (1992). Assessment of cognitive, psychiatric, and behavioral disturbances in patients with dementia: the Neurobehavioral Rating Scale. *Journal of the American Geriatrics Society*, **40**(6), 549–55.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., *et al*. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, **15**(1), 273–89.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, **30**(3), 829–58.

van't Wout, M., Kahn, R.S., Sanfey, A.G., Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport*, **16**(16), 1849–52.

Wechsler, D. (1997). *The Wechsler Memory Scale*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

Woods, R.P., Mazziotta, J.C., Cherry, S.R. (1993). MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, **17**(4), 536–46.

Yamada, M., Camerer, C.F., Fujie, S., *et al*. (2012). Neural circuits in the brain that are activated when mitigating criminal sentences. *Nature Communications*, **3**, 759.

Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(15), 6753–8.