Psychology. Journal of the Higher School of Economics.
2019. Vol. 16. N 3. P. 529–550. DOI: 10.17323/1813-8918-2019-3-529-550

THE ROLE OF THE TEMPOROPARIETAL AND PREFRONTAL CORTICES IN A THIRD-PARTY PUNISHMENT: A TDCS STUDY

O.O. ZINCHENKO[a], A.V. BELIANIN[a], V.A. KLUCHAREV[a]

[a] *National Research University Higher School of Economics, 20 Myasnitskaya Str., Moscow, 101000, Russian Federation*

Abstract

Recent studies have demonstrated that the right dorsolateral prefrontal cortex and the right temporoparietal junction are causally involved in social norm compliance: its activation corresponds with the third-party norm enforcement behavior, known as third-party punishment. The current study aimed to address the inconsistencies in effects of brain stimulation methods on right dorsolateral prefrontal cortex and right temporoparietal junction to clarify its role on third-party punishment. Despite a decade of neuroimaging research, the interaction between the right temporoparietal junction and right dorsolateral prefrontal cortex in third-party punishment remained unclear. Here, we tested the hypothesis that a third party's decision to punish norm violations depends on the activity of the entire right dorsolateral prefrontal cortex-right temporoparietal junction network. We used transcranial direct current stimulation to independently or jointly modulate right temporoparietal junction and right dorsolateral prefrontal cortex activity during the third-party dictator game. We found a significant effect of anodal transcranial direct current stimulation of the right temporoparietal junction, which decreased the third-party punishment of moderately unfair splits. Joint stimulation of the right temporoparietal junction (by anodal transcranial direct current stimulation) and right dorsolateral prefrontal cortex (by cathodal transcranial direct current stimulation) produced a marginal effect on third-party punishment. Our results suggested that the right temporoparietal junction could modulate the perceived moral costs of third-party punishment.

Keywords: dorsolateral prefrontal cortex, transcranial direct current stimulation, temporoparietal junction, third-party punishment, social norms.

Introduction

Human societies crucially depend on social norms that often regulate appropriate actions in various situations and can be reinforced by "second" parties that are directly affected by the norm violators and "third" parties that are not directly affected (Fehr & Fischbacher, 2004). Since norm violations often do not directly

The article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project '5-100'.

hurt other people, third-party sanctions are especially critical in reinforcing social norms (Bendor & Swistak, 2001; Fehr & Fischbacher, 2004). More than a decade of neuroimaging research has established that several distinct brain networks are consistently recruited during social punishment; that is, the cooperative individuals' propensity to spend part of their resources to penalize norm violators (Krueger & Hoffman, 2016). Here, we further investigate the neural underpinnings of third parties' punishment of a fairness norm violation.

The social norm of fair distribution implies a rejection of the distribution of goods that violates the equality principle (Elster, 1989; Kahneman, Knetcsh, & Thaler, 1986). The norm of fairness is often investigated using economic games, allowing different distributions of financial transfers between players. Importantly, behavioral studies have robustly demonstrated that many players (including third parties) in economic games not only prefer fair distributions to unequal ones (Guth, Schmittberger, & Schwarze, 1982; Engel, 2011), but they also tend to spend personal resources to punish unfair distributions (norm violations) on their own accord (Fehr & Fischbacher, 2004; Ruff, Ugazio, & Fehr, 2013).

Functional magnetic resonance imaging (fMRI) and brain stimulation studies have suggested that the right dorsolateral prefrontal cortex (rDLPFC) controls selfish impulses (Strang et al., 2015) and responds to inequity (Fliessbach et al., 2012),where individual differences in sanction-induced norm compliance correlate with rDLPFC activity (Spitzer, Fichbacher, Hernberger, Grön, & Fehr, 2007). Brüne and colleagues (2012) showed that inhibitory rTMS of the rDLPFC increased third-party punishment during the dictator game, which suggests that the rDLPFC associated third parties' emotional responses to observed unfairness of dictators. In contrast, rTMS of the rDLPFC resulted in decreased third-party punishment when participants where shown criminal scenarios ranging from simple theft to murder (Buckholtz et al., 2015). Inconsistencies in effects of rTMS on rDLPFC require further work to clarify the role of the rDLPFC on third-party punishment. Like Brüne and colleagues (2012), the current project utilized the third-party dictator game but in contrast to their experiment manipulations, we aimed to apply excitatory anodal tDCS on the rDLPFC, and predict that the opposed effects would ensue and therefore decrease third-party punishment.

Another neuroanatomical structure that has been found to play a critical role in third parties' punishment decisions is the right temporoparietal junction (rTPJ) (Baumgartner, Götte, Gügler, & Fehr, 2012; Baumgartner, Schiller, Rieskamp, Giantti, & Knoch, 2014). Importantly, the ability to make inferences about other people's mental states is associated with TPJ activation, that is crucial for the ability to blame others for violations of complex context-dependent social norms. Increased rTPJ activity has been associated with reduced punishment of defecting in-group members during the prisoner's dilemma game (Baumgartner et al., 2012). Here, we hypothesized that excitatory anodal tDCS of the rTPJ should reduce third-party punishment during the third-party dictator game.

Recently, Krueger and Hoffman (2016) argued that during third-party punishment, the TPJ integrates the inference of intentions into an assessment of blame. The DLPFC converts the blame signal into a specific punishment decision. Thus, the DLPFC plays an executive role, while the TPJ drives processes associated with blame and initiates punishment. Although these mechanistic actions are neurobiologically plausible, the exact interaction between the rTPJ and rDLPFC for the function of third-party punishment remain unclear, because it is unknown whether the rDLPFC entrains the rTPJ or that rTPJ activates independently (for a detailed discussion, see Zinchenko & Klucharev, 2017). It has also been shown that, TPJ activity during third-party punishment is paralleled by an initial deactivation of the DLPFC which indicates functionally opposed neural activity in these two regions (Buckholtz et al., 2008). The DLPFC demonstrates biphasic neural activity—after the initial deactivation, it later increases in activity—when subjects make the final decision to punish "based on assessed responsibility and blameworthiness" (Buckholtz et al., 2008, p. 935). Overall, Buckholtz and colleagues (2008) suggested that this pattern of reciprocal activation could reflect a crucial mentalizing process before an appropriate punishment is determined and a decision is made. Therefore, it would be important to further study the functional interaction of the rTPJ and rDLPFC during third-party punishment through the use of joint stimulation of the rDLPFC and rTPJ in a reciprocal manner.

In the current study, we further investigated the role of the DLPFC and TPJ in third-party punishment with an overarching aim of understanding neural resource activation plays a role in social norm reinforcement. Our motivation was based on previous studies (Baumgartner et al., 2012; Brüne et al., 2012) and sought to replicate their results, but by uniquely testing the opposed effect of independent excitatory anodal tDCS of the rTPJ or rDLPFC and predict that decreased third-party punishment of unfair splits would be resultant (Hypothesis I, Study 1) compared to sham stimulation. On the other hand, based on the seminal fMRI study (Buckholtz et al., 2008), we hypothesized that a joint anodal tDCS of the rTPJ and cathodal tDCS of the rDLPFC of third parties might make third-party punishment of unfair splits stronger comparing to sham stimulation (Hypothesis II, Study 2). Therefore, in Study 1, we stimulated the rDLPFC and rTPJ independently, while in Study 2, we jointly stimulated the rDLPFC and rTPJ in a reciprocal, antagonistic manner. Importantly, according to the theory of inequity aversion, individuals dislike outcomes that are perceived as inequitable (Fehr & Schmidt, 1999). Therefore, we expected to find the strongest effect of tDCS on third-party punishment in trials with a payoff structure, where *sanctioners* (third parties) were able to establish the equality between all players (Hypothesis III). Overall, we used tDCS to further investigate the role of the rDLPFC and rTPJ in third parties' punishment of a fairness norm violation. According to our hypotheses, an independent or joint stimulation of the rDLPFC and rTPJ could lead to different behavioral effects.

Methods

Subjects

Twenty-three healthy, right-handed subjects (mean age = 21.5 years, range = 18–27 years, 7 males) participated in Study 1. Twenty-one healthy, right-handed subjects (mean age = 22.79 years, range = 18–27 years, 10 males) participated in Study 2. Each subject participated in only one of the two studies. All subjects gave written informed consent to participate in the study. Subjects (n = 5) who did not punish at least once or demonstrated only antisocial punishment in fair trials (20:20 split condition) were excluded from the analysis, resulting in 20 (n = 20, Study 1) and 19 (n = 19, Study 2) subjects respectively. The studies conformed to the Declaration of Helsinki, and the experimental protocol was approved by the university ethics committee. The sample size was based on the previous study of Brüne and colleagues (2012), which included 20 subjects.

Procedure

Each subject participated in three sessions of the dictator game that were separated by 7±2 days. Next, tDCS was applied for 15 minutes. The third-party dictator game lasted approximately 20–25 minutes. A structured debriefing after each session revealed that the subjects believed the instructions and their behavior were comparable to those in “real-life” situations.

Study Design

Dictator Game with Third-Party Punishment

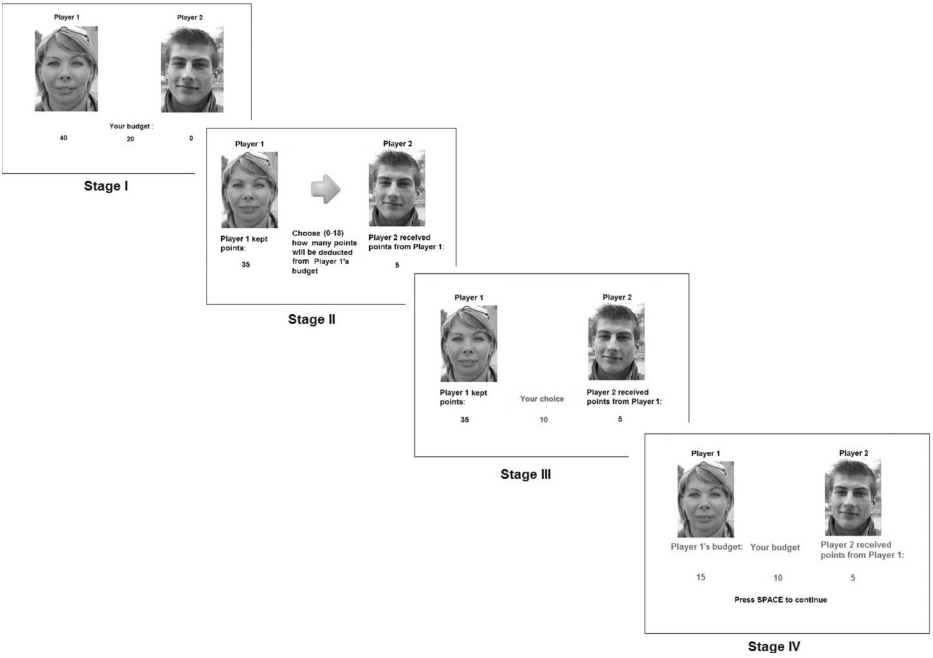
The subjects participated in multiple rounds of a preprogrammed dictator game as *sanctioners* (third parties). In the instructions, the dictator distributed 40 experimental monetary units (MUs; 1 MU ≈ 0.26 Russian rubles, or 0.004 U.S. dollars) between herself and the *recipient.*

Figure 1 demonstrates the details of the trial structure. To make the game more social, in each trial the participants first observed pictures of two individuals (a *dictator* and a *recipient*). The genders of the *dictators* and *recipients* were counterbalanced across subjects. Participants (*sanctioners*) were able to punish *dictators* using a budget of 20 MUs. The budget was renewed for each round, and all points not invested in punishment were converted into a monetary payoff and paid to the participant after the experiment. To avoid demand effects, the instructions described the task using neutral language, such as “You will be able to deduct the first player’s earnings.” *Sanctioners* could use 0–18 MUs out of their 20 MUs budget to punish the *dictator,* and were able to use an even number of MUs, such as 2, 4, 6, etc. These MUs were multiplied by two and deducted from the *dictator*’s budget. For example, if the *sanctioner* used 10 MUs to punish the *dictator,* 20 MUs (2 × 10 MUs) were deducted from *dictator*’s budget.

The photos of *dictators* and *recipients* were preselected from 300 photos of young adults. The images were retrieved from the Internet from open access sources, such as popular social media without being logged in. For ethical reasons, it was carefully ensured that the photos stayed anonymous — no personal information was stored. Similar to the study of Brüne and colleagues (2012), we pretested stimuli: photos were evaluated for attractiveness, trustworthiness, and cooperativeness on a

Figure 1

Trial Structure of the Dictator Game with Third-Party Punishment



Note. At the beginning of each trial, a dictator (Player 1) received 40 monetary units (MUs; Stage I) to choose whether to give some MUs to the recipient (Player 2; Stage II). Next, the subject (Player 3, sanctioner) received 20 MUs (Stage III) to choose how much (if any) to spend on punishing the dictator (Stage IV), in which every MU spent by the sanctioner reduced the dictator's payoff by 2 MUs.

seven-point Likert-type scale by 17 subjects (10 females) prior to the study. We calculated the average rating of each measure (attractiveness, trustworthiness, and cooperativeness) for each photo. Similar to the previous rTMS study (Brüne et al. 2012), only photos with a mean average rating between 2.5 and 5.5 points were used in the current study. If the average rating for at least one measure was higher or lower than this range, the photo was excluded and never used in the study.

The following information was emphasized to the participants: (1) their partners were real people participating in the game at the same time, located in different rooms; (2) the partners varied in each round; and (3) both the participants and their partners would be paid real money, as all points that had not been invested during the game would be paid out at the end of the study. Although the subjects believed that they were playing an "online" game, they were, in fact, playing with prerecorded human players (*dictators* and *recipients*) who had played the same game before against other human opponents (see Brüne et al., 2012, for the same approach). Therefore, each session consisted of 48 trials per split condition, with shares of 0:40 (n = 2), 15:25 (n = 1), 20:20 (n = 26), 25:15 (n = 4); 30:10 (n = 6), 35:5 (n = 3) and 40:0 (n = 6). The trials were randomized in each session. All the subjects were native Russians recruited via email. The number of trials in each split condition was defined based on a behavioral pilot study (n = 178).

tDCS

The tDCS is a noninvasive brain stimulation technique that can modulate activity in specific regions of the cortex (Nitsche, Paulus, & 2001, Nitsche et al. 2003; Paulus, 2011). During tDCS, weak electrical currents are applied to the scalp surface from the anode to cathode: anodal tDCS typically depolarizes (excites) and cathodal tDCS typically hyperpolarizes (inhibits) neurons. In the current study, a direct current was induced using two saline-soaked surface sponge electrodes (active electrode area = 25 cm^2) and delivered by a battery-driven, constant current StarStim 8 stimulator (Neuroelectrics). The stimulation intensity was set at 1.5 mA and lasted 15 minutes, with ramping up and ramping down time equal to 30 seconds. Impedances were kept below 10 kOhm.

After 15 minutes of tDCS, participants immediately participated in the dictator game as a third-party. Importantly, several methodological studies demonstrated that even tDCS (1 mA) delivered for a short time (5–13 minutes) induced long-lasting changes of cerebral excitability: up to 90 minutes after the end of stimulation (Nitsche & Paulus, 2001) for anodal tDCS and up to one hour for the cathodal tDCS (Nitsche et al., 2003). Therefore, a 15-minute tDCS in our study should modulate cortical excitability during the entire dictator game.

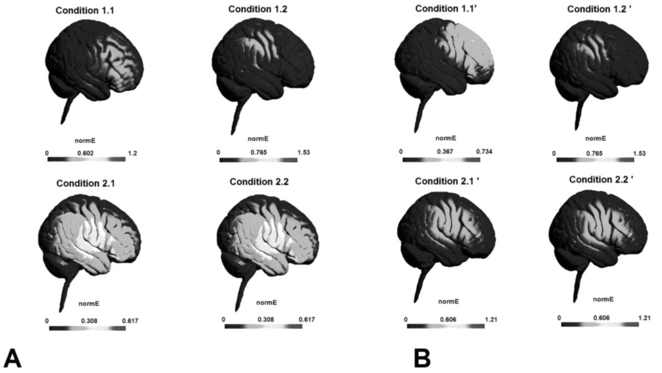
The stimulation point for the rDLPFC was defined using the MNI coordinates reported by Spitzer et al. (2007) for rDLPFC activity (x = 52, y = 28, z = 14), which showed both stronger fMRI activation for punishment condition minus baseline condition as well as a correlation of brain activity with the transfer difference between punishment and baseline conditions (see Ruff et al., 2013 for a similar approach). To further clarify the optimal electrode position, we simulated tDCS using SimNIBS software, version 2.1.1. (see Figure 2; www.simnibs.de/start; Thielscher et al., 2015).

Overall, the results of the simulation indicated that the F8 electrode position was an adequate target for the rDLPFC stimulation. To stimulate the rTPJ, the target electrode was located over CP6 region (Santiesteban, Banissy, Catmur, & Bird, 2012; Sellaro et al., 2015). For the sham stimulation, the intensity and position of the electrodes were the same as during a real stimulation, but the stimulator was only turned on for 30 seconds. The positions of the electrodes for the sham stimulation in Study 1 and Study 2 were randomized and counterbalanced, as was the order of the stimulation sessions (see Figure 3).

In Study 1, we applied the anodal tDCS of the rDLPFC and rTPJ independently, as follows: (1) rDLPFCa condition (Condition 1.1) – anodal tDCS of the rDLPFC; (2) rTPJa condition (Condition 1.2) – anodal tDCS of the rTPJ; and (3) sham condition (Condition 1.3). In all conditions of Study 1, the cathodal electrode was placed over the vertex (Cz electrode position). We expected (Hypothesis I) to decrease third-party punishment in Conditions 1.1 and 1.2 as compared with the control (Condition 1.3).

Figure 2

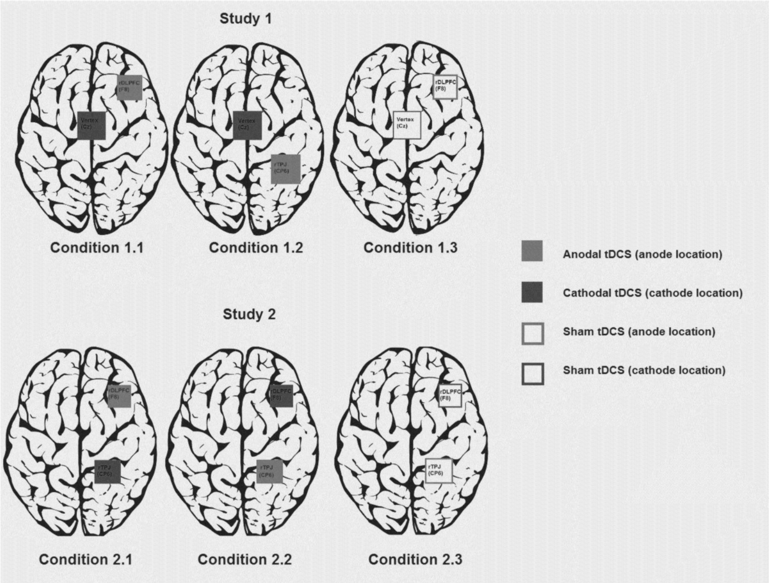
Simulations of Electric Current Distributions (E-Field, V/M) for Different tDCS Protocols



Note. **A:** Simulations of the tDCS protocols used in the current study (F8 electrode for rDLPFC stimulation, CP6 electrode for rTPJ stimulation). **B:** Simulations of the alternative tDCS protocol, where F8 electrode is replaced with F4 (F4 electrode for rDLPFC stimulation, CP6 electrode for rTPJ stimulation).

Figure 3

Transcranial Direct Current Stimulation (tDCS) Set-Ups for Study 1 and Study 2



Note. Study 1: Condition 1.1 — anodal tDCS of the rDLPFC; Condition 1.2 — anodal tDCS of the rTPJ; Condition 1.3 — sham condition. In all conditions of Study 1, the cathodal electrode was placed over vertex. Study 2: Condition 2.1 — simultaneous anodal tDCS of the rDLPFC and cathodal tDCS of the rTPJ; Condition 2.2 — simultaneous cathodal tDCS of the rDLPFC and anodal tDCS of the rTPJ; Condition 2.3 — sham.

In Study 2, we simultaneously modulated rDLPFC and rTPJ activity, as follows: (1) rDLPFCa/rTPJc condition (Condition 2.1) – simultaneous anodal tDCS of the rDLPFC and cathodal tDCS of the rTPJ; (2) rDLPFCc/rTPJa condition (Condition 2.2) – simultaneous cathodal tDCS of the rDLPFC and anodal tDCS of the rTPJ, and (3) sham condition (Condition 2.3), which was the same as in Study 1. Following the findings of Buckholtz and colleagues (2008), which demonstrated a reciprocal activation of the rDLPFC and rTPJ, we expected (Hypothesis II) to increase third-party punishment in Condition 2.2 as compared with Conditions 1.2 and 2.3.

There is evidence suggesting that tDCS could effectively modulate within-network and between-network interactions. For example, the simultaneous anodal tDCS of the DLPFC, together with cathodal tDCS of the supraorbital region, led to changes in the default mode network (Keeser et al., 2011; Peca-Gómez et al., 2012). Here, we used a simultaneous application of tDCS to the rTPJ and rDLPFC to modulate their interaction during third-party punishment. Thus, we developed a mixed design that allows exogenous modulation of between-network interaction.

Statistical Analysis

For each experimental condition, we calculated a sum of MUs, which were used to punish the dictator, to estimate the *punishment level*, or total investment in punishment (see Brüne et al., 2012 for the same approach). According to the payoff matrix of the dictator game, our participants would experience *advantageous inequity* (when they receive more than others) or *disadvantageous inequity* (when they receive less than others) after all splits except for the 20:20 split. Importantly, only when the *dictator* chose a 30:10 split were participants able to restore equality by spending 10 of their own MUs on punishing the dictator. Interestingly, only in the case where participants observed moderately unfair 30:10 splits, the conflict between material (selfish) and moral (prosocial) costs was minimal, since participants either did not punish or punished extremely little if the material costs were high. Therefore, we could expect that the strongest effect of tDCS would be observed in the 30:10 split condition, when participants were able to restore equality and protect their own material interests (Hypothesis III).

To test Hypothesis III, we aggregated split conditions of the game into three trial types: *FS-trials*—fair splits (20:20 and 25:15 splits), *US_equal-trials*—unfair splits (30:10 splits), where third-party punishment was able to establish equality between all players, and *US_inequal-trials* (35:5 and 40:0 splits), where participants were unable to establish such equality. Due to a very low number of observations, 0:40, 15:25 split conditions were not included in the main analysis. However, Table 1 provides descriptive statistics for all split conditions.

Since the punishment levels were not normally distributed, behavioral results were analyzed using the Wilcoxon signed-rank test and the Friedman test, and p-values < .05 were considered significant. To correct for multiple comparisons, the false discovery rate (FDR) correction at 10% level using the Benjamini-Hochberg procedure (1995) was computed to compare the effects of three types of stimulation

Table 1

The Mean and Standard Deviations of Punishment Level in Study 1 and Study 2

Condition/ Types of splits	**0:40 + 15:25 (not included into main analysis)**	**20:20+25:15 [FS-trials]**	**30:10 [US_equal-trials]**	**35:5+40:0 [US_inequal-trials]**
Study 1				
Condition 1.1				
Mean	0.5	22	48.60	118.30
SD	1.43	10.05	19.22	45.17
Condition 1.2				
Mean	2.7	20.40	46.20	112
SD	6.53	10.59	17.96	43.61
Condition 1.3				
Mean	1.5	21.30	49.50	119.30
SD	3.55	10.02	16.78	39.70
Study 2				
Condition 2.1				
Mean	0.21	13.79	37.05	112.53
SD	0.92	11.25	20.56	33.65
Condition 2.2				
Mean	1.16	13.37	40.52	109.37
SD	4.18	10.67	19.43	32.19
Condition 2.3				
Mean	2.11	11.26	34.21	111.37
SD	8.23	8.92	20.36	28.86

(Conditions 1.1, 1.2, and 1.3) and three types of splits (*FS-trials, US_equal-trials, and US_inequal-trials*) in Study 1. To compute the FDR correction, p-values obtained in the statistical analysis were ranked from the lowest to the highest and then compared to FDR-corrected alpha levels (Benjamini-Hochberg critical value). Only p-values not exceeding FDR-corrected alpha levels were considered significant. To control individual differences in third-party punishment, we normalized *punishment levels*: for each trial type, the *punishment level* was divided by the punishment level in the sham condition and multiplied by 100%. Between-group differences were further evaluated using the Kruskal–Wallis H test, which was applied to normalized data.

Results

Study 1: Independent modulation of the rDLPFC and rTPJ

In total, in the sham condition (Condition 1.3), participants spent only 1.5 (±3.6) MUs for the punishment of generous 0:40 and 15:25 splits, 21.3 (±10.0) MUs — for the punishment of fair *FS-trials*, while for unfair *US_equal-trials* they used 49.5 (±16.8) MUs and for *US_inequal-trials*, 119.3 (±39.7) MUs. Due to a very low number of observations, 0:40, 15:25 split conditions were not included in further analyses.

The lowered punishment level of *FS-trials* compared with *US_equal-trials* and *US_inequal-trials* was observed in all experimental tDCS conditions. Table 1 represents the mean and standard deviations of punishment level of third-party punishment for each. As expected, the participants punished unfair splits much more strongly than they did fair splits.

rDLPFCa condition. We observed a trend of a stronger third-party punishment in the rDLPFCa condition (Condition 1.1) than in the rTPJa condition (Condition 1.2): Z = −2.177; p = .029, which did not survive FDR correction (see Table 2 for FDR-corrected alpha levels).

rTPJa condition. We found that the third-party punishment in *US_equal-trials* (30:10 splits) in the rTPJa condition (Condition 1.2) was significantly smaller than it was in the sham condition (Condition 1.3): Z = −2.746, p = .006 (see Table 1 and Table 3 for details). We also observed a trend of a smaller third-party punishment in *US_inequal-trials* (35:5 and 40:0 splits) in the rTPJa condition (Condition 1.2) compared with the sham condition (Condition 1.3): Z = −2.006, p = .045, which did not survive FDR correction (see Table 2 for FDR-corrected alpha levels).

Table 2

The False Discovery Rate Computation (Study 1)

Condition	p-values obtained in the statistical analysis	Rank	FDR-corrected alpha levels (Benjamini-Hochberg critical values)
Condition 1.2 – Condition 1.3 (30:30)	0.006*	1	0.011
Condition 1.1 – Condition 1.2 (35:5+40:0)	0.029	2	0.022
Condition 1.2 – Condition 1.3 (35:5+40:0)	0.045	3	0.033
Condition 1.1 – Condition 1.2 (20:20+25:15)	0.347	4	0.044
Condition 1.1 – Condition 1.2 (30:10)	0.450	5	0.055
Condition 1.1 – Condition 1.3 (20:20+25:15)	0.459	6	0.066
Condition 1.1 – Condition 1.3 (30:10)	0.649	7	0.077
Condition 1.1 – Condition 1.3 (35:5+40:0)	0.678	8	0.088
Condition 1.2 – Condition 1.3 (20:20+25:15)	1.000	9	0.100

Table 3

Effect of Unilateral Transcranial Direct Current Stimulation on Third-Party Punishment (Study 1)

Split	20:0+25:5	30:10	35:5+40:0
Rank	(1.90; 1.88; 2.23)	(2.30; 1.58; 2.13)	(2.20; 1.60; 2.20)
χ^2	1.937	7.508	5.408
df	2	2	2
p	.380	.023*	.067
Note. Friedman test for 3 samples. * Significant at the level of *p* = .05.			
Condition 1.2 – Condition 1.3			
Split	20:0+25:5	30:10	35:5+40:0
Z	0.000	−2.746	−2.006
p	1	.006*	.045
Condition 1.1 – Condition 1.3			
Split	20:0+25:5	30:10	35:5+40:0
Z	−0.741	−0.456	−0.415
p	.459	.649	.678
Condition 1.1 – Condition 1.2			
Split	20:0+25:5	30:10	35:5+40:0
Z	−0.940	−0.755	−2.177
p	.347	.450	.029
Note. Wilcoxon signed-rank test for sums of punishment points. * Significant at the level of *p* = .05.			

We found no other significant effects of tDCS on third-party punishment. Therefore, Hypotheses I and III were partly supported: anodal tDCS of the rTPJ significantly decreased third-party punishment, but only in *US_equal-trials* (unfair 30:10 splits), where third-party punishment was able to establish equality between all players.

Study 2: Simultaneous Modulation of the rDLPFC and rTPJ

Similar to Study 1, the participants in Conditions 2.1, 2.2, and 2.3 punished unfair splits (*US_equal-trials* and *US_inequal-trials*) more strongly than they did fair splits. We found no significant effects of tDCS in Study 2. Interestingly, the third-party punishment for 30:10 splits in the rDLPFCc/rTPJa condition (Condition 2.2) tended to be higher than that in the sham condition (Condition 2.3), Z = −1.917, *p* = .055 (uncorrected; see Table 4 for details). Therefore, our results did not support Hypothesis II.

Table 4

Effect of Reciprocal Transcranial Direct Current Stimulation on Third-Party Punishment (Study 2)

Split	20:20+25:5	30:10	35:5+40:0
Rank	(1.71; 2.16; 2.13)	(1.66; 2.24; 2.11)	(1.92; 2.08; 2.00)
χ^2	2.984	5.115	0.269
df	2	2	2
p	.225	.077	.874
Note. Friedman test for 3 samples.			
Condition 1.2 – Condition 1.3			
Split	20:0+25:5	30:10	35:5+40:0
Z	-0.999	−1.917	−0.028
p	.318	.055	.977
Condition 1.1 – Condition 1.3			
Split	20:0+25:5	30:10	35:5+40:0
Z	−1.307	−1.401	−0.087
p	.191	.161	.930
Condition 1.1 – Condition 1.2			
Split	20:0+25:5	30:10	35:5+40:0
Z	−0.140	−1.337	−0.570
p	.888	.181	.569

Between-group analysis

A Kruskal–Wallis H test showed that the normalized punishment levels for 30:10 splits differed in Study 1 and Study 2: $\chi^2 = 4.481$, $p = .034$ (Condition 1.2 versus Condition 2.2; see Table 5). Third-party punishment for 30:10 splits was significantly lower in Study 1 than in Study 2. This could indicate an opposite effect of the anodal tDCS of the rTPJ (rTPJa condition) compared with the anodal tDCS of the rTPJ when paralleled with cathodal tDCS of rDLPFC.

Inequity aversion model

To assess the effect of stimulation on third-party punishment from theoretical viewpoint, we used a modified inequity aversion model (Fehr & Schmidt, 1999) extended to third parties (Svedsater & Johannsson, 2005). The model assumes that each player in the game generally dislikes unfair outcomes reducing their utility. First, consider the utility function of *sanctioner*, who observes the dictator game between *dictator* and *recipient*, but who has no punishment option. The *sanctioner* receives an endowment and feels unhappy whenever any other players receive

Table 5

Comparison of Transcranial Direct Current Stimulation Effects on Third-Party Punishment in Study 1 And Study 2 (Condition 1.2–Condition 2.2 [Normalized Data]).

Split	25/15	30/10	35/5	40/0
Rank	(23.28; 16.55)	(16.25; 23.95)	(23.10; 16.74)	(20.23; 19.76)
χ^2	3.420	4.481	3.105	0.016
df	1	1	1	1
p	.064	.034*	.078	.898

either more (with parameter α) or less than she does (with parameter β). The *sanctioner* also experiences moral loss when the payoffs of the *dictator* and *recipient* are unequal (with parameter γ). In terms of inequity aversion model of Fehr and Schmidt (1999), the resulting utility is:

$$U_3N = w_3 - \alpha max(X_1 - w_3,0) - \alpha max(X_2 - w_3,0) - \beta\, ax(w_3 - X_1,0) \; - \beta max(w_3 - X_2,0) - \gamma|X_1 - X_2|. \qquad \text{(Eq. 1)}$$

Here, X_1 and X_2 are MUs collected by *dictator* and *sanctioner*, respectively, $X_1 + X_2 = 40$ (MUs); $w_3 = 20$ is the endowment of *sanctioner*, and α, β, and γ are parameters of inequity aversion. Only if both other players receive exactly as much as the third player (and hence, their respective payoffs are equal) the third player experiences no utility loss, receiving just w_3.

Fehr and Schmidt's (1999) canonical two-player inequity aversion model typically assumes that $\alpha > 1 > \beta > 0$; this captures the envy of each player who dislikes being treated unfairly more than (s)he dislikes being unfair towards another player. In our study, *dictator*'s decision does not materially affect the *sanctioner*, who may want to punish the former player only for violations of ethical standards, but not because of personal material losses. Hence, moral loss of the *sanctioner* can be assumed to be larger than the cost of punishment; that is, $1 > \alpha > \beta > 0$. Furthermore, in our application, we may separate two feelings of the *sanctioner*, as follows:

(1) The *sanctioner's discomfort*/disapproval of unfair actions of the *dictator*, which she may compensate for by the third-party punishment. The level of this discomfort is proportional to the extent of unfairness, and its strength is captured by parameter γ (we assume that $\gamma > 1$); and

(2) *The costs of punishment*, which consist of two elements, namely the monetary cost of punishment and the *sanctioner's* wellbeing relative to that of the other players. Inequity-averse *sanctioners* are concerned about fairness of the terminal distribution; hence, these feelings are proportional to the realized differences between the revenues of Players 3 and 1 and 3 and 2, taken with strengths α and β, respectively. We assume that the *sanctioner* does not distinguish between her residual income relative to Players 1 and 2's terminal incomes; hence, parameters α and β of the *sanctioner* are the same when applied to Players 1 and 2.

In total, the *sanctioner's* utility in the case of punishment is

$$U_3P = w_3 - x_3 - \alpha max((X_1 - kx_3 - (w_3 - x_3), 0)) - \alpha max((X_2 - (w_3 - x_3), 0)) - \beta max((w_3 - x_3 - -(X_1 - kx_3), 0)) - \beta max((w_3 - x_3) - X_2, 0)) - \gamma|X_1 - kx_3 - X_2|, \quad \text{(Eq. 2)}$$

where k (=2) is the punishment efficiency parameter — the number of MUs taken from Player 1 if that player is punished by x_3 ($x_3 \leqslant w = 18$). This utility function is maximized with respect to x_3 (punishment size), and it reaches a maximum when all terms involving x_3 are brought to 0, that is, when the shares of Players 1 and 2 are exactly equal. A rational Player 3 (*sanctioner*) with these preferences will punish if Eq.2 > Eq.1.

Proposition: A unique equilibrium punishment strategy of Player 3 with utilities given by Eq.1 and Eq.2 is

$x_3 = 0$ *if* $X_1 < 20$,

$x_3 = X_1 g/(g - 1) - 20(g + 1)/(g - 1)$ *if* $20 < X_1$ *and* $x_3 < 18$, *where* $g = \alpha + \beta + 2\gamma$,

$x_3 = 18$ *if* $x_3 \geqslant 18$.

Equilibrium punishment strategy and model predictions are the following: the more unfair the split the *sanctioner* observes, the stronger the punishment she assigns to Player 1, until the maximum number of MUs allowed, 18.

Positive punishment should take place whenever Eq. 1 < Eq. 2. By construction, $X_1 > w_3$ whenever $w_3 > X_2$, hence Eq.1 equals either

$$U_3^{Na} = w_3 - \alpha max(X_1 - w_3, 0) - \beta max(w_3 - X_2, 0) - \gamma(X_1 - X_2) \quad \text{(Eq.3)}$$

or

$$U_3^{Nb} = w_3 - \alpha max(X_2 - w_3, 0) - \beta max(w_3 - X_1, 0) - \gamma(X_2 - X_1) \quad \text{(Eq.4)}$$

provided $X_1 \neq X_2$ (otherwise, there is no inequity and no reason to punish at all). When punishment takes place, Eq. 2 becomes either

$$U_3^{Pa} = w_3 - x_3 - \alpha max((X_1 - kx_3 - (w_3 - x_3), 0)) - \beta max((w_3 - x_3) - X_2, 0)) - \gamma(X_1 - kx_3 - X_2) \quad \text{(Eq.5)}$$

or

$$U_3^{Pb} = w_3 - x_3 - \alpha max((X_2 - (w_3 - x_3), 0)) - \beta max((w_3 - x_3 - (X_1 - kx_3), 0)) - \gamma(X_2 - X_1 + kx_3) \quad \text{(Eq.6)}$$

given the definition of $X_1 = 40 - X_2$, and k, it is straightforward to see that $X_1 - kx_3 - (w_3 - x_3)$ and $X_2 - (w_3 - x_3)$ cannot be both > 0.

Hence, we can limit attention to two possible cases:

$X_1 > X_2$, and choice between Eq.3 and Eq.5, and

$X_1 < X_2$, and choice between Eq.4and Eq.6.

Consider them in turn:

Case 1: $X_1 > X_2$

Decision to punish takes place when $U_3^{Na} < U_3^{Pa}$, i. e.

$$w_3 - \alpha(X_1 - w_3) - \beta(w_3 - X_2) - \gamma(X_1 - X_2) < w_3 - x_3 - \alpha(X_1 - kx_3 - (w_3 - x_3)) - \beta(w_3 - x_3) - X_2) - \gamma(X_1 - kx_3 - X_2)$$
$$=> 0 < x_3(\alpha + \beta + 2\gamma - 1) \quad \text{(Eq.7)}$$

which holds true whenever α, β, γ are all positive and large enough. Hence in this case player 3 should punish until the condition is satisfied, i.e. up to the tipping point when Eq.7 becomes violated. From Eq.5, it is straightforward to see that utility increases in x_3 and is given by

$$x_3 = X_1(\alpha + \beta + 2\gamma)/(\alpha + \beta + 2\gamma - 1) - (w_3(\alpha - \beta + 1) + 40(\beta + 2\gamma))/(\alpha + \beta + 2\gamma - 1) =$$
$$= X_1(\alpha + \beta + 2\gamma)/(\alpha + \beta + 2\gamma - 1) - 20(\alpha + \beta + 2\gamma + 1)/(\alpha + \beta + 2\gamma - 1) \quad \text{(Eq.8)}$$

as stated in the proposition. If $g = \alpha + \beta + 2\gamma > 1$, this strategy increases from $X_1 = 20$ to the maximum punishment allowed of 18 at a rate greater than 1, up to the point where $X_1 - 20 - x_3 > 0$. This last condition is satisfied provided $X_1 - 20 - X_1\gamma/(\gamma - 1) - 20(\gamma + 1)/(\gamma - 1) = (40 - X_1)/(\gamma - 1) > 0$, which is true for any value of X_1.

Case 2: $X_1 < X_2$

Punishment takes place whenever $U_3^{Na} < U_3^{Pb}$, i.e.

$$w_3 - \alpha(X_2 - w3) - \beta(w_3 - X_1) - \gamma(X_2 - X_1) < w_3\ x_3 - \alpha(X_2 - (w_3 - x_3)) - \beta(w_3 - x_3 - (X_1\ kx_3)) - \gamma(X_2\ X_1 + kx_3)$$
$$=> 0 < - x_3(\alpha + \beta + 2\gamma + 1) \quad \text{(Eq.9)}$$

which condition can never be true if the parameters are positive.

In sum, the inequity aversion model (Fehr & Schmidt, 1999; Svedsater & Johansson, 2005) predicts that third-party punishment (as given by Eq.8) increases linearly if $X_1 > 20$, up to the maximum allowed amount of 18, and is zero otherwise.

Model Predictions

To elaborate on our results, we used a computational model: participants' material costs were defined by parameters α and β, where α represents disadvantageous inequality and β represents advantageous inequality. The model predicts no third-party punishment of equal (20:20) splits and a stronger third-party punishment of more unfair splits.

The model implies that the third-party punishment decision depends on two key components: *material costs* (underlined by rDLPFC activity) and *moral costs* (underlined by rTPJ activity). *Material costs* of third-party punishment increase as the *sanctioner* pays higher amounts to punish the unfair behavior of the *dictator*. In contrast, for *moral costs*, the *sanctioner* is better off the more she pays to punish the *dictator's* unfair behavior. Both components are depicted in Figure 4, which also shows the following hypothetical effect of stimulation as predicted by the model: anodal tDCS of the rTPJ, where the activation of moral feelings increases the parameter and changes the slope of the moral cost line, but only up to the point where the *material* costs do not exceed the *moral* ones. This may suggest an optimum equilibrium of *material* and *moral costs* for third-party punishment, where the further increase of punishment increases the *material costs*, while a *decrease* of punishment would increase the *moral costs*.

The discussion above suggests that anodal tDCS of the rTPJ could increase the marginal utility of the moral costs, which could shift the utility function of such costs (Figure 4, optimal punishment point x_3 changes to x_3^*), and consequently, decrease third-party punishment. This effect of anodal tDCS of the rTPJ was indeed observed in Study 1 (Condition 1.2). In contrast, when paralleled with cathodal stimulation of the rDLPFC, anodal tDCS of the rTPJ should increase the marginal utility of the moral costs while decreasing the marginal utility of the material costs, consequently increasing third-party punishment. Interestingly, we observed a trend of this tDCS effect in Study 2 (Condition 2.2).

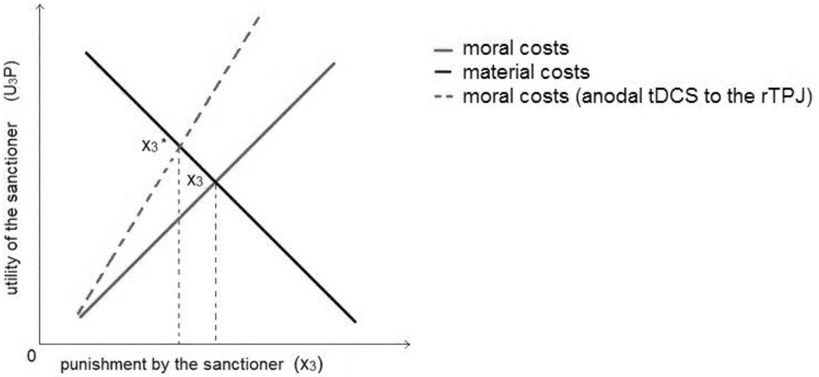
Importantly, in our version of the dictator game, third-party punishments of extremely unfair splits are quite costly. Accordingly, the model implies a strong conflict between *material* and *moral costs* when participants punish such splits. Interestingly, only in the case where participants observe moderately unfair 30:10 splits, the conflict between material and moral costs is minimal, since participants either do not punish or punish extremely little if the material costs are high. Therefore, tDCS could have the strongest effect on the third-party punishment of 30:10 splits, as increased *moral costs* would not conflict with marginally affected *material costs*. Thus, the model could explain our findings of the significant effect of tDCS on third-party punishment of slightly unfair 30:10 splits. Importantly, the model suggests that only for moderate (30:10) splits were the subjects able to maximize the utility of third-party punishment, and at the same time, minimize all players' inequity. Interestingly, third-party punishment of 30:10 splits creates a Pareto optimal distribution of MUs (10:10:10), where it is impossible to improve the income of one player without worsening the incomes of the other players.

Discussion

Our study demonstrated that anodal tDCS to the rTPJ decreased third-party punishment of moderately unfair splits during the dictator game. Our finding is

Figure 4

Hypothetical Scenario of Moral and Material Costs' Interaction During Third-Party Punishment


— moral costs
— material costs
-- moral costs (anodal tDCS to the rTPJ)
utility of the sanctioner (U3P)
punishment by the sanctioner (X3)
0
X3*
X3

Note. The intersection of lines representing material costs and moral costs indicates an optimal punishment decision (x_3), while x_3* represents an optimal punishment decision when the moral costs are affected by anodal tDCS to the rTPJ.

consistent with the recent TMS study (Baumgartner et al., 2014), which demonstrated that an inhibition of the rTPJ decreased the parochial punishment of out-group members.

A previous study showed that anodal tDCS applied to the rTPJ led to less blame for accidental harms during a moral judgment task (Sellaro et al., 2015). This suggests that rTPJ is involved in processing the agent's moral intentions. Recent meta-analysis suggests that rTPJ showed significant activation when making one's own moral decisions (Garrigan, Adlam, & Langdon, 2016). Thus, rTPJ activity in our study could underlie the processing of the *dictator*'s mental state — her moral intentions. Alternatively, it could reflect thinking about the consequences of the third-party's own decision and how harmful it would be for others. Thus, anodal stimulation of this area could exaggerate the latter process and consequently diminish the punishment. Overall, anodal tDCS of the rTPJ could affect the perceived degree of the moral norm violation, and consequently decrease the assigned blame and punishment of the *dictator*.

Our results only partly support Hypothesis I, since we did not find a significant effect of anodal tDCS of the rDLPFC on third-party punishment. A previous study showed that the suppression of rDLPFC by TMS leads to increased third-party punishment (Brüne et al., 2012). Importantly though, in this previous study, third-party punishment was combined with helping behavior — the *recipients*' payoffs increased by the same amount that was taken from the *dictators*' budget by the *sanctioners*. A recent study suggested that the rDLPFC is especially activated during the helping behavior of third parties (Hu, Strang, & Weber, 2015). Thus, one possible explanation for the discrepancy with our results is that, in our paradigm, third-party punishment was not associated with helping behavior.

Interestingly, anodal tDCS of the rTPJ significantly decreased third-party punishment only in *US_equal-trials*, where third-party punishment was able to establish equality between all players. In our study, third-party punishment of 30:10 splits created a Pareto optimal distribution of MUs (10:10:10), where it was impossible to improve the income of one player without worsening the incomes of the other players. . Our model predicted a minimal conflict between the material and moral costs of third-party punishment in moderately unfair (30:10) splits. Thus, only the punishment of 30:10 splits could maximize the utility of third-party punishment and minimize the inequity of all players. Overall, we speculate that anodal tDCS of the rTPJ could increase the marginal utility of moral costs, which could shift the utility function of the moral costs and decrease third-party punishment.

In Study 2, we simultaneously applied cathodal tDCS to the rDLPFC and anodal tDCS to the rTPJ. A previous fMRI study demonstrated that TPJ activity during third-party punishment is paralleled by an initial deactivation of the DLPFC (Buckholtz et al., 2008). We found only a trend of effect of the reciprocal stimulation protocol on third-party punishment and failed to confirm Hypothesis II. A recent meta-analysis showed that the excitatory effect of anodal tDCS is replicable in cognitive studies, while the cathodal stimulation effect is not stable and rarely leads to inhibition (Jacobson, Koslowsky, & Lavidor, 2012). In our study, cathodal tDCS of the rDLPFC could have led to a marginal effect of rDLPFCc/rTPJa stimulation. Overall, the trend of an increment of third-party punishment after rDLPFCc/rTPJa stimulation in our study could indicate an effect of tDCS on the interaction of the default mode network (TPJ) and central executive network (rDLPFC). Additional studies are needed to investigate the effects of the rDLPFCc/rTPJa tDCS protocol.

To further probe the frontoparietal interactions during third-party punishment, follow-up studies could combine brain stimulation and brain imaging techniques. Electroencephalography coherence as a measure of functional cortical connectivity on a centimeter scale (Srinivasan, Nunez, & Silberstein, 1998; Nunez & Srinivasan, 2006) could offer a tool for studying TPJ/DLPFC synchronization during third-party punishment decisions.

Conclusion

Our study demonstrates that anodal tDCS of the rTPJ decreases third-party punishment of moderately unfair behavior when the participants have an opportunity to restore equality in their social groups. We found only a small, insignificant trend in the effect of simultaneous anodal tDCS of the rTPJ and cathodal tDCS of the rDLPFC on third-party punishment. Overall, our findings support the critical role of the rTPJ in third-party punishment.

Acknowledgement

We thank Maartje Elisa Dankbaar, MSc, for her assistance in stimulus preparation; Dr. Matteo Feurra for his expertise in tDCS; and Dr. Alexei Zakharov, who provided the original behavioral data of the dictator game. This study used the HSE Synchronous Eye-tracking, Brain Signal Recording and Non-Invasive Brain Stimulation System.

References

Baumgartner, T., Götte, L., Gügler, R., & Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping, 33*(6), 1452–1469. doi:10.1002/hbm.21298

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R., & Knoch, D. (2014). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Social Cognitive Affective Neuroscience, 9*(5), 653–660. doi:10.1093/scan/nst023

Bendor, J., & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology, 106*(6), 1493–1545. doi:10.1086/321298

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57*(1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Brüne, M., Scheele, D., Heinisch, C., Tas, C., Wischniewski, J., & Güntürkün, O. (2012). Empathy moderates the effect of repetitive transcranial magnetic stimulation of the right dorsolateral prefrontal cortex on costly punishment. *PloS ONE,* 7(9), e44747. doi:10.1371/ journal.pone.0044747

Buckholtz, 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

Buckholtz, J. W., Martin, J. W., Treadway, M. T., Jan, K., Zald, D. H., Jones, O., & Marois, R. (2015). From blame to punishment: Disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron, 87*(6), 1369–1380. doi:10.1016/j.neuron.2015.08.023

Elster, J. (1989). Social Norms and Economic Theory. *The Journal of Economic Perspectives, 3*(4), 99–117.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics, 14*(4), 583–610. doi:10.1007/s10683-011-9283-7

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

Fehr, E., & Schmidt, K. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868. Retrieved from https://www.jstor.org/stable/2586885

Fliessbach, K., Phillipps, C. B., Trautner, P., Schnabel, M., Elger, C. E., Falk, A., & Weber, B. (2012). Neural responses to advantageous and disadvantageous inequity. *Frontiers in Human Neuroscience*, *8*(6), 165. doi:10.3389/fnhum.2012.00165

Garrigan, B., Adlam, A. L., & Langdon, P. E. (2016). The neural correlates of moral decision-making: A systematic review and meta-analysis of moral evaluations and response decision judgements. *Brain and Cognition, 108,* 87–97. doi:10.1016/j.bandc.2016.07.007

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining.*Journal of Economic Behavior and Organization, 3*(4), 367–388.

Hu, Y., Strang, S., & Weber, B. (2015). Helping or punishing strangers: neural correlates of altruistic decisions as third-party and of its relation to empathic concern. *Frontiers in Behavioral Neuroscience*, *9*, 24. doi:10.3389/fnbeh.2015.00024

Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Experimental Brain Research*, 216(1), 1-10. doi:10.1007/s00221-011-2891-9

Kahneman, D., Knetsch, J., & Thaler, R. (1986). Fairness and the Assumptions of Economics.*Journal of Business, 59*, 5285–5300.

Keeser, D., Meindl, T., Bor, J., Palm, U., Pogarell, O., Mulert, C., ... Padberg, F. (2011). Prefrontal transcranial direct current stimulation changes connectivity of resting-state networks during fMRI. *Journal of Neuroscience*, *31*, 15284–15293.

Krueger, F., & Hoffman, M. (2016). The emerging neuroscience of third-party punishment. *Trends in Neurosciences*, *39*(8), 499–501. doi: 10.1016/j.tins.2016.06.004

Nitsche, M. A., Nitsche, M. S., Klein, C. C., Tergau, F., Rothwell, J. C., & Paulus, W. (2003). Level of action of cathodal DC polarisation induced inhibition of the human motor cortex. *Clinical Neurophysiology, 114*(4), 600–604. doi: 10.1016/S1388-2457(02)00412-1

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology*, *57*(10), 1899–1901. doi:10.1212/ WNL.57.10.1899

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG* (2nd ed.). New York: Oxford University Press.

Paulus, W. (2011). Transcranial electrical stimulation (tES - tDCS; tRNS, tACS) methods. *Neuropsychological Rehabilitation, 21*(5), 602–617 doi: 10.1080/09602011.2011.557292

Peña-Gómez, C., Sala-Lonch, R., Junqué, C., Clemente, I. C., Vidal, D., Bargalló, N., & Bartrés-Faz, D. (2012). Modulation of large-scale brain networks by transcranial direct current stimulation evidenced by resting-state functional MRI. *Brain Stimulation, 5*(3), 252–263. doi:10.1016/j.brs.2011.08.006

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342*(6157), 482–484. doi:10.1126/science.1241399

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology, 22*(23), 2274–2277. doi:10.1016/j.cub.2012.10.018

Sellaro, R., Güroğlu, B., Nitsche, M. A., van den Wildenberg, W. P., Massaro, V., Durieux, J., ... Colzato, L. S. (2015). Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction. *Neuropsychologia, 77,* 400–408. doi:10.1016/j.neuropsychologia.2015.09.016

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56*(1), 185–196. doi:10.1016/j.neuron.2007.09.011

Srinivasan, R., Nunez, P. L., & Silberstein, R. B. (1998). Spatial filtering and neocortical dynamics: estimates of EEG coherence. *IEEE Transactions on Biomedical Engineering, 45*(7), 814–826. doi:10.1109/10.686789

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A.T. (2015). Be nice if you have to – the neurobiological roots of strategic fairness. *Social Cognitive and Affective Neuroscience, 10*(6), 790–796. doi:10.1093/scan/nsu114

Svedsater, H., & Johansson, L. O. (2005). Beyond egocentric judgments of fairness: Advantageous, disadvantageous, and third-party inequality aversion. *IACM 18th Annual Conference,* 30. doi:10.2139/ssrn.736225

Thielscher, A., Antunes, A., & Saturnino, G. B. (2015). Field modeling for transcranial magnetic stimulation: a useful tool to understand the physiological effects of TMS? *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015,* 222–225. doi:10.1109/EMBC.2015.7318340

Zinchenko, O., & Klucharev, V. (2017). Commentary: The emerging neuroscience of third-party punishment. *Frontiers in Human Neuroscience, 11,* 512. doi: 10.3389/fnhum.2017.00512

Oksana Zinchenko – junior research fellow, Institute of Cognitive Neuroscience, Centre for Cognition and Decision Making, National Research University Higher School of Economics.
Research area: cooperation, social norms, social punishment.
E-mail: ozinchenko@hse.ru

Alexis Belianin – adviser, International College of Economics and Finance; laboratory head, senior research fellow, International Laboratory for Experimental and Behavioural Economics, National Research University Higher School of Economics, Ph.D.
E-mail: abelianin@hse.ru

Vasily Klucharev – director, Institute of Cognitive Neuroscience; leading research fellow, Centre for Cognition and Decision Making, National Research University Higher School of Economics, Ph.D., professor.
E-mail: vklucharev@hse.ru

Роль височно-теменно-затылочной области и дорсолатеральной префронтальной коры правого полушария в социальном наказании третьей стороной: исследование с применением транскраниальной электрической стимуляции

О.О. Зинченко[a], А.В. Белянин[a], В.А. Ключарев[a]

[a] Национальный исследовательский университет «Высшая школа экономики», 101000, Россия, Москва, ул. Мясницкая, д. 20

Резюме

Предыдущие исследования показали, что дорсолатеральная префронтальная кора и височно-теменно-затылочная область правого полушария вовлечены в поддержание поведения подчинения социальным нормам: их активация связана с поведением третьей стороны по упрочению социальной нормы, известным как социальное наказание третьей стороной. Ввиду противоречивых результатов нейроимиджинговых исследований и исследований с использованием методов стимуляции мозга, настоящее исследование ставит своей целью прояснение роли дорсолатеральной префронтальной коры и височно-теменно-затылочной области правого полушария в социальном наказании третьей стороной. Несмотря на значительный прогресс в изучении нейрональных основ социального наказания третьей стороной, паттерны взаимодействия дорсолатеральной префронтальной коры и височно-теменно-затылочной области правого полушария остаются не до конца изученными. В связи с этим была выдвинута следующая гипотеза: решение третьей стороны о необходимости социального наказания связано с активностью нейрональной сети, включающей дорсолатеральную префронтальную кору и височно-теменно-затылочную область правого полушария. Мы использовали метод транскраниальной электрической стимуляции для независимой или одновременной (реципрокной) стимуляции данных областей мозга в ходе выполнения игры «Диктатор» с наказанием третьей стороной. Нами был обнаружен эффект анодной транскраниальной электрической стимуляции височно-теменно-затылочной области правого полушария, приводящей к уменьшению социального наказания третьей стороной. Применение реципрокной стимуляции (анодной стимуляции височно-теменно-затылочной области правого полушария и катодной стимуляции дорсолатеральной префронтальной коры правого полушария) привело к увеличению социального наказания третьей стороной на уровне статистического тренда. Активность височно-теменно-затылочной области правого полушария может быть связана с модуляцией воспринимаемых моральных потерь в социальном наказании третьей стороной.

Ключевые слова: дорсолатеральная префронтальная кора, височно-теменно-затылочная область, социальное наказание третьей стороной, социальные нормы, транскраниальная электрическая стимуляция.

Зинченко Оксана Олеговна — младший научный сотрудник, Институт когнитивных нейронаук, Центр нейроэкономики и когнитивных исследований, Национальный исследовательский университет «Высшая школа экономики».
Сфера научных интересов: кооперация, социальные нормы, социальное наказание.
Контакты: ozinchenko@hse.ru

Белянин Алексей Владимирович — советник, Международный институт экономики и финансов; заведующий лабораторией, Международная лабораториая экспериментальной и поведенческой экономики, Национальный исследовательский университет «Высшая школа экономики», Ph.D.
Контакты: abelianin@hse.ru

Ключарев Василий Андреевич — директор, Институт когнитивных нейронаук, Национальный исследовательский университет «Высшая школа экономики», кандидат биологических наук, профессор.
Контакты: vklucharev@hse.ru