



Effective connectivity of brain regions underlying third-party punishment: Functional MRI and Granger causality evidence

Gabriele Bellucci, Sergey Chernyak, Morris Hoffman, Gopikrishna Deshpande, Olga Dal Monte, Kristine M. Knutson, Jordan Grafman & Frank Krueger

To cite this article: Gabriele Bellucci, Sergey Chernyak, Morris Hoffman, Gopikrishna Deshpande, Olga Dal Monte, Kristine M. Knutson, Jordan Grafman & Frank Krueger (2017) Effective connectivity of brain regions underlying third-party punishment: Functional MRI and Granger causality evidence, *Social Neuroscience*, 12:2, 124-134, DOI: [10.1080/17470919.2016.1153518](https://doi.org/10.1080/17470919.2016.1153518)

To link to this article: <https://doi.org/10.1080/17470919.2016.1153518>



View supplementary material [↗](#)



Accepted author version posted online: 04 Mar 2016.
Published online: 10 Mar 2016.



Submit your article to this journal [↗](#)



Article views: 422



View Crossmark data [↗](#)



Citing articles: 13 View citing articles [↗](#)

Effective connectivity of brain regions underlying third-party punishment: Functional MRI and Granger causality evidence

Gabriele Bellucci^{a,b}, Sergey Chernyak^a, Morris Hoffman^{c,d}, Gopikrishna Deshpande^{e,f}, Olga Dal Monte^g, Kristine M. Knutson^h, Jordan Grafmanⁱ and Frank Krueger^{a,j}

^aMolecular Neuroscience Department, George Mason University, Fairfax, VA, USA; ^bBerlin School of Mind and Brain, Humboldt-Universitaet zu Berlin, Berlin, Germany; ^cSecond Judicial District, Denver, CO, USA; ^dJohn D. and Catherine T. MacArthur Foundation's Research Network on Law and Neuroscience, Nashville, TN, USA; ^eAU MRI Research Center, Department of Electrical and Computer Engineering, Auburn University, Auburn, AL, USA; ^fDepartment of Psychology, Auburn University, Auburn, AL, USA; ^gDepartment of Psychology, Yale University, New Haven, CT, USA; ^hBehavioral Neurology Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA; ⁱBrain Injury Research Program, Rehabilitation Institute of Chicago, Chicago, IL, USA; ^jDepartment of Psychology, George Mason University, Fairfax, VA, USA

ABSTRACT

Third-party punishment (TPP) for norm violations is an essential deterrent in large-scale human societies, and builds on two essential cognitive functions: evaluating legal responsibility and determining appropriate punishment. Despite converging evidence that TPP is mediated by a specific set of brain regions, little is known about their effective connectivity (direction and strength of connections). Applying parametric event-related functional MRI in conjunction with multivariate Granger causality analysis, we asked healthy participants to estimate how much punishment a hypothetical perpetrator deserves for intentionally committing criminal offenses varying in levels of harm. Our results confirmed that TPP legal decisions are based on two domain-general networks: the mentalizing network for evaluating legal responsibility and the central-executive network for determining appropriate punishment. Further, temporal pole (TP) and dorsomedial prefrontal cortex (PFC) emerged as hubs of the mentalizing network, uniquely generating converging output connections to ventromedial PFC, temporo-parietal junction, and posterior cingulate. In particular, dorsomedial PFC received inputs only from TP and both its activation and its connectivity to dorsolateral PFC correlated with degree of punishment. This supports the hypothesis that dorsomedial PFC acts as the driver of the TPP activation pattern, leading to the decision on the appropriate punishment. In conclusion, these results advance our understanding of the organizational elements of the TPP brain networks and provide better insights into the mental states of judges and jurors tasked with blaming and punishing legal wrongs.

ARTICLE HISTORY

Received 28 July 2015
Accepted 6 February 2016
Published online
10 March 2016

KEYWORDS



Third-party punishment;
social cognition; prefrontal
cortex; temporal pole;
effective connectivity

Introduction


Third-party punishment (TPP) in response to social norm violations is a social behavior that enables security and cooperation in modern large-scale societies of genetically heterogeneous people (Fehr & Fischbacher, 2004a; Hoffman & Goldsmith, 2004). All modern criminal justice systems are systems of impartial TPP, because deliberations about guilt and acquittal are entrusted to a third party—the jury and judge—rather than to the victim of the crime. With few exceptions, to be held legally responsible under criminal law, perpetrators must have committed a prohibited act (*actus reus*) inflicting actual harm on the victim and must

have acted with a guilty intent (*mens rea*) (Shen, Hoffman, Jones, Greene, & Marois, 2011). TPP in criminal contexts includes two essential cognitive mechanisms: evaluation of legal responsibility and determination of appropriate punishment (Buckholtz et al., 2008).

Previous neuroimaging studies investigating moral and legal decision-making have demonstrated that the cognitive mechanisms required for those kinds of decisions build on two domain-general brain networks (Buckholtz et al., 2008; Krueger, Hoffman, Walter, & Grafman, 2013; Schleim, Spranger, Erk, & Walter, 2011; Yamada et al., 2012). The mentalizing network is engaged to evaluate responsibility (i.e., default-mode network, e.g., medial prefrontal cortex (mPFC), posterior

CONTACT Frank Krueger  FKrueger@gmu.edu  Molecular Neuroscience Department/ Department of Psychology, George Mason University, 4400 University Drive, Mail Stop 2A1, Fairfax, VA 22030, USA

F.K. and J.G. designed research; F.K., O.M., and K.K. performed research; F.K., G.B., S.C., and G.D. analyzed data; G.D. provided analysis tools; G.B., S.C., M.H., and F.K. wrote the paper. All authors discussed the results and commented on the manuscript.

 Supplemental data for this article can be accessed [here](#).

© 2016 Taylor & Francis

cingulate (PC), temporal pole (TP), temporo-parietal junction (TPJ)), whereas the central-executive network is engaged to determine the appropriate punishment (e.g., dorsolateral prefrontal cortex (dlPFC), posterior parietal cortex) (Buckholz & Marois, 2012). Despite converging evidence that these networks mediate TPP, little is known about the effective connectivity (direction and strength of connections) of the underlying brain regions.

The goal of this study was to investigate the effective connectivity of the TPP networks by combining parametric event-related functional MRI (fMRI) with multivariate Granger causality analysis (GCA). We asked healthy participants to determine how much punishment a hypothetical perpetrator named “John” deserved for having committed intentional offenses ranging from property destruction and theft to kidnapping, rape, and murder (Robinson & Darley, 2007; Robinson & Kurzban, 2007). Based on a proposed neurocognitive model of impartial TPP (Buckholz & Marois, 2012), we tested the hypothesis that mPFC is the driver of TPP brain regions in that it evaluates information about the criminal intent of the perpetrator and is coupled with dlPFC for determination of the deserved punishment. First, we attempted to replicate the core brain regions previously reported for TPP. Second, we parametrically analyzed the linear relationships between degree of punishment and brain activations. Finally, we applied multivariate GCA to capture the effective connectivity among the neural components of TPP.

Materials and methods

Subjects

Twenty-six normal healthy native English speakers (13 females, 13 males, mean years of age: 26.0 ± 5.7 , mean years of education: 16.9 ± 2.6) participated in the study for financial compensation. Participants were right-handed, as determined by the Edinburgh Handedness Inventory (mean \pm SD: 93.0 ± 10.4), and had normal or corrected-to-normal vision. None of the participants was taking medication and a neurological examination by a board-certified neurologist during the previous 12 months determined no history of medical, psychiatric, or neurological diagnoses. Based on a preliminary brief psychological survey, participants were excluded from this study if they have had any physical or emotional trauma involving either injury or threat of injury to themselves or a close family member/friend, or been the victim of or had witnessed a violent crime. The study has been conducted according to the principles

expressed in the Declaration of Helsinki. Participants gave written informed consent after understanding the study procedures, which were approved by the National Institute of Neurological Disorders and the Stroke Institutional Review Board in Bethesda, MD.

Stimuli and experimental task

A set of criminal scenarios ($n = 44$), consisting of two- to three-sentence vignettes, was used as stimuli for the experiment (Robinson & Kurzban, 2007) (Supplemental data, Table 1). Scenarios featured a hypothetical perpetrator named “John”, who always purposefully engaged in criminal offenses, which varied in harm inflicted on the victim from property destruction and theft to kidnapping, rape, and murder.

The E-Prime (Psychology Software Tools, Sharpsburg, PA, USA) software was used for stimuli presentation. In each trial, a fixation cross (+) (0.5 s) was presented followed by a vignette (i.e., header and scenario) (reading phase = 12 s) (Supplemental data, Figure 1). Then, the type of judgment task (legal task or syllable task) and a Likert scale were presented. In the legal task (experimental condition), participants were asked to determine how much punishment John deserved (Likert scale: 0 = no punishment to 100 = extreme punishment). In the syllable task (control condition), participants were asked to estimate (but not count) the number of syllables in each vignette (Likert scale: approximately 30 to approximately 95). Participants were instructed to respond as quickly and accurately as possible and their response times and ratings were recorded. Two fiber optic response pads (Lumina LP-400 System, Cedrus Corporation, San Pedro, CA, USA) were provided on which participants placed the index and middle fingers of their right and left hands. They answered by pressing with either their right index (to move left) or right middle finger (to move right) until the marker reached the desired scale-value and then pressing with their left index finger to mark their final decision (decision phase = 6 s). Note that the marker was always initially placed in the middle of the Likert scale. Finally, before the next trial, a blank screen with two fixation crosses (++) was displayed for a jittered inter-stimulus interval (jitter mean = 4 s, range 2–6 s). The fMRI experiment consisted of three runs and each vignette was presented twice—once in the experimental condition and once in the control condition—but never within the same run.

Before the scan, participants familiarized themselves with the task using a separate set of stimuli ($n = 4$; Supplemental data, Table 2). After the scan, they completed psychological surveys in randomized order to (i) list what kind of punishment they imagined for

punishment scores of 1, 25, 50, 75, and 100 and (ii) rate the affect elicited by the criminal vignettes on a 5-point Likert scale (valence: 1 = negative to 5 = positive; arousal: 1 = low to 5 = high) of the Self-Assessment Manikin instrument (Bradley & Lang, 1994). The fMRI experiment lasted for approximately 2 hours (1 hour for scanning and 1 hour for post-scan questionnaires).

Data acquisition

Structural and functional data were collected on a 1.5 Tesla GE MRI whole-body scanner equipped with an 8-channel receiver head coil (General Electric, Milwaukee, WI, USA). Anatomical (T1-weighted 3D MP-RAGE sequence: TR, 8.9 ms; flip angle, 12°; number of slices, 124; FOV, 240 mm; matrix size, 256 × 256; voxel size, 1 × 1 × 1 mm³) and functional (2D gradient EPI sequence: TR, 2000 ms; TE, 28 ms; flip angle, 90°; thickness, 3.5 mm; number of slices, 31; FOV, 240 mm; matrix size, 64 × 64) images were acquired. Functional images were taken parallel to the AC-PC line and the first five volumes were discarded due to T1 equilibration effects.

Data analysis

Behavioral analysis

Behavioral analyses were carried out in SPSS 22.0 (Statistical Package for the Social Sciences, Armonk, NY: IBM Corp.). *P* value was set to *P* < 0.05 (two-tailed). Data were normally distributed (Kolmogorov–Smirnov test) and assumptions for analyses of variance (Bartlett's test) were not violated. Bivariate Pearson correlation analyses were performed among participants' reaction times, punishment ratings, and affective ratings to explore relationships between these experimental measures.

fMRI analysis

BrainVoyager QX 2.0 (Brain Innovation, Maastricht, The Netherlands) was used to perform fMRI data analyses. Preprocessing of anatomical data included reassembling into 1 mm resolution and normalizing into Talairach space using a piecewise linear transformation. Preprocessing of functional data included slice-scan time correction (sinc interpolation), small head movement correction by spatially aligning all volumes to the first volume (rigid body transformation), removal of linear trends and low-frequency non-linear drifts of three or fewer cycles from the time series (temporal high-pass filtering), and spatial smoothing of the functional images (Gaussian filter of 8 mm full-width at half-

maximum). Functional data were co-registered with the individual's 3D anatomical images and then reassembled into 3 × 3 × 3 mm³ isotropic voxels.

A General Linear Model (GLM) corrected for first-order serial correlation was applied (K. J. Friston, Holmes, & Worsley, 1999). Random-effect analyses were performed at the multi-subject level to explore brain regions associated with the decision phases of the experimental and control conditions. The GLM consisted of 11 regressors: two categorical regressors for the main scenario reading phase (Experimental, Control), two categorical regressors for the main decision phase (Experimental, Control), one parametric regressor for the degree of punishment (Punishment), and six parametric regressors of no interest for the 3D motion correction (translations in X, Y, Z directions, rotations around X, Y, Z axes). The regressor time courses were adjusted for the hemodynamic response delay by convolution with a dual-gamma hemodynamic response function (Buchel, Holmes, Rees, & Friston, 1998).

Two statistical models were fit at the multi-subject level after computing the coefficients (betas) for all regressors: (i) main modulation (Decision Phase: Experimental > Control) to explore brain activation associated with TPP and (ii) parametric modulation (Experimental > Control ∩ (in conjunction with) Parameter: Punishment) to reveal linear relationship (positive, negative) between the degree of punishment and blood oxygenation level-dependent (BOLD) responses of brain regions. Note that the conjunction for the parametric modulation (Experimental > Control ∩ Punishment) provided the specificity about the linear parametric effect and requires a significant effect in each of these two contrasts ("minimum *t*-statistic") (Nichols, Brett, Andersson, Wager, & Poline, 2005).

For the main modulation, a whole brain analysis approach with a threshold of *q*(FDR) < 0.05 (*k* > 10) was chosen (Genovese, Lazar, & Nichols, 2002) to determine without an a priori hypothesis the underlying regions of interest (ROIs) of the TPP network. For the parametric modulation, a cluster-level statistical threshold corrected for multiple comparisons was determined by calculating the minimum cluster size to achieve a false activation probability $\alpha = 0.05$. The cluster-level statistical threshold approach was based on a 3D extension of the randomization procedure described in Forman et al. (1995). First, the voxel-level threshold was set at *t* (25) = 3.08 (*P* < 0.005, uncorrected) for the parametric contrast. Then, the thresholded map was submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure (Monte Carlo simulation) for estimating

cluster level false-positive rates. The implemented method corrected for multiple cluster tests across space. For each simulated image, all “active” clusters in the imaged volume were considered and used to update a table reporting the counts of all the clusters above this threshold for each specific size. After 1,000 iterations, an α -value was assigned to each cluster size based on its observed relative frequency. From this information, the minimum cluster size threshold was specified to yield a cluster-level false-positive rate of $\alpha = 5\%$. The estimated minimum cluster size of 135 mm^3 was applied to the statistical maps for the parametric contrasts.

Brodmann areas were determined by employing the Talairach Daemon Client software (Research Imaging Center, San Antonio, TX, USA) and the coplanar stereotaxic atlas of the human brain (Talairach & Tournoux, 1988). For display purposes, statistical images were overlaid onto a surface brain map in Talairach space.

Effective connectivity analysis

Effective connectivity analysis was implemented using multivariate GCA as a statistical method for assessing directional influences among simultaneously recorded BOLD time series (C. W. J. Granger, 1969; Deshpande, Sathian, & Hu, 2010; Roebroeck, Formisano, & Goebel, 2005). The GCA concept draws on the principle of temporal predictability, which assumes that if the current temporal progression of brain activity in one brain region allows the prediction of future temporal progression of activity in another brain region, then the first brain region is assumed to have a causal influence on the second brain region (C. Granger, 1969). GCA is a data-driven approach statistical model, which is capable of determining directional influences among BOLD-signal time series of brain regions without a priori brain connectivity models as opposed to more traditional hypothesis-driven approaches such as dynamic causal modeling (DCM) (Deshpande & Hu, 2012; Deshpande, LaConte, James, Peltier, & Hu, 2009; K. Friston, Moran, & Seth, 2013; Roebroeck et al., 2005).

In recent years, theoretical developments and simulations (Katwal, Gore, Gatenby, & Rogers, 2013; Sreenivasan, Havlicek, & Deshpande, 2015; Wen, Rangarajan, & Ding, 2013), along with experimental studies (Hutcheson et al., 2015; Krueger, Landgraf, van der Meer, Deshpande, & Hu, 2011; Sathian, Deshpande, & Stilla, 2013), have demonstrated the utility and validity of GCA. For example, an fMRI study employing a simple auditory-motor paradigm (Abler et al., 2006) demonstrated that Granger causality correctly estimates

expected causal influences from the auditory cortex to the motor cortex even with data acquired with long TRs (2,440 ms in their study). Moreover, previous evidence indicates that hemodynamic deconvolution removes the intersubject and interregional variability of the hemodynamic response function (Handwerker, Ollinger, & D’Esposito, 2004), thus increasing the effective temporal resolution of the signal. Finally, effective connectivity inferred from GCA and DCM on deconvolved fMRI data agreed with those obtained from intra-cerebral EEG, indicating converging evidence from these different methods (David et al., 2008).

For the GCA, only those regions that survived the fMRI analysis threshold of the contrast Experimental > Control were selected as ROIs. The ROIs were limited round the peak of activation with a maximum size of $6 \times 6 \times 6 \text{ mm}^3$. For the selected ROIs, the time series of the BOLD signal intensities from both the experimental and control conditions were extracted and averaged across voxels and then normalized across participants per run. Finally, the normalized times series from the decision phase were selected and collapsed across all runs and participants to obtain a single time series per ROI for the experimental and control conditions (Deshpande et al., 2009).

A multivariate autoregressive (MVAR) approach was employed to assess directed interactions between multiple nodes while factoring out influences mediated indirectly in the set of regions selected (Deshpande, Hu, Stilla, & Sathian, 2008; Deshpande et al., 2009; Havlicek, Friston, Jan, Brazdil, & Calhoun, 2011). The MVAR model was applied to hidden neuronal variables obtained after parametric hemodynamic deconvolution of the extracted BOLD time series from selected ROIs (Havlicek et al., 2011). The MVAR model’s coefficients were allowed to vary as a function of time to obtain condition-specific connectivity values (Grant et al., 2014, 2015; Sathian et al., 2013).

Boxcar functions that represented the decision phases of the legal and syllable tasks were used to extract connectivity values corresponding to the experimental and control conditions. The resulting dynamic GC path weights were populated into two samples, and paired t -tests ($q(\text{FDR}) < 0.05$) were performed to assess the condition-specific modulations of connectivity (Experimental > Control). Bivariate Spearman correlations were computed to determine the associations between subjects’ punishment ratings and connectivity values among ROIs. BrainNet Viewer as a graph-theoretical network visualization toolbox under MATLAB (www.mathworks.com) was used to display the effective connectivity networks (i.e., nodes, edges) on a brain surface (Xia, Wang, & He, 2013).

Results

Behavioral results

For the experimental task, the punishment ratings (Supplemental data, Figure 2) were independent of the response times (mean \pm SD, 3713 \pm 497 ms) ($r = -0.22$, $P = 0.291$). In contrast, the syllable ratings (63.4 \pm 10.7) were associated with the number of syllables contained in the legal vignettes ($r = 0.96$, $P < 0.001$) for the control task, but were independent of the response times (3419 \pm 515 ms) ($r = -0.06$, $P = 0.767$). After scanning, participants listed the kind of punishment they imagined during the experiment and demonstrated a strong agreement about their internal scale of justice for selected punishment scores (1, 25, 50, 75, and 100): low punishment scores for financial or social crimes (1, 25), higher punishment scores were associated with longer jail times (50, 75), and the highest punishment scores (100) led to life imprisonment or the death penalty (Supplemental data, Table 3). Furthermore, participants' punishment ratings were significantly correlated with valence and arousal ratings, indicating that the higher the punishment ratings, the more negative the valence ratings ($r = 0.89$, $P < 0.001$) and the higher the arousal ratings ($r = -0.92$, $P < 0.001$) (Supplemental data, Figure 3).

fMRI results

A GLM analysis was applied to identify those brain regions activated by TPP decisions. The contrast for the main modulation (Decision Phase: Experimental > Control) revealed activations in the two hypothesized brain networks. For the mentalizing network, we found activations in ventromedial PFC ((vmPFC), medial frontal gyrus (MFG), Brodmann Area (BA 9)), dorsomedial PFC ((dmPFC), MFG, BA 10/32), PCC (BA 31), left TP (middle temporal gyrus; BA 21/38), and left TPJ (superior temporal gyrus; BA 39), whereas for the central-executive network, we found activations in left dlPFC (inferior frontal gyrus; BA 9/47) (Table 1, Figure 1A). To explore the parametric modulation of the TPP

activation pattern, analysis of linear relationships between BOLD responses and punishment ratings revealed a negative correlation between degree of punishment and dmPFC activation (MFG; BA 9; peak coordinates: $x = 5$, $y = 41$, $z = 23$; peak activation: $t(25) = -3.89$; $P < 0.05$, cluster-threshold correction) (Figure 1B).

Effective connectivity results

Finally, multivariate GCA was applied to investigate the effective connectivity among brain regions identified in the previous main modulation as significantly more active during TPP decisions (Figure 2, Table 2). We found that left TP and dmPFC emerged as hubs of the TPP networks. The dmPFC hub received input only from left TP and both hubs sent outputs to all other regions. Moreover, dmPFC bound the two networks with a connection to dlPFC that was marginally significantly positive correlated with participants' degree of punishment ($\rho = 0.39$, $P = 0.051$).

Discussion

We combined parametric event-related fMRI with multivariate GCA to investigate the effective connectivity of the neural correlates of TPP. We asked healthy participants to estimate how much punishment a perpetrator deserved for intentionally committing hypothetical criminal offenses varying in harm levels. Our results demonstrated that left TP and dmPFC behave as hubs of the TPP activation pattern in that they were the only regions sending outputs to all others (vmPFC, dlPFC, TPJ, PCC). Receiving input only from the TP hub, dmPFC represented the driver of the TPP brain regions, specifically mediating the punishment decisions.

Behavioral results

TPP decisions are considered to rely on two principal cognitive mechanisms: evaluation of legal responsibility and selection of proper punishment (Buckholtz & Marois, 2012). Our post-scan debriefing results demon-

Table 1. Brain regions activated for TPP decisions.

Region	Anatomical area	Brodmann area (BA)	Talairach coordination of local maxima (mm)			Local maxima
			x	y	z	T
vmPFC	Medial frontal gyrus (MFG)	9	-5	58	5	7.25
dmPFC	Medial frontal gyrus (MFG)	10/32	-4	48	26	9.35
PCC	Posterior cingulate gyrus (PCG)	31	-6	-55	22	7.97
L dlPFC	L inferior frontal gyrus (IFG)	9/47	-42	27	-2	9.15
L TP	L Middle temporal gyrus (MTG)	21/38	-50	-1	-21	7.91
L TPJ	L Superior temporal gyrus (STG)	39	-48	-60	18	7.10

Results from the main modulation. Regions significantly more activated in the experimental condition vs. the control condition ($q(\text{FDR}) < 0.05$) with Talairach coordinates and T values for local maxima. L, left.

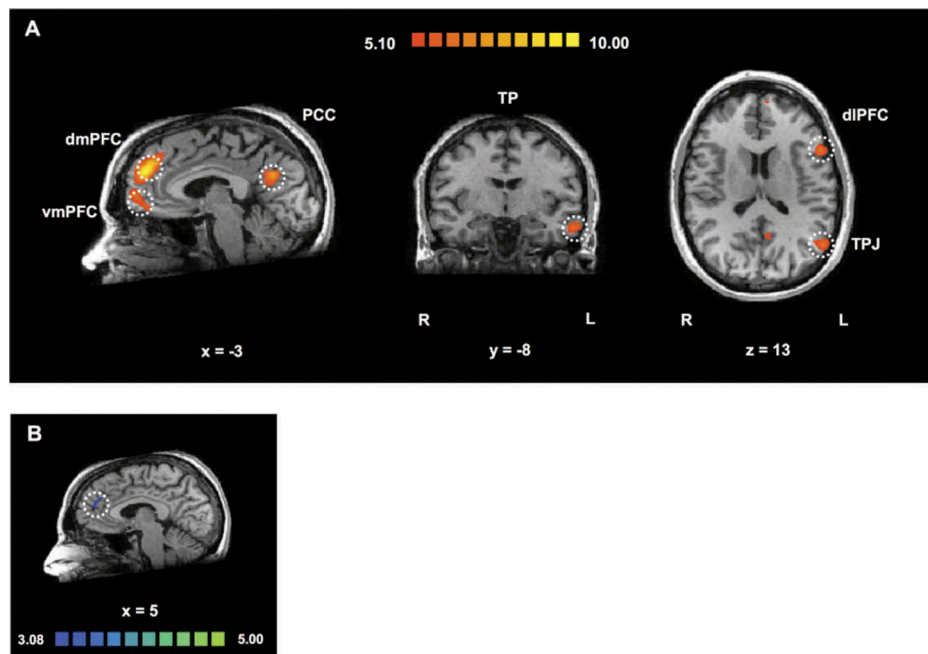


Figure 1. Brain activations. A, Main modulation. Contrasting the experimental condition with the control condition activated dorsomedial prefrontal cortex (dmPFC), ventromedial PFC (vmPFC), left dorsolateral PFC (dlPFC), left temporal pole (TP), left temporo-parietal junction (TPJ), and posterior cingulate cortex (PCC) ($q(\text{FDR}) < 0.05$). B, Parametric modulation. Changes in the degree of punishment showed a negative linear relationship with dorsomedial prefrontal cortex (dmPFC) activation ($P < 0.05$, cluster-level threshold corrected, minimum cluster size of 135 mm^3). For display purposes, statistical images were overlaid onto a surface brain map in Talairach space and were reversed left to right according to radiologic convention. L, left; R, right.

strated that participants' punishment ratings were a good proxy measure of their real-world TPP decisions. Although participants' hypothetical decisions had no direct, real-world consequences for criminal offenders, participants implicitly applied a typical internal punishment scale: shorter jail times were associated with low

punishment scores (1, 25), longer jail time with medium punishment scores (50, 75), and life imprisonment or death penalty with highest punishment scores (100). Further, punishment ratings were related to subjects' affective reports, revealing that more harmful criminal acts were found both more severe and more emotion-

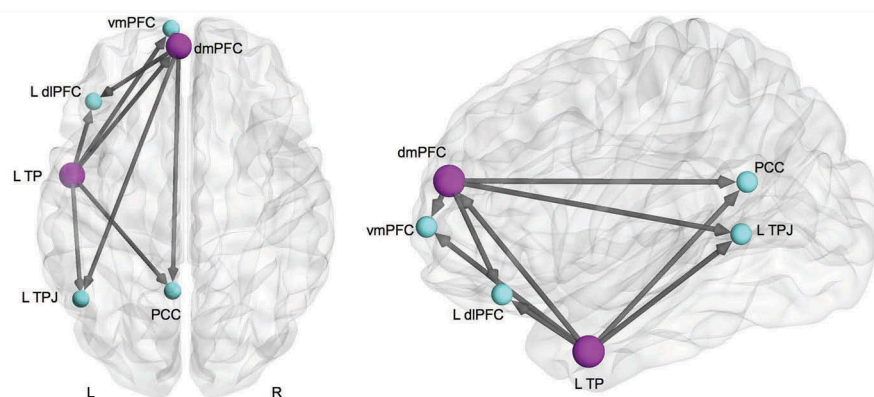


Figure 2. Multivariate Granger causality analysis. Results from the effective connectivity analysis showing connectivity patterns and direction of connections among ROIs previously defined on the basis of our fMRI results. After estimating effective connectivity values for both the experimental and control conditions, values were extracted and contrasted to assess condition-specific modulations of connectivity (paired t-test, $q(\text{FDR}) < 0.05$). For the contrast Experimental > Control, dorsomedial prefrontal cortex (dmPFC) received input from left temporal pole (TP) and both regions emerged as hubs that sent inputs to all other regions (vmPFC, ventromedial PFC; TPJ, temporo-parietal junction; PCC, posterior cingulate cortex; dlPFC, dorsolateral PFC; L, left; R, right).

Table 2. Path weights of Granger causality connections.

Source region	Target region	Path weight	<i>T</i> value	<i>P</i> value
dmPFC	L dlPFC	0.04	3.76	8.5×10^{-5}
	vmPFC	0.07	3.34	4.2×10^{-4}
	L TPJ	0.06	3	1.3×10^{-3}
	PCC	0.03	2.94	1.6×10^{-3}
L TP	L TPJ	0.16	3.82	6.64×10^{-5}
	PCC	0.16	3.26	5.5×10^{-4}
	vmPFC	0.12	3.25	5.7×10^{-4}
	L dlPFC	0.15	3.23	6.2×10^{-4}
	dmPFC	0.15	2.8	2.5×10^{-3}

Significant effective connectivity paths stronger in the experimental condition than in the control condition ($q(\text{FDR}) < 0.05$). The directionality of the connectivity is exhibited in the first two columns: The first column represents source regions whose activity predicted activation in the target regions listed in the second column. Connectivity strength is presented in the third column by mean path-weight values. *T* and *P* values are contained in the fourth and fifth columns, respectively. DmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial PFC; L dlPFC, left dorsolateral PFC; L TP, left temporal pole; L TPJ, left temporo-parietal junction; PCC, posterior cingulate cortex.

ally charged (Darley & Pittman, 2003; Fehr & Fischbacher, 2004a, 2004b; Seymour, Singer, & Dolan, 2007).

Neural results

We next investigated the brain regions underlying TPP decisions and their effective connectivity by contrasting the control condition with the experimental condition for the decision period. Significant brain activations were revealed in brain regions associated with the mentalizing network (dmPFC, vmPFC, PC, TPJ, TP) and the central-executive network (dlPFC), providing further evidence to previous work on normative decision-making (Heekeren, Wartenburger, Schmidt, Schwintowski, & Villringer, 2003; Schleim et al., 2011; Yamada et al., 2012). Moreover, a unique effective connectivity among those brain regions was uncovered: left TP and dmPFC emerged as hubs of the TPP activation pattern, with dmPFC receiving input only from left TP and both regions sending outputs to all other regions. These hubs as part of the mentalizing network possibly accomplish complementary functions to enable the evaluation of legal responsibility.

Overall, in legal decision-making, the mentalizing network is associated with evaluating whether a criminal perpetrator intentionally committed a prohibited act by harming an innocent victim in order to establish her guilty (Buckholtz & Marois, 2012; Krueger, Hoffman, Walter, & Grafman, 2014; Van Overwalle, 2009). Similar to a real-world trial in which jurors and juries are first provided with evidence about the crime, as third-party decision-makers, participants were first required to understand the criminal offense in the form of a social script. The TP hub properly represents information processing related to understanding the criminal offense

and retrieving social concepts to comprehend the perpetrator's action. Previous research has shown that left TP is not only associated with narrative comprehension and social concept knowledge, but it is also believed to enable inferential thinking about the behavior of others (Ferstl, Rinck, & von Cramon, 2005; Ferstl & von Cramon, 2001; Frith & Frith, 2006; Zahn et al., 2007, 2009).

After understanding the criminal scenario, evaluations of legal responsibility have to be performed before decisions of the proper punishment are made. TP was found to send input to the dmPFC hub, presumably providing information about the crime and the criminal act to be integrated with inferences on perpetrator's guilty and blameworthiness. Indeed, a functional connectivity study has previously shown that connectivity between TP and medial PFC mediates inferences about cognitive and motivational dispositions (Moll et al., 2002). The dmPFC hub properly evaluates the mental state of the perpetrator leading to the assessment of the responsibility she bears for her criminal act and thus acting as the driver of the TPP activation pattern. Previous evidence revealed that dmPFC is associated with social inferences (Harris, Todorov, & Fiske, 2005; Mitchell, Macrae, & Banaji, 2006), not only in tasks ranging from reading stories about social interaction to online economic exchange in neuroeconomic games (Amodio & Frith, 2006; Krueger, Grafman, & McCabe, 2008; Krueger et al., 2007) but also in tasks requiring legal or moral decision-making (Buckholtz et al., 2008; Schleim et al., 2011; Yamada et al., 2012). Moreover, dmPFC is implicated in processing social schemata, thus enabling inferences about goals and dispositional trait attributions to identify other people as causal agents (Krueger, Barbey, & Grafman, 2009; Krueger, Moll, Zahn, Heinecke, & Grafman, 2007; Mitchell et al., 2006). Finally, the negative linear relationship between dmPFC activation and the degree of punishment further indicates that dmPFC was less involved when participants decided to punish more. A decision to punish a perpetrator for his blameworthy act is intertwined with the understanding of his goals by inferences based on the described behavior. Understanding perpetrator's goals is necessary to establish whether the behavior is accompanied by a bad intention, a key element of legal punishment (Harris et al., 2005). Less harmful crimes, allowing for diverse possible goals as explanation of the criminal act, place thus more burden on the social inference system and so require a higher engagement of dmPFC.

During the decision phase, all other regions of the mentalizing network (vmPFC, TPJ, and PCC) manifested as input-only nodes of both hubs (dmPFC, TP). Output connections to vmPFC, TPJ, and PCC may be implicated

in representation processes supporting the legal decision by integration of information about the criminal act and the established responsibility of the perpetrator. Specifically, vmPFC likely encodes harm implied by the committed crime due to its general role in monitoring expected values of behavioral outcomes attributed to both punishment and reward during normative decision-making (Amodio & Frith, 2006; Blair, 2008; Buckholz & Marois, 2012; de Quervain Et Al., 2004; Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005; Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010). On the other hand, TPJ most likely allows inferences on the mental state of the perpetrator while committing the crime, based on its general role for inferences of transitory states such as intentions and beliefs of others (Decety & Sommerville, 2003; Saxe, 2006; Saxe & Powell, 2006; Van Overwalle & Baetens, 2009), whereas the PCC presumably provides crime-related context, since it is implicated in contextual and episodic thinking (Addis, Wong, & Schacter, 2007; Szpunar, Chan, & McDermott, 2009).

On the other hand, output connections to dlPFC are presumably related with functions of the central-executive network (i.e., integration and selection of goal-directed actions) (D'Esposito, Postle, & Rypma, 2000; Miller & Cohen, 2001; Ochsner et al., 2004; Wagner, Maril, Bjork, & Schacter, 2001), and therefore involved in choosing the appropriate punishment among an array of punishment options (Buckholz et al., 2008). In particular, stronger connectivity from dmPFC to dlPFC was related to harsher punishments. This finding demonstrates that the dmPFC-dlPFC connectivity is involved in weighing the magnitude of the punishment that fits the crime and supports previous evidence that dlPFC is consistently engaged in explicit rule processing for appropriate forms of two-party punishment (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003) and TPP (Buckholz et al., 2008; Schleim et al., 2011; Yamada et al., 2012).

Limitations and conclusions

Our study had a couple limitations. First, we used a control task (estimate of syllable count) that differed from the experimental task (estimate of punishment) in aspects other than punishment estimation. For both tasks, we separated the reading phase from the decision phase and participants were informed about the type of the task (legal task vs. syllables task) only after the reading period. In this way, participants were expected to make TPP inferences only during the decision period in the experimental but not in the control condition. On

the behavioral level, our control task was meant to ensure that participants paid attention to the tasks. The fact that subjects' syllable ratings were highly significantly associated with the number of syllables in the vignettes shows that they did. On the neural level, we were able to replicate previous findings on the neural components of legal TPP (Buckholz et al., 2008; Schleim et al., 2011; Yamada et al., 2012). Future studies with a more sophisticated control condition are necessary to confirm our findings. Second, since our study was the first study to investigate the effective connectivity of the TPP networks, we employed GCA as a data-driven statistical method that is capable of determining directional influences among BOLD-signal time series of brain regions without a priori hypotheses (Deshpande et al., 2009, 2010; Havlicek et al., 2011). However, future studies need to confirm the identified effective connectivity model using more traditional hypothesis-driven methods such as DCM (Deshpande et al., 2009, 2010; Friston et al., 2013; Havlicek et al., 2011; Wen et al., 2013).

Despite these limitations, the present study revealed for the first time the effective connectivity of the brain regions underlying TPP. We demonstrated that TPP builds on two domain-general networks: mentalizing network (dmPFC, vmPFC, TP, TPJ, PCC) for the evaluation of legal responsibility and central-executive network (dlPFC) for the determination of appropriate punishment. As hub of the mentalizing network, dmPFC received input from TP and both regions uniquely projected outputs to all other regions. Finally, dmPFC emerged as the driver of the TPP networks and the association of its activation and its connectivity to dlPFC with the degree of punishment confirmed the hypothesis that mPFC processes important information related to the punishment decision, which is then conveyed to dlPFC to enable that the decision be made.

In conclusion, these findings advance our understanding of the composition and interaction of the neural correlates of TPP, thereby contributing to a more comprehensive picture of the cognitive mechanisms involved when impartial TPP decision-makers, such as judges and jury members, inflict punishment on wrongdoers.

Acknowledgement

The authors are grateful to E. Wassermann for performing the neurological exams.

Disclosure statement

The authors are unaware of any conflicts of interest, financial, or otherwise.

Funding

The work was supported by the Intramural Research Program of the CNS/ NINDS/ NIH.

References

- Abler, B., Roebroeck, A., Goebel, R., Hose, A., Schonfeldt-Lecuona, C., Hole, G., & Walter, H. (2006). Investigating directed influences between activated brain areas in a motor-response task using fMRI. *Magnetic Resonance Imaging*, 24(2), 181–185. doi:10.1016/j.mri.2005.10.022
- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7), 1363–1377. doi:10.1016/j.neuropsychologia.2006.10.016
- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7(4), 268–277. doi:10.1038/nrn1884
- Blair, R. J. (2008). The amygdala and ventromedial prefrontal cortex: Functional contributions and dysfunction in psychopathy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1503), 2557–2565. doi:10.1098/rstb.2008.0027
- Bradley, M. M., & Lang, P. J. (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25, 49–59. doi:10.1016/0005-7916(94)90063-9
- Buchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage*, 8(2), 140–148. doi:10.1006/nimg.1998.0351
- Buckholz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, 60(5), 930–940. doi:10.1016/j.neuron.2008.10.016
- Buckholz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, 15(5), 655–661. doi:10.1038/nn.3087
- D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: Evidence from event-related fMRI studies. *Experimental Brain Research*, 133(1), 3–11. doi:10.1007/s002210000395
- Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, 7(4), 324–336. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14650389>
- David, O., Guillemain, I., Sallet, S., Reyt, S., Deransart, C., Segebarth, C., ... Depaulis, A. (2008). Identifying neural drivers with functional MRI: An electrophysiological validation. *PLoS Biology*, 6, e315. doi:10.1371/journal.pbio.0060315
- de Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305(5688), 1254–1258. doi:10.1126/science.1100735
- Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive Sciences*, 7(12), 527–533. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14643368
- Deshpande, G., & Hu, X. (2012). Investigating effective brain connectivity from fMRI data: Past findings and current issues with reference to Granger causality analysis. *Brain Connectivity*, 2(5), 235–245. doi:10.1089/brain.2012.0091
- Deshpande, G., Hu, X. P., Stilla, R., & Sathian, K. (2008). Effective connectivity during haptic perception: A study using Granger causality analysis of functional magnetic resonance imaging data. *Neuroimage*, 40(4), 1807–1814. doi:10.1016/j.neuroimage.2008.01.044
- Deshpande, G., LaConte, S., James, G. A., Peltier, S., & Hu, X. (2009). Multivariate Granger causality analysis of fMRI data. *Human Brain Mapping*, 30(4), 1361–1373. doi:10.1002/hbm.20606
- Deshpande, G., Sathian, K., & Hu, X. (2010). Effect of hemodynamic variability on Granger causality analysis of fMRI. *Neuroimage*, 52(3), 884–896. doi:10.1016/j.neuroimage.2009.11.060
- Fehr, E., & Fischbacher, U. (2004a). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4), 185–190. doi:10.1016/j.tics.2004.02.007
- Fehr, E., & Fischbacher, U. (2004b). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63–87. doi:10.1016/S1090-5138(04)00005-4
- Ferstl, E. C., Rinck, M., & von Cramon, D. Y. (2005). Emotional and temporal aspects of situation model processing during text comprehension: An event-related fMRI study. *Journal of Cognitive Neuroscience*, 17(5), 724–739. doi:10.1162/0898929053747658
- Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: An event-related fMRI study. *Cognitive Brain Research*, 11(3), 325–340. doi:10.1016/S0926-6410(01)00007-6
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance Medicine*, 33(5), 636–647. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7596267
- Friston, K., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, 23(2), 172–178. doi:10.1016/j.conb.2012.11.010
- Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? *Neuroimage*, 10(1), 1–5. doi:10.1006/nimg.1999.0439
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531–534. doi:10.1016/j.neuron.2006.05.001
- Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4), 870–878. doi:10.1006/nimg.2001.1037
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, 424–438. doi:10.2307/1912791
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37(3), 414. doi:10.2307/1912791
- Grant, M. M., White, D., Hadley, J., Hutcheson, N., Shelton, R., Sreenivasan, K., & Deshpande, G. (2014). Early life trauma

- and directional brain connectivity within major depression. *Human Brain Mapping*, 35(9), 4815–4826. doi:10.1002/hbm.22514
- Grant, M. M., Wood, K., Sreenivasan, K., Wheelock, M., White, D., Thomas, J., ... Deshpande, G. (2015). Influence of early life stress on intra- and extra-amygdaloid causal connectivity. *Neuropsychopharmacology*, 40(7), 1782–1793. doi:10.1038/npp.2015.28
- Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4), 1639–1651. doi:10.1016/j.neuroimage.2003.11.029
- Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. *Neuroimage*, 28(4), 763–769. doi:10.1016/j.neuroimage.2005.05.021
- Havlicek, M., Friston, K. J., Jan, J., Brazdil, M., & Calhoun, V. D. (2011). Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. *Neuroimage*, 56(4), 2109–2128. doi:10.1016/j.neuroimage.2011.03.005
- Heekeren, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H.-P., & Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport*, 14(9), 1215–1219. doi:10.1097/00001756-200307010-00005
- Hoffman, M., & Goldsmith, T. (2004). The biological roots of punishment. *Ohio State Journal of Criminal Law*, 1(2), 627–641.
- Hutcheson, N. L., Sreenivasan, K. R., Deshpande, G., Reid, M. A., Hadley, J., White, D. M., ... Lahti, A. C. (2015). Effective connectivity during episodic memory retrieval in schizophrenia participants before and after antipsychotic medication. *Human Brain Mapping*, 36(4), 1442–1457. doi:10.1002/hbm.22714
- Katwal, S. B., Gore, J. C., Gatenby, J. C., & Rogers, B. P. (2013). Measuring relative timings of brain activities using fMRI. *Neuroimage*, 66, 436–448. doi:10.1016/j.neuroimage.2012.10.052
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829–832. doi:10.1126/science.1129156
- Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, 13(3), 103–109. doi:10.1016/j.tics.2008.12.005
- Krueger, F., Grafman, J., & McCabe, K. (2008). Review. Neural correlates of economic game playing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1511), 3859–3874. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18829425
- Krueger, F., Hoffman, M., Walter, H., & Grafman, J. (2013). An fMRI investigation of the effects of belief in free will on third-party punishment. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nst092
- 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), 1143–1149. doi:10.1093/scan/nst092
- 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), 1419–1431. doi:10.1002/hbm.21119
- Krueger, F., McCabe, K., Moll, J., Kriegenkorte, N., Zahn, R., Strenziok, M., ... Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104(50), 20084–20089. doi:10.1073/pnas.0710103104
- 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), 2346–2353. doi:10.1093/cercor/bhl143
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202. doi:10.1146/annurev.neuro.24.1.167
- Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, 50(4), 655–663. doi:10.1016/j.neuron.2006.03.040
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *The Journal of Neuroscience*, 22(7), 2730–2736. doi:10.1523/JNEUROSCI.2002-02.2002
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). Opinion: The neural basis of human moral cognition. *Nat Rev Neurosci*, 6(10), 799–809. doi:10.1038/nrn1768
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, 25(3), 653–660. doi:10.1016/j.neuroimage.2004.12.005
- Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage*, 23(2), 483–499. doi:10.1016/j.neuroimage.2004.06.030
- Robinson, P., & Darley, J. (2007). Intuitions of justice: Implications for the criminal law. *S Cal L Reviews*, 81, 1–66.
- Robinson, P., & Kurzban, R. (2007). Concordance and conflict in institutions of justice. *Minnesota Law Review*, 91, 1829–1907.
- Roebroeck, A., Formisano, E., & Goebel, R. (2005). Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage*, 25(1), 230–242. doi:10.1016/j.neuroimage.2004.11.017
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, 300(5626), 1755–1758. doi:10.1126/science.1082976
- Sathian, K., Deshpande, G., & Stilla, R. (2013). Neural changes with tactile learning reflect decision-level reweighting of perceptual readout. *Journal of Neuroscience*, 33(12), 5387–5398. doi:10.1523/JNEUROSCI.3482-12.2013
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16(2), 235–239. doi:10.1016/j.conb.2006.03.001
- Saxe, R., & Powell, L. J. (2006). It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science*, 17(8), 692–699. doi:10.1111/j.1467-9280.2006.01768.x
- 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. doi:[10.1093/scan/nsq010](https://doi.org/10.1093/scan/nsq010)
- Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nat Rev Neurosci*, 8(4), 300–311. doi:[10.1038/nrn2119](https://doi.org/10.1038/nrn2119)
- 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), 1306–1360. Retrieved from <Go to ISI>://000297236300002
- Sreenivasan, K. R., Havlicek, M., & Deshpande, G. (2015). Nonparametric hemodynamic deconvolution of fMRI using homomorphic filtering. *IEEE Trans Med Imaging*, 34(5), 1155–1163. doi:[10.1109/TMI.2014.2379914](https://doi.org/10.1109/TMI.2014.2379914)
- Szpunar, K. K., Chan, J. C., & McDermott, K. B. (2009). Contextual processing in episodic future thought. *Cerebral Cortex*, 19(7), 1539–1548. doi:[10.1093/cercor/bhn191](https://doi.org/10.1093/cercor/bhn191)
- Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers.
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, 30(3), 829–858. doi:[10.1002/hbm.20547](https://doi.org/10.1002/hbm.20547)
- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage*, 48(3), 564–584. doi:[10.1016/j.neuroimage.2009.06.009](https://doi.org/10.1016/j.neuroimage.2009.06.009)
- Wagner, A. D., Maril, A., Bjork, R. A., & Schacter, D. L. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral Prefrontal cortex. *Neuroimage*, 14(6), 1337–1347. doi:[10.1006/nimg.2001.0936](https://doi.org/10.1006/nimg.2001.0936)
- Wen, X., Rangarajan, G., & Ding, M. (2013). Is Granger causality a viable technique for analyzing fMRI data? *PLoS One*, 8(7), e67428. doi:[10.1371/journal.pone.0067428](https://doi.org/10.1371/journal.pone.0067428)
- Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS One*, 8(7), e68910. doi:[10.1371/journal.pone.0068910](https://doi.org/10.1371/journal.pone.0068910)
- Yamada, M., Camerer, C. F., Fujie, S., Kato, M., Matsuda, T., Takano, H., ... Takahashi, H. (2012). Neural circuits in the brain that are activated when mitigating criminal sentences. *Nature Communications*, 3, 759. doi:[10.1038/ncomms1757](https://doi.org/10.1038/ncomms1757)
- 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*, 107(15), 6753–6758. doi:[10.1073/pnas.0914826107](https://doi.org/10.1073/pnas.0914826107)
- Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(15), 6430–6435. doi:[10.1073/pnas.0607061104](https://doi.org/10.1073/pnas.0607061104)
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., & Grafman, J. (2009). The neural basis of human social values: Evidence from functional MRI. *Cerebral Cortex*, 19(2), 276–283. doi:[10.1093/cercor/bhn080](https://doi.org/10.1093/cercor/bhn080)