The role of ventromedial prefrontal cortex and temporo-parietal junction in third-party punishment behavior

The role of ventromedial prefrontal cortex and temporo-parietal junction in third-party punishment behavior.

Emanuele Lo Gerfo [1, 2, 8, 9] *†, Alessia Gallucci [3,8,†], Rosalba Morese,[4,5], Alessandra Vergallito [3,8] Stefania Ottone [1,8,9] Ferruccio Ponzano [7,9] Gaia Locatelli [3], Francesca Bosco [4,6], Leonor Josefina Romero Lauro [3,8,9]

1) *Department of Economic, Management and Statistics, University of Milano Bicocca, Italy.*
2) *Clinical Psychology Service of Mediterranean Institute for Transplantation and Advanced Specialized Therapies (IRCSS IsMeTT)*
3) *Department of Psychology, University of Milano Bicocca, Italy*
4) *Department of Psychology, University of Turin, Turin, Italy*
5) *Faculty of Communication Sciences, Università della Svizzera italiana, Lugano, Switzerland*
6) *Neuroscience Institute of Turin, University of Turin, Italy*
7) *University of Eastern Piedmont, Department of Law and Political, Economic and Social Sciences, Italy.*
8) *NeuroMi - Milan Center for Neuroscience, Italy*
9) *CISEPS, University of Milano Bicocca, Italy*

**Corresponding author: Emanuele Lo Gerfo, Economic, Management and Statistics, University of Milano Bicocca, Piazza dell'Ateneo Nuovo 1, Milan, Italy. Email: emanuelelogerfo@gmail.com.*
†These authors contributed equally to the manuscript.

ABSTRACT

Third parties punish, sacrificing personal interests, offenders who violate either fairness or cooperation norms. This behavior is defined altruistic punishment and the degree of punishment typically increases with the severity of the norm violation. An opposite and apparently paradoxical behavior, namely anti-social punishment, is the tendency to spend own money to punish cooperative or fair behaviors. Previous fMRI studies correlated punishment behavior with increased activation of reward system areas (e.g. the ventromedial prefrontal cortex, VMPFC), the mentalizing (e.g. the temporoparietal junction, TPJ) and central-executive networks. In the present study, we aimed at investigating the causal role of VMPFC and TPJ in punishment behaviors through the application of anodal transcranial direct current stimulation (tDCS).

Sixty healthy participants were randomly assigned to three tDCS conditions: (1) anodal tDCS over VMPFC, (2) anodal tDCS over TPJ, (3) sham stimulation. At the end of the stimulation, participants played a third-party punishment game, consisting in viewing a series of fair or unfair monetary allocations between unknown proposers and recipients. Participants were asked whether and how much punishing proposers using their own monetary endowment. To test membership effects, proposers and recipients could be either Italian or Chinese.

Anodal tDCS over VMPFC increased altruistic punishment behavior whereas anodal tDCS over TPJ increased anti-social punishment choices compared with sham condition, while membership did not influence participant's choices. Our results support the idea that the two types of punishment behaviors rely upon different brain regions, suggesting that reward and mentalizing systems, underlie respectively altruistic and anti-social punishment behaviors.

Key words: tDCS, VMPFC, TPJ, altruistic punishment, antisocial punishment.

Highlight:

Altruistic punishment and antisocial punishment rely upon different brain regions
Anodal tDCS influences punishing behaviors
VMPFC and TPJ are implicates in altruistic and antisocial punishment
Reward and mentalizing systems, underlie altruistic and anti-social punishment

1.1 Introduction

Complex social norm systems, which regulate small and big social groups, distinguish human beings from other animal species and are fundamental for survival and for the functioning of human society (Fehr & Rockenbach, 2004; Helbing et al., 2010). Indeed, respecting spoken or unspoken shared social rules promotes cooperation and leads human social behavior (Bendor & Swistak, 2001; Elster, 1989; Ostrom, 2000). It has been suggested that "evolution built us to punish cheater" (Hoffman, 2014, pag 1), stressing how punishment instinct enabled us to live in small groups, allowing to benefit

of the mutual defense, division of the labor and revealing its fundamental role in preserving cooperation (Boyd et al., 2010).

Two puzzling sanctioning behaviours are altruistic and antisocial punishment. Both of them represent a costly form of punishment whose effect is not a direct material benefit maximization (Sääksvuori et al., 2011). The term altruistic punishment originates within the field of behavioral experimental economics and describes a scenario where punishment is addressed to people who violate shared norms (i.e. they behave unfairly) (Fehr & Fischbacher, 2003; Goette et al., 2012; Riedl et al., 2012). The most relevant form of altruistic punishment is represented by Third-Party Punishment (TPP). TPP occurs when people implement sanctioning mechanisms even when they are impartial bystanders, so called "third parties", that is when they are not directly affected by others' unfair behavior (Fehr & Fischbacher, 2003; Ostrom, 2000; Riedl et al., 2012). TPP has been acknowledged as a relevant "social norm enforcement device" (Fehr et al., 2002).

Although the altruistic punishment, by definition, does not involve any overt benefit for the punisher, actually the satisfaction of revenge, the experience of power and the expectation of future rewards, as secondary advantages, could enforce it (Jordan et al., 2016; Strobel et al., 2011). Moreover, altruistic punishment differs in interactions with in-group members and out-group members, namely the in-group condition protects or favors the members of own group from those of the others (De Dreu et al., 2010; Goette et al., 2012; Halevy et al., 2012; Henrich et al., 2005; Levine et al., 2005; Tajfel & Turner, 1986). In a recent study, Rabellino et al. (2016) investigated the altruistic punishment using TPP game in in-group and out-group contexts in which the membership differed for nationality (Chinese or Italian). Behavioral results demonstrated that this kind of punishment behavior emerged as a tendency to protect in-group victims of unfair behavior. Bernhard et al. (2006) defined this difference in altruistic behavior between in-group and out-group interactions as parochial altruism.

The opposite behavior, called antisocial punishment, is instead the tendency to spend own resources to punish cooperative or fair behaviors (Nikiforakis, 2008). Even if the attempts to explain antisocial

punishment are still seminal, the fact that usually non-cooperative subjects implement it lies on some possible motivations. It may represent a form of retaliation on cooperators who punished free riders, as well as an attempt to discourage cooperative behavior due either to preferences for competition or to preferences for conformism when cooperation is not the shared rule (Herrmann et al., 2008). According to Herman et al. 2008, some bargaining experiments (Bahry & Wilson, 2006; Henrich et al 2007; Hennig-Schmidt et al 2008) showed antisocial punishment that could also considered as a form of do-gooder derogation. In these researches people reject fair and hyperfair proposes. According to the authors, people might be suspicious of others who appear too generous.

Economic games represent valid means to explore humans' punishment behaviors, finding one of their main application as behavioral tasks in neuroimaging studies interested in shedding light on neural substrates of punishment behaviors (Boyd et al., 2010; Strobel et al., 2011). In particular, the TPP game has been effectively used. In a typical TPP game an impartial bystander (a third party, player C) can decide, spending part of his endowment, to punish a player (a dictator, player A) who allocates fair or unfair amount of money to a dummy player (a receiver, player B) (Fehr & Fischbacher, 2004; Ottone et al., 2015).

1.1.2 Neural correlates of sanctioning behavior

A growing number of social neuroscience and neuroeconomics evidence converged in showing correlations between participants' punishment responses in economic games and functional activity of different cerebral networks (Buckholtz et al., 2008). Particularly, the implicated networks include: the salience network, which detect the risk or the presence of norm violations, composed by the anterior insula, dorsal anterior cingulate cortex, amygdala and putamen (Feng et al., 2016; Güroğlu et al., 2011; Harlé et al., 2012; Krueger & Hoffman, 2016; Sanfey et al., 2003); the default mode network, which modulates the emotional processing of harming a victim and the representation of others' intentions, includes the medial prefrontal cortex; the mentalizing network, comprising the

dorsomedial prefrontal cortex and the temporoparietal junction (TPJ) (Feng et al., 2018; Güroğlu et al., 2011; Krueger & Hoffman, 2016; Bosco et al., 2017); the central-executive network, anchored to the posterior parietal cortex and the dorsolateral prefrontal cortex (DLPFC), which transforms signals coming from the default mode network into punishment behaviors (Krueger & Hoffman, 2016; Zinchenko & Arsalidou, 2018); the reward network, involving the nucleus accumbens and the ventromedial prefrontal cortex (VMPFC) (De Quervain et al., 2004; Hu et al., 2015). These neural hubs seem to have a general role in both norms' representation and violation processing (Zinchenko & Arsalidou, 2018). Differences in neural responses were reported when the role of group membership was investigated. Indeed, a recent fMRI study (Morese et al., 2016), comparing subjects' punishing behavior between in-group vs out-group settings in a TPP game, showed that observing in-group norm violation was associated with an increased activation of mentalizing network. Interestingly, a previous study (Baumgartner et al., 2012) converged in supporting the hypothesis that the recruitment of mentalizing network could be explained by subjects' attempts to understand or justify in-group norm violation. Concerning TPJ role, some authors speculated an antagonistic relationship between this region and the DLPFC during TPP (Krueger & Hoffman, 2016). Indeed, the DLPFC showed an initial deactivation when increased activity of TPJ was recorded, immediately followed by increased responses when subjects decided to punish. As previously mentioned, TPJ is involved in assessing the blame of violators, while the DLPFC, being part of the central executive network, is responsible of converting the evaluation into the decision to punish. Therefore, the biphasic activity of the DLPFC could underlie the inhibitory action of executive network over the mentalizing system when planning punishment behaviors is needed. The TPJ's right portion especially was generally found to have a high specialization for mentalizing (Saxe & Powell, 2006; Saxe et al., 2009; Young et al., 2010) that is a crucial to interact in the social environment. For instance, a study showed that, compared with healthy controls, right TPJ responses of autism spectrum patients were similar for both mentalizing and physical judgments, with anomalous right TPJ activations of patients correlating with the degree of their social impairment (Lombardo et al.,

2011). With a focus on sanctioning behavior, several neuroimaging studies support the specific role of the right TPJ in altruism, highlighting the involvement of this region when considering the tradeoff between the spontaneous altruistic tendencies and the costs of the altruistic actions (Morishima et al., 2012). Moreover, scholars reported correlations between responses of right TPJ and the subjective value of sanctioning (Zhong et al., 2016) as well as a causal relationship between right TPJ activity and parochial punishment (Baumgartner et al., 2013) in TPP.

Regarding antisocial punishment, Morese et al. (2016), through exploratory analysis, reported a specific role of ventromedial prefrontal cortex. Hence, this region together with the right TPJ seem to be crucial for punishing unfair (i.e. altruistic punishment) and fair (i.e. antisocial punishment) behaviors (Baumgartner et al., 2012; Bellucci et al., 2017; Morese et al., 2016). However, the attempts to explore neural correlates of punishing behaviors are still at the beginning.

With the present study, we aimed to fill this gap by investigating, through the application of the transcranial direct current stimulation (tDCS), the causal role of reward and mentalizing networks, in particular of VMPFC and TPJ, in the altruistic and the antisocial punishment behaviors. TDCS has already been shown to be an effective tool in modulating punishment behaviors (Civai et al., 2014; Hämmerer et al., 2016; Keeser et al., 2011; Peña-Gómez et al., 2012; Polanía et al., 2015). Moreover, to the best of our knowledge none study so far has investigated the tDCS effects on punishment behaviors in the context of TPP as well as the mechanisms of antisocial punishment.

Building on previous evidence, we expect that tDCS would modulate altruistic punishment when applied to both VMPFC and right TPJ whereas only tDCS over VMPFC would modulate antisocial punishment. In line with previous evidence on parochial altruism we expected participants to punish more frequently an outgroup member as dictator making unfair offers to an ingroup member as receiver in the sham conditions and that stimulation over right TPJ would modulate such behavior, as suggested by Morese et al. (2016) results.

Finally, recent studies showed that people reactions to other player's choices in economics game are affected by the concern people have for others, that is prosociality (Bieleke et al., 2017; Camerer & Fehr, 2003), empathy and racial prejudice (Kirman et al., 2010; Morese et al., 2016; Stanley et al., 2011). Hence, we administered the Social Value Orientation slider measure (SVO; Murphy et al., 2011), the Interpersonal Reactivity Index (IRI; Davis, 1980) and the Implicit Association Test (IAT; Greenwald et al., 1998) to measure individual prosociality, empathy and prejudice and their impact on punishment behaviors.

1.2. Material and methods

1.2.1 Participants

Sixty healthy Italian students participated in the experiment (25 males, mean age = 23, SD ± 2.5). Participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected to normal vision, no clinical history of neurological or psychiatric disorders nor other specific contraindications to non-invasive brain stimulation (Rossi et al., 2009). Each participant completed the Adult Safety Screening Questionnaire (Keel et al., 2001) and gave informed written consent prior to study procedures. The experiment took place at the University of Milano-Bicocca with the approval of the local Ethic Committee and was carried out in accordance with the ethical standards of the revised Helsinki Declaration.

1.2.2 Experimental Design and Procedures

The experimental procedure was divided in two different sessions.

Session one: Psychological traits

In the first session participants were asked to carry out two computerized tasks: the IAT and the SVO, plus a self-report questionnaire, i.e. the IRI.

The IAT requires to categorize stimuli belonging to two opposite categories associated with attributes with positive or negative valence. First, the stimuli and attributes are presented separated, then they are associated in pairs that can be congruent or incongruent relatively to the common feeling. The IAT assumes that a higher implicit association causes a greater difficulty in categorizing the stimuli when presented in the incongruent condition. The difference in reaction times and accuracy between the congruent and incongruent condition is considered a measure of the strength of the implicit association, assessed through the D score (Greenwald et al., 2003). In this study, the IAT was used to evaluate implicit attitudes towards Caucasian and Asian human faces. Specifically two sets of trials were compared. In the first set, Caucasian faces were paired with positive valence attributes, while Asian faces were coupled with negative valence words (congruent trials); in the second set the opposite couple were made, i.e. Caucasian/negative stimuli and Asian/positive stimuli (in-congruent trials).

In order to control the impact of prosociality, we used the SVO slider measure which calculates the choices during a series of dictator games between participants and another person. Participants' decisions lead to different SVO scores and thereby to four different categories: perfect altruism, that is maximizing the other participant's payoff; prosociality, that is sacrificing a part of one's own pay-off to give something to the other; individualism, that is maximizing one's own payoff; competitiveness, that is maximizing the difference between one's own and the other person's payoff.

After completing IAT and SVO, subjects had to fill the IRI, a 28-items questionnaire that measures empathy skills. For each item participants completed a five-points Likert scale from 1 (it does not describe my behavior) to 5 (it totally describes my behavior). IRI is divided into four subscales: the perspective-taking scale, the fantasy scale, the empathic concern scale, the personal distress scale.

Session two: tDCS procedure and Third Party Punishment game

The second session took place after a week. At the beginning of the sessions, the tDCS stimulation was performed, followed by the execution of the Third Party Punishment game.

tDCS Procedures

Participants were randomly assigned to one of three experimental conditions (20 participants for each group): 1) sham stimulation (6 males, 14 females): in this condition, half participants received a placebo stimulation over right TPJ and half over VMPFC (it had been considered as unique group) (Mattavelli et al 2019), 2) anodal tDCS over right TPJ (10 males and 10 females), 3) anodal tDCS over VMPFC (9 males and 11 females). To control for gender distribution across the three conditions we run a X square, indicating no significant difference (X (3) = 2.16, p =.33). A series of one way anova was run to control for eventual difference among the three groups for age, IAT score, IRI total score and SVO orientation (see Table 1 in supplemental material for F values and p values). Significant difference emerged among the three groups indicating that the randomization was successful. The sample size was estimated using G*Power3 (Faul et al., 2007). A sample size of 19 participants per group (a total of 57 for the entire study) would have been required to detect an effect size of 0.20 with 90% power and a = 0.05. We performed a double-blind study design, therefore participants and experimenters were blinded about the condition they were assigned.

TDCS was applied through a Brain Stim stimulator (Newronica, Milan, Italy). Electrodes' position was established through the EEG 10-20 International System. For the stimulation of the VMPFC the center of the anode was positioned in a middle point between Fp1 and Fp2 (Chib et al., 2013), while to stimulate the right TPJ it was centered on Cp6 (Santiesteban et al., 2012). In both conditions the cathode was positioned on O1 (see Figure 1 for the simulated tDCS-induced electrical field distribution in the two experimental conditions). A constant current of 1.5 mA intensity was delivered

with a 5 x 5 cm anode and a 10 x 5 cathode, in order to increase the focality of the stimulation (Nitsche et al., 2008). In the two real stimulation conditions tDCS was applied for 20 minutes. In the sham tDCS, instead, the stimulator turned off automatically after 30 seconds, a procedure which has been shown to be effective in blinding participants from their assigned condition (sham vs real tDCS, Gandiga et al., 2006; Ambrus et al., 2012; Woods et al., 2016). TDCS was delivered while participants watched a cartoon video in order to standardize the procedure (Giustolisi et al., 2018).

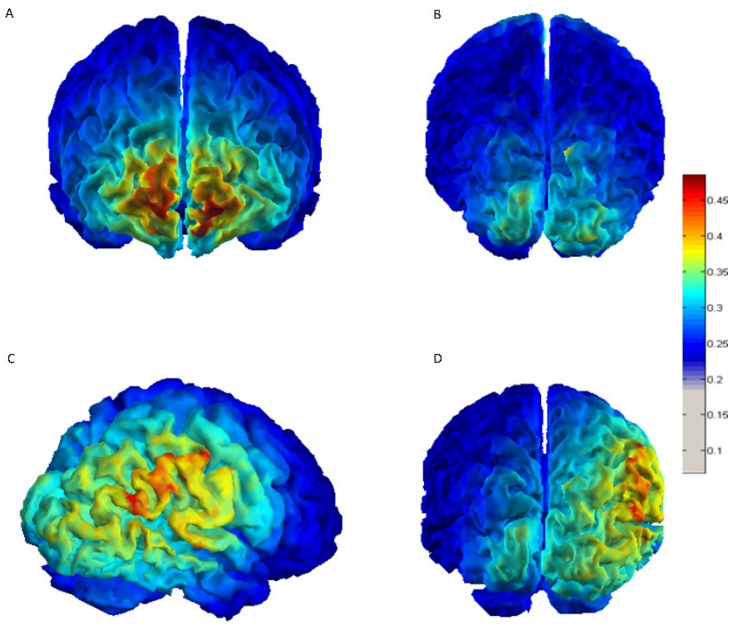

Figure 1. Computational model of tDCS-induced electric field. A simulation of the electrical field induced by the tDCS protocol used in the study was computed using Comets. Following EEG international 10-20 system, in the anodal VMPFC condition (upper figure A & B), the anode (25 cm2) was placed between Fp1 and Fp2 (panel A) and the cathode (50 cm2) was placed over O1 (panel B). In the anodal TPJ condition (panel C) the anode was placed over Cp6 and the cathode over O1 (panel D). Red colour indicates the strongest electrical field occurring over the VMPFC and rTPJ.

Third party punishment game (TPP)

Immediately after the stimulation, participants took part at a TPP game (Fehr & Fischbacher, 2004) with 160 trials (Figure 2). TPP is a modified version of the Dictator Game (Strobel et al., 2011), in

which typically two players, named A and B, interact during an economical exchange. In TPP paradigm a third player, named player C, is added. Participants were told that they would have played with other five players, being randomly assigned at the beginning of the experiment to a different role, i.e. player A, player B or player C. Actually, a preset computer program controlled everything and our experimental subjects were always player C. Player C watched sharing choices of two players, i.e. player A, the dictator, and player B, the receiver. At the beginning of each trial, player A and player C had an endowment of 20 tokens, while player B had only 10 tokens (20 tokens corresponded to 0.05 €). Each trial started with a fixation cross (with a random duration between 1000-1500 ms) followed by a first screen where player A could decide to transfer part of his endowment to player B (4000 ms duration). Actually, tokens transfers were controlled by the computer program and they, in random order, included from 0 to 5 tokens in half of trial and from 6 to 10 in the other one (but subjects were informed that the transfer could be from 0 to 20 tokens). In a second screen, player C, observing player A's transfer, could choose not to intervene in the sharing, subtracting 0 tokens to player A, or to punish him by subtracting 2, 4, 6, 8 or 10 tokens; then, as a feedback, the number of subtracted token was framed by a red square. Participants were informed that punishing player A resulted in an economic cost for them, who had to pay 1 token for each couple of tokens subtracted from player A's endowment (for example, if they wanted to subtract 6 tokens to player A, they would have a cost of 3 tokens).

Moreover, we used two combinations to manipulate group membership: an in-group condition, in which both players A and B, as the subjects enrolled in this experiment, were Italian, and out-group condition, in which at least one of the players was Chinese. During the game, national flags were shown to signal the nationality of player A and player B. To keep the attention on the task, in the 8% of trials, participants were asked to remember the nationality of player A and player B of the previous trial; a wrong answer caused a tokens' lost for player C.

At the end of the session, a questionnaire aiming at assessing subjects' fairness reference point was presented (see also Ottone et. al, 2015): participants were asked to indicate, according to their sense, the number of tokens that player A ought ideally transfer to player B. Then subjects were debriefed about tDCS and TPP game real procedures and experimental aims. They also received 2 € for the participation and an additional payment based on the amount of money earned during the SVO and the TTP game (mean = 12.5 €, SD = 2.7).

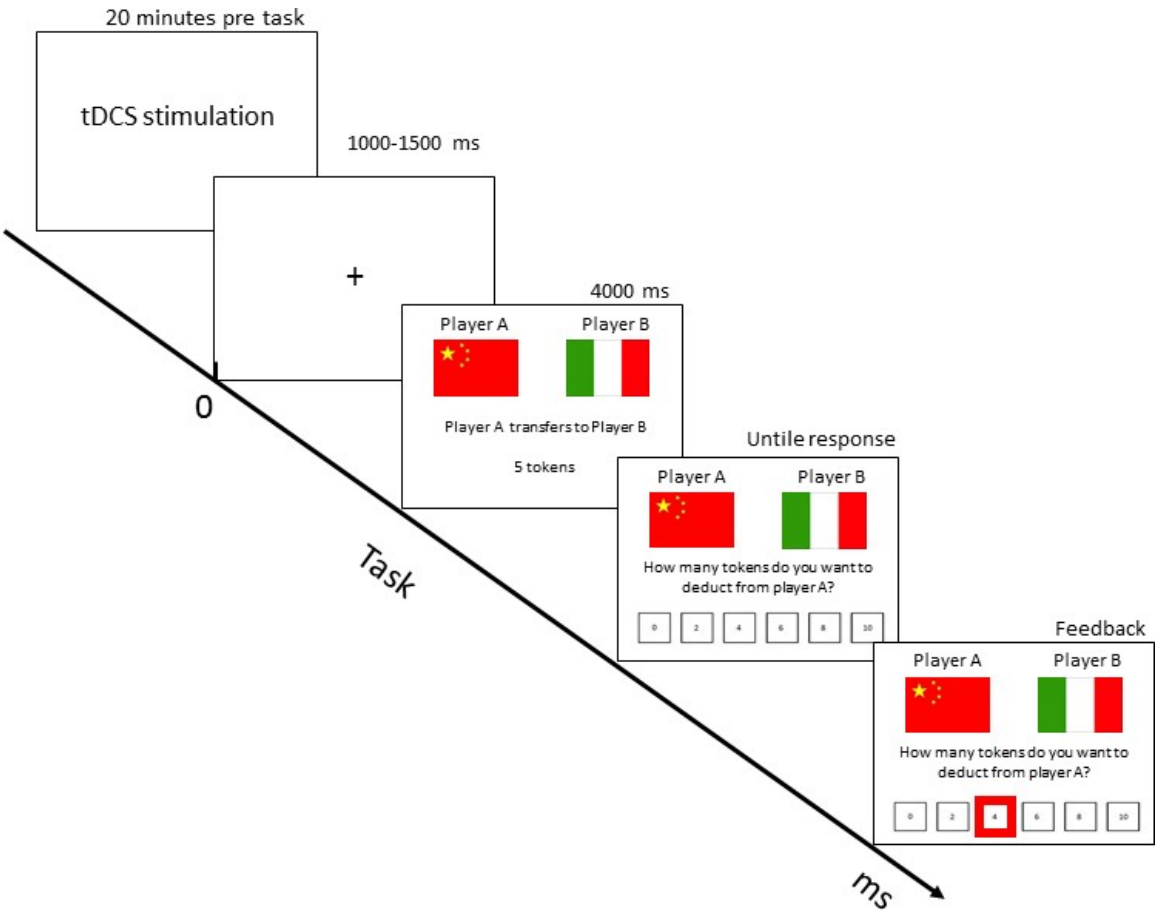

Figure 2. Schematic experimental procedure. TPP started after 20 minutes of sham or real tDCS. Each trial started with a fixation cross followed by a first screen where player A could decide to transfer part of his endowment (from 0 to 5 tokens) to player B. In a second screen, player C, observing player A's transfer, could choose not to intervene in the sharing, subtracting 0 tokens to player A, or to punish him by subtracting 2, 4, 6, 8 or 10 tokens; then, as a feedback, in the third screen the number of subtracted token was framed by a red square.

1.2.3 Data analysis

The TPP trials were classified as fair and unfair based on the participants' subjective fairness level, as assessed by the questionnaire at the end of the second session (mean of subjective fairness was 7.13 tokens; median and mode were equal to 5). We identified as fair those trials in which player A's transfer was equal or higher than the participants' reference point and as unfair those trials in which player A's transfer was lower than the participants' reference point. We calculated for each subject the decision to punish (equal to 1 if player C subtracted tokens to player A, 0 otherwise) and the amount of punishment (whose values were censored between 0 and 10) as dependent variables. We run two typical regression models: random-effect probit and a random-effect tobit regressions (McDonald & Moffitt, 1980; Gibbons & Hedeker, 1994) for decision to punish and amount of punishment respectively. In a first analysis, we considered, in both models, the following regressors: unfairness (a dummy variable equal to 1 when player A's transfer was lower than player C's fairness reference point); VMPFC and TPJ (two dummy variables for tDCS-VMPFC and tDCS- TPJ condition respectively; the control condition was SHAM tDCS); age; female (a dummy variable equal to 1 if player C was a woman); SVO angle (the higher the SVO angle value, the higher the player C's concern for the others); IAT index1. In each regression, the variable unfairness had a positive and significant effect (see Table 1 and 2). In order to isolate and study antisocial punishment (within fair trials) and altruistic punishment (within unfair trials), we performed separate second analyses for fair and unfair trials. We run a series of random-effect probit and tobit regressions with the following regressors: unfairness/fairness; in-group (a dummy variable equal to 1 if both player A and player B are Italian, 0 otherwise); FRONT and TPJ (two dummy variables for tDCS-VMPFC and tDCS-TPJ condition respectively; the control condition was SHAM tDCS); age; female (a dummy variable equal to 1 if player C was a woman); SVO angle (the higher the SVO angle value, the higher the player C's concern for the others); IAT index. (see Table 3 and Table 4).

1.3.1 Results

In the first analysis, comparing subjects' punishing choices, we found that both the percentage of decisions to punish and the average level of punishment were higher when player A's transfers were classified as unfair by player C. This difference was observed across the three tDCS conditions (see Figure 3).

	SHAM	TPJ	VMPFC
Unfairness	0.185***	0.149***	0.287***
Age	-0.033	0.141	-0.014
Female	-0.433	-0.071	-1.171
SVO	0.024	-0.000	-0.01
IAT	-1.87**	1.125	-1.246
Costant	-0.502	-4.255	0.678
N	3200	3200	3200

Tab 1. Random – effect Probit Regression. Decision to punish as dependent variable (equal to 1 if C subtracts tokens to A, 0 otherwise). 20 Subjects for each tDCS condition (SHAM, TPJ, VMPFC). ***1% significance **5%significance *10% significance.

	SHAM	TPJ	VMPFC
Unfairness	0.732***	0.698***	1.129***
Age	-0.158	0.665	-0.005
Female	1.95	-0.707	-5.103
SVO	0.116	0.012	0.05
IAT	-7.547**	4.766	-3.77
Costant	-1.708	-19.757	1.047
Left-censored obs	2053	1639	1754
Uncensored obs	1045	1387	1267
Right-censored obs	102	174	179

Tab 2. Random – effect Tobit Regression. Amount of punishment as dependent variable (from 0 to 10 tokens). 20 Subjects for each t-DCS condition (SHAM-TPJ-VMPFC). ***1% significance **5%significance *10% significance.

In the second analysis, considering the decision to punish in subjective unfair trials, analyses showed a significant effect of: unfairness ($p < 0.001$), SVO angle ($p = 0.026$) and VMPFC ($p = 0.056$). This implies that when players C faced higher levels of subjective unfairness, they were more likely to punish. Moreover, higher player C's concern for the others (measured by means of SVO) increased the probability of sanctioning unfair players A. Finally, under the VMPFC condition, punishment of

unfair players A was more likely in VMPFC condition. That is, the stimulation of VMPFC seemed to positively affect altruistic punishment. In subjective fair trials, analyses showed a significant effect of fairness ($p < 0.001$) and TPJ ($p = 0.037$) (see Table 3). This means that the higher Player A's transfer with respect to Player C's fairness reference point, the lower the probability to punish. Unlike, tDCS of TPJ positively affects the probability to punish and therefore the presence of antisocial punishment.

Considering the amount of punishment in unfair trials, analyses showed significant effect of: unfairness ($p < 0.001$), FRONT ($p = 0.056$) and SVO angle ($p= 0.024$). This means that a higher level of subjective unfairness, player C's concern for the others and stimulation of VMPFC increased the intensity of the sanction when player A behaved unfairly. In fair trails, analyses showed a significant effect of the fairness ($p < 0.001$) and TPJ ($p = 0.029$) (see Table 4). That is, when player A was more than fair, stimulation of TPJ increased the level of antisocial punishment.

	Fair trials	**Unfair Trials**
Level of unfairness		0.156***
Level of fairness	-0.181***	
TPJ	1.355**	0.514
VMPFC	0.677	0.844*
Age	-0.102	0.087
Female	-0.263	-0.22
SVO	-0.009	0.029**
IAT	-1.37	-0.674
Ingroup	-0.068	-0.022
Costant	1.532	-3.201*

Tab 3. Random – effect Probit Regression for Fair and Unfair trials. Decision to punish as dependent variable (equal to 1 if C subtracts tokens to A, 0 otherwise). ***1% significance **5%significance *10% significance.

	Fair trials	**Unfair Trials**
Level of unfairness		0.77***
Level of fairness	-0.411***	
TPJ	4.926**	1.671
VMPFC	2.56	3.038*
Age	-0.38	0.202
Female	-1.293	-1.266
SVO	-0.032	0.106**
IAT	-4.647	-2.433
Ingroup	-0.154	-0.021

Costant	5.297	-9.555
Left-censored obs	2831	2615
Uncensored obs	938	2761
Right-cesored obs	107	348

Tab 4. Random – effect Tobit Regression for Fair and Unfair trials. Amount of punishment as dependent variable (from 0 to 10 tokens). ***1% significance **5%significance *10% significance.

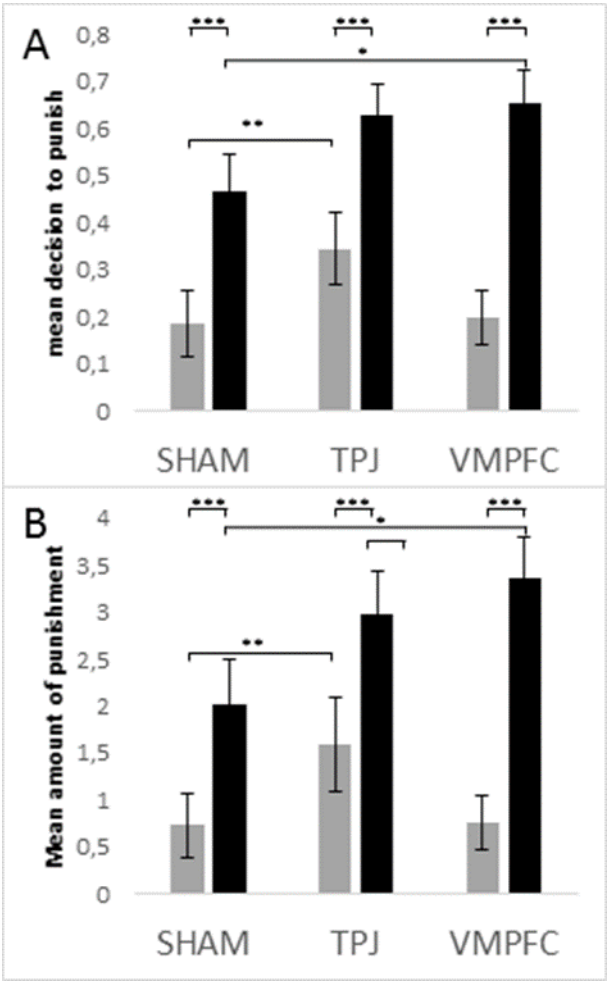

Figure 3. **A**, random-effect probit with decision to punish as dependent variable. **B**, random-effect tobit regressions with amount of punishment as dependentvariable. Bars represent the standard error. Gray and Black indicate Fair and Unfair trials

1.3.2 Results of Psychological traits

The D score of IAT was calculated as the difference of reaction times between congruent and incongruent trials. This index evaluate implicit race bias towards European and Asian faces. In the present study, participants showed a mean IAT-D score of 0.46 (sd = 0.29). A one-way ANOVA between groups (sham, TPJ, VMPFC) did not show significant differences in IAT-D score p = 0.296.

Results showed that, on average, participants associated pleasant words to own racial group, showing a moderate preference toward their group (Greenwald et al 1998, 2003).

The SVO mean score of the sample was 26.39 (sd = 13.67). A one-way ANOVA between groups (sham, TPJ, VMPFC) did not show a significant difference in SVO score, p = .828 revealing that participants were characterized by prosociality with no differences among the three assigned tDCS conditions.

The IRI mean score of the sample was 67.66 (sd = 14.09). The one-way ANOVA between groups (sham, TPJ, VMPFC) did not show a significant difference in IRI score p = .781.

1.4.1 Discussion

The current study is the first attempt to investigate the casual role of VMPFC and TPJ on altruistic and antisocial punishment, by combining TPP game with anodal tDCS. These two sites were chosen because they are parts of reward system and mentalizing systems respectively.

Furthermore, it addressed whether in-group vs out-group nationality membership modulates punishment behaviors of third party.

At the behavioral level, results showed the feasibility of applying a TPP game to trigger both altruistic punishment and antisocial punishment. Indeed, firstly unfair trials increased decisions to punish and amounts of punishment of players C, compared with fair trials. This finding is in line with previous literature, which showed that third parties tended to spend own resources to punish unfair behaviors even when they were not directly involved in the unfair economic exchanges (Ciaramidaro et al., 2018; Fehr & Fischbacher, 2004; Jensen et al., 2010; Morese et al., 2016; Rabellino et al., 2016). Secondly, the TPP game, even if less frequently, enhanced players C's antisocial punishment in fair trials. Also this result converges with the evidence, described by Rabellino et al. (2016) and Morese

et al. (2016), that most of participants spent small amounts of money to punish fair behaviors during a TPP game.

Regarding the modulatory role of VMPFC and TPJ in punishing behaviors, our data revealed for the first time that reward and mentalizing networks differently modulate altruistic and antisocial punishment. Indeed, while anodal tDCS over VMPFC shows a trend increased altruistic punishment, anodal tDCS over TPJ increased anti-social punishment choices. Particularly, the result that anodal tDCS over right TPJ significantly triggered, during fair trials, an increase of punishment behaviors, is a novelty for the research on neural correlates of punishment behaviors. However, recent neuroimaging studies demonstrated a critical role of TPJ and mentalizing system in TPP (Baumgartner et al 2012; Bellucci et al 2017; Zinchenko & Klucharev, 2017). The ability to attribute mental states to ourselves and others is an important aspect in social cognition. This ability is often referred to as "mentalizing", "mindreading" or "theory of mind" (Frith & Frith, 2006; Saxe, 2006) and plays a crucial role in altruistic decision making because it allows to understand the mental (affective) states of others, their beliefs and intentions. Neuroimaging and lesion studies showed that the recruitment of the bilateral TPJ is fundamental in people's metalizing ability (Samson et al., 2004; Schaafsma et al., 2015; Schurz et al., 2014). Indeed, a recent fMRI study revealed an increased activity of TPJ while third parties observed victim receiving help (Hu et al., 2016). The involvement of the TPJ could be explained not only by the process of mentalization but also by the attentional charge that is required when making a specific choice against the norm (David et al., 2017). In addition, other neuroimaging studies showed an increased activation of bilateral TPJ during competitive economic games (Halko et al., 2009; Votinov et al., 2015). These data support the possible competitive nature of antisocial punishment. This behavior indeed might be a perfidious way of reducing other's pay off in order to obtain the higher pay off (Fliessbach et al., 2007). Barclay (2013) also proposed that antisocial punishment could be a means to discredit competitors, preventing them from cooperating (Barclay, 2013, 2016; Pleasant & Barclay, 2018). Indeed, according to the

biological markets theory (Noë & Hammerstein, 1994, 1995), cooperation helps the development of a reputation that makes the cooperator more likely to be chosen for a beneficial cooperative partnership (Sylwester & Roberts, 2010). Pleasant & Barclay (2018) in a recent study demonstrated that antisocial punishment was used during a public good game as a means to be chosen for a following cooperative task (trust game). Authors suggested that antisocial punishment was used to avoid looking bad when cooperation was needed. We speculated that the enhancement of the neuronal activity of TPJ could increase the capability to infer Player A's mental state. Consequently, Players C could punish first part's altruistic behavior because they could interpret it as a way of player A to develop his reputation for a possible next competition.

Furthemore, analyses showed that anodal stimulation of VMPFC, compared with sham tDCS, increased both the decision to punish and the amount of punishment in unfair trials, confirming the role of VMPFC in mediating the altruistic punishment. These findings, even if at limits of significance, are in line with a recent tDCS study reporting enhanced altruistic behaviors in a Dictator game after anodal stimulation of VMPFC, while none significant effect followed cathodal stimulation, compared with sham (Zheng et al., 2016). The association between activity of VMPFC and cooperative behaviors such as altruism, emerged in our study, is further confirmed by clinical lesions studies (Krajbich et al., 2009; Moretto et al., 2013), showing that patients with demage to the VMPFC divided less equally their endowment when acting as dictators in a Dictator game. Our results support also recent fMRI data (Mathur et al., 2010; Morese et al. 2016; Waytz et al., 2012). Particularly, Morese and her colleagues demonstrated increased activation of VMPFC when subjects punished the unfair condition in a TPP paradigm. However, in this study the VMPFC was also involved in punishing player A fair transfers, suggesting that the VMPFC might have a key role in both altruistic and antisocial punishment. Regarding TPJ, Morese et al. (2016) showed that this region mediated punishing unfair trials acted by in-group members (i.e. when player A was chinese in TPP

paradigm), that is the parochial altruism. Our study extends these results, showing that stimulating VMPFC and TPJ differently affected punishment behaviors, with anodal tDCS over VMPFC increasing altruistic punishment and anodal tDCS over TPJ icreasing antisocial punishment. We can speculate that even though the mentalizing network, which includes both the VMPFC and TPJ, is crucially involved when social interactions require punishing (Baumgartner et al., 2012; Morese et al., 2016), as those occuring during TPP, these two regions have specific and discernable roles.

However, it is possible that the increased altruism following anodal stimulation of VMPFC observed in our study, rather than being associated to the specific role of VMPFC in altruistic punishment, is due to tDCS affecting adjacent structures, such as DLPFC. Indeed, this region, as part of central executive network, has been demonstrated to be crucial in TPP induced behaviors (Knoch et al., 2009; Krueger & Hoffman, 2016; Nihonsugi et al., 2015; Zinchenko & Klucharev, 2017). Interestingly, a tDCS study by Zheng et al. (2016) allows us to rule out this possibility. Indeed, stimulating the VMPFC and the DLPFC in two different experiments, authors showed that only tDCS over the VMPFC, and not over the DLPFC, significantly increased cooperative choices in a Dictator game. These results support our hypothesis, providing evidence that altruism specifically depends on VMPFC activity. By contrast, the DLPFC might more likely mediate economic exchanges when self-interested motives are involved, as demonstrated by previous neurostimulation studies (Knoch et al., 2006, 2008).

Moreover, our data revealed that high scores in SVO significantly increase the probability to punish and the amount of punishment during unfair trials. This result is in line with prior research (De Cremer & Van Lange, 2001; Stouten et al., 2005; Van Dijk et al., 2004; Van Lange, 1999) which showed that Prosocials (people with high scores in SVO) endorsed a true norm of fairness and therefore they should reject unfair proposes or punish unfair behavior more likely than Proselfs (people with low scores in SVO).

Our results, differently by previous studies (Baumgartner et al 2012; Morese et al., 2016; Rabellino et al., 2016), did not report substantial impact of group membership in punishment behavior. This result could be related to a different distribution of initial amount of the players. In Morese et al. (2016) Player A had more tokens than Palyer C and this not equal distribution could emphasize the group membership and the competition between groups. In contrast, in our study Player A and Player C had the same amount of tokens and this could mitigate the group membership. However, further studies are needed to corroborate this hypothesis and confirm the results with other social groups of different nationality.

To conclude, our findings highlight that tDCS differently modulates the VMPFC and the TPJ activity when people observe unfair and fair economic interactions, suggesting that different brain networks, namely the reward and mentalizing systems, underlie altruistic and anti-social punishment behaviours. However, further studies are needed to corroborate our results. Mainly, a better understanding of the nature of altruistic and antisocial punishment could drive to more exhaustive conclusions on whether and how these brain networks work in synergy when social interactions stimulate punishing behaviours.

Authors Contributions: E.L.G, L.J.R.L, S.O, F.P, R.M, A.G, F.B, conceived and designed the study. E.L.G, A.G, G.L run the experiments. S.O, conducted the statistical analysis, A.G, E.L.G, A.V, S.O, R.M prepared the draft and E.L.G, A.G, A.V jointly produced the final draft.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ambrus, G. G., Al-Moyed, H., Chaieb, L., Sarp, L., Antal, A., & Paulus, W. (2012). The fade-in–short stimulation–fade out approach to sham tDCS–reliable at 1 mA for naive and experienced subjects, but not investigators. Brain stimulation, 5(4), 499-504.

Bahry, D. L., & Wilson, R. K. (2006). Confusion or fairness in the field? Rejections in the ultimatum game under the strategy method. Journal of Economic Behavior & Organization, 60(1), 37-54.

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans.

Evolution and Human Behavior, 34(3), 164-175.

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. Current Opinion in Psychology, 7, 33-38.

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.

Baumgartner, T., Schiller, B., Rieskamp, J., Gianotti, L. R., & Knoch, D. (2013). Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. Social cognitive and affective neuroscience, 9(5), 653-660.

Bellucci, G., Chernyak, S., Hoffman, M., Deshpande, G., Dal Monte, O., Knutson, K. M. & Krueger, F. (2017). Effective connectivity of brain regions underlying third-party punishment: functional MRI and Granger causality evidence. Social neuroscience, 12(2), 124-134.

Bendor, J., & Swistak, P. (2001). The evolution of norms. American Journal of Sociology, 106(6), 1493-1545.

Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. Nature, 442(7105), 912.

Bieleke, M., Gollwitzer, P. M., Oettingen, G., & Fischbacher, U. (2017). Social value orientation moderates the effects of intuition versus reflection on responses to unfair ultimatum offers. Journal of Behavioral Decision Making, 30(2), 569-581.

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science, 328(5978), 617-620.

Bosco, F. M., Parola, A., Valentini, M. C., & Morese, R. (2017). Neural correlates underlying the comprehension of deceitful and ironic communicative intentions. Cortex, 94, 73-86.

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.

Camerer, C., & Fehr, E. (2003). The roundtable series in behavioral economics. Russel Sage Foundations.

Chib, V. S., Yun, K., Takahashi, H., & Shimojo, S. (2013). Noninvasive remote activation of the ventral midbrain by transcranial direct current stimulation of prefrontal cortex. Translational Psychiatry, 3(6), e268.

Ciaramidaro, A., Toppi, J., Casper, C., Freitag, C. M., Siniatchkin, M., & Astolfi, L. (2018). Multiple-Brain Connectivity During Third Party Punishment: an EEG Hyperscanning Study. Scientific reports, 8(1), 6822.

De Cremer, D., & Van Lange, P. A. (2001). Why prosocials exhibit greater cooperation than proselfs: The roles of social responsibility and reciprocity. European Journal of personality, 15(S1), S5-S18.

Civai, C., Miniussi, C., & Rumiati, R. I. (2014). Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. Social cognitive and affective neuroscience, 10(8), 1054-1060.

De Dreu, C. K., Greer, L. L., Handgraaf, M. J., Shalvi, S., Van Kleef, G. A., Baas, M., ... & Feith,

S. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. Science, 328(5984), 1408-1411.

David, B., Hu, Y., Krüger, F., & Weber, B. (2017). Other-regarding attention focus modulates third-party altruistic choice: an fMRI study. Scientific reports, 7, 43024.

Davis, M. H. (1980). Interpersonal reactivity index. Edwin Mellen Press.

De Quervain, D. J., Fischbacher, U., Treyer, V., & Schellhammer, M. (2004). The neural basis of altruistic punishment. Science, 305(5688), 1254.

Elster, J. (1989). Social norms and economic theory. Journal of economic perspectives, 3(4), 99- 117.

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. American Economic Review, 90(4), 980-994.

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. Human nature, 13(1), 1-25.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. Nature, 425(6960), 785.

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. Evolution and human behavior, 25(2), 63-87.

Fehr, E., & Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. Current opinion in neurobiology, 14(6), 784-790.

Feng, C., Deshpande, G., Liu, C., Gu, R., Luo, Y. J., & Krueger, F. (2016). Diffusion of responsibility attenuates altruistic punishment: a functional magnetic resonance imaging effective connectivity study. Human brain mapping, 37(2), 663-677.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C. E., & Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. science, 318(5854), 1305-1308.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. Neuron, 50(4), 531-534.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods, 39(2), 175-191

Gibbons, R. D., & Hedeker, D. (1994). Application of random-effects probit regression models.

Journal of consulting and clinical psychology, 62(2), 285.

Giustolisi, B., Vergallito, A., Cecchetto, C., Varoli, E., & Lauro, L. J. R. (2018). Anodal transcranial direct current stimulation over left inferior frontal gyrus enhances sentence comprehension. Brain and language, 176, 36-41.

Goette, L., Huffman, D., Meier, S., & Sutter, M. (2012). Competition between organizational groups: Its impact on altruistic and antisocial motivations. Management science, 58(5), 948-960.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. Journal of personality and social psychology, 74(6), 1464.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. Journal of personality and social psychology, 85(2), 197.

Güroğlu, B., van den Bos, W., van Dijk, E., Rombouts, S. A., & Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. Neuroimage, 57(2), 634-641.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. Journal of economic behavior & organization, 3(4), 367-388.

Halevy, N., Weisel, O., & Bornstein, G. (2012). “In- group love” and “out- group hate” in repeated interaction between groups. Journal of Behavioral Decision Making, 25(2), 188-195.

Halko, M. L., Hlushchuk, Y., Hari, R., & Schürmann, M. (2009). Competing with peers: Mentalizing-related brain activity reflects what is at stake. Neuroimage, 46(2), 542-548.

Hämmerer, D., Bonaiuto, J., Klein-Flügge, M., Bikson, M., & Bestmann, S. (2016). Selective alteration of human value decisions with medial frontal tDCS is predicted by changes in attractor dynamics. Scientific reports, 6, 25160.

Harlé, K. M., Chang, L. J., van't Wout, M., & Sanfey, A. G. (2012). The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula. Neuroimage, 61(1), 32-40.

Helbing, D., Szolnoki, A., Perc, M., & Szabó, G. (2010). Evolutionary establishment of moral and double moral standards through spatial interactions. PLoS computational biology, 6(4), e1000758.

Hennig-Schmidt, H., Li, Z. Y., & Yang, C. (2008). Why people reject advantageous offers—Non-monotonic strategies in ultimatum bargaining: Evaluating a video experiment run in PR China. Journal of Economic Behavior & Organization, 65(2), 373-384.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., ... & Henrich, N. S. (2005). "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behavioral and brain sciences, 28(6), 795-815.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., ... & Lesorogol, C. (2006). Costly punishment across human societies. Science, 312(5781), 1767-1770.

Herrmann, B., Thöni, C., & Gächter, S. (2008). Antisocial punishment across societies. Science, 319(5868), 1362-1367.

Hoffman, M. B. (2014). The Punisher's Brain: The Evolution of Judge and Jury. Cambridge University Press.

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.

Hu, Y., Scheele, D., Becker, B., Voos, G., David, B., Hurlemann, R., & Weber, B. (2016). The effect of oxytocin on third-party altruistic decisions in unfair situations: An fMRI study. Scientific reports, 6, 20236.

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365(1553), 2635-2650.

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. Nature, 530(7591), 473.

Keel, J. C., Smith, M. J., & Wassermann, E. M. (2001). A safety screening questionnaire for transcranial magnetic stimulation. Clinical neurophysiology, 112(4), 720.

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(43), 15284-15293.

Kirman, A., & Teschl, M. (2010). Selfish or selfless? The role of empathy in economics. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365(1538), 303-317.

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.

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2007). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. Cerebral Cortex, 18(9), 1987-1990.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. Proceedings of the National Academy of Sciences, pnas-0911619106.

Krajbich, I., Adolphs, R., Tranel, D., Denburg, N. L., & Camerer, C. F. (2009). Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. Journal of Neuroscience, 29(7), 2188-2192.

Krueger, F., & Hoffman, M. (2016). The emerging neuroscience of third-party punishment. Trends in neurosciences, 39(8), 499-501.

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. Personality and Social Psychology Bulletin, 31(4), 443-453.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Baron-Cohen, S., & MRC AIMS Consortium. (2011). Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. Neuroimage, 56(3), 1832-1838.

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. Neuroimage, 51(4), 1468-1475.

Mattavelli, G., Gallucci, A., Schiena, G., D'Agostino, A., Sassetti, T., Bonora, S., ... & Sassaroli, S. (2019). Transcranial direct current stimulation modulates implicit attitudes towards food in eating disorders. International Journal of Eating Disorders.

McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. The review of economics and statistics, 318-321.

Morese, R., Rabellino, D., Sambataro, F., Perussia, F., Valentini, M. C., Bara, B. G., & Bosco, F. M. (2016). Group membership modulates the neural circuitry underlying third party punishment. PloS one, 11(11), e0166357.

Moretto, G., Sellitto, M., & di Pellegrino, G. (2013). Investment and repayment in a trust game after ventromedial prefrontal damage. Frontiers in human neuroscience, 7, 593.

Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. Neuron, 75(1), 73-79.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation.

Nihonsugi, T., Ihara, A., & Haruno, M. (2015). Selective increase of intention-based economic decisions by noninvasive brain stimulation to the dorsolateral prefrontal cortex. Journal of Neuroscience, 35(8), 3412-3419.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., ... & Pascual-Leone, A. (2008). Transcranial direct current stimulation: state of the art 2008. Brain stimulation, 1(3), 206-223.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves?. Journal of Public Economics, 92(1-2), 91-112.

Noë, R., & Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. Behavioral ecology and sociobiology, 35(1), 1-11.

Noë, R., & Hammerstein, P. (1995). Biological markets. Trends in Ecology & Evolution, 10(8), 336-339.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory.

Neuropsychologia, 9(1), 97-113.

Ostrom, E. (2000). Collective action and the evolution of social norms. Journal of economic perspectives, 14(3), 137-158.

Ottone, S., Ponzano, F., & Zarri, L. (2015). Power to the People? An experimental analysis of bottom-up accountability of third-party institutions. The Journal of Law, Economics, & Organization, 31(2), 347-382.

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.

Pleasant, A., & Barclay, P. (2018). Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. Psychological science, 0956797617752642.

Polanía, R., Moisa, M., Opitz, A., Grueschow, M., & Ruff, C. C. (2015). The precision of value-based choices depends causally on fronto-parietal phase coupling. Nature communications, 6, 8090.

Rabellino, D., Morese, R., Ciaramidaro, A., Bara, B. G., & Bosco, F. M. (2016). Third-party punishment: altruistic and anti-social behaviours in in-group and out-group settings. Journal of Cognitive Psychology, 28(4), 486-495.

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees.

Proceedings of the national academy of sciences, 109(37), 14824-14829.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clinical neurophysiology, 120(12), 2008- 2039.

Sääksvuori, L., Mappes, T., & Puurtinen, M. (2011). Costly punishment prevails in intergroup conflict. Proceedings of the Royal Society of London B: Biological Sciences, rspb20110252.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. Nature neuroscience, 7(5), 499.

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.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. Current Biology, 22(23), 2274-2277.

Saxe, R. (2006). Uniquely human social cognition. Current opinion in neurobiology, 16(2), 235- 239.

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.

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. Child development, 80(4), 1197-1209.

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. Trends in cognitive sciences, 19(2), 65-72.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neuroscience & Biobehavioral Reviews, 42, 9-34.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., & Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proceedings of the National Academy of Sciences, 108(19), 7710-7715.

Stouten, J., De Cremer, D., & Van Dijk, E. (2005). All is well that ends well, at least for proselfs: Emotional reactions to equality violation as a function of social value orientation. European Journal of Social Psychology, 35(6), 767-783.

Strobel, A., Zimmermann, J., Schmitz, A., Reuter, M., Lis, S., Windmann, S., & Kirsch, P. (2011). Beyond revenge: neural and genetic bases of altruistic punishment. Neuroimage, 54(1), 671-680.

Sylwester, K., & Roberts, G. (2010). Cooperators benefit through reputation-based partner choice in economic games. Biology letters, rsbl20100209.

Tajfel, H., & Turner, J. C. (1986). The Social Identity Theory of Intergroup Behavior, Psychology of Intergroup Relations, edited by Stephen Worchel and William G. Austin. Chicago: Nelson-Hall, 724.

Van Dijk, E., De Cremer, D., & Handgraaf, M. J. (2004). Social value orientations and the strategic use of fairness in ultimatum bargaining. Journal of experimental social psychology, 40(6), 697-707.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. Journal of personality and social psychology, 77(2), 337.

Votinov, M., Pripfl, J., Windischberger, C., Sailer, U., & Lamm, C. (2015). Better you lose than I do: neural networks involved in winning and losing in a real time strictly competitive game. Scientific reports, 5, 11017.

Waytz, A., Zaki, J., & Mitchell, J. P. (2012). Response of dorsomedial prefrontal cortex predicts altruistic behavior. Journal of Neuroscience, 32(22), 7646-7650.

Woods, A. J., Antal, A., Bikson, M., Boggio, P. S., Brunoni, A. R., Celnik, P., ... & Knotkova, H. (2016). A technical guide to tDCS, and related non-invasive brain stimulation tools. Clinical Neurophysiology, 127(2), 1031-1048.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. Neuropsychologia, 48(9), 2658-2664.

Zheng, H., Huang, D., Chen, S., Wang, S., Guo, W., Luo, J., ... & Chen, Y. (2016). Modulating the activity of ventromedial prefrontal cortex by anodal tDCS enhances the trustee's repayment through altruism. Frontiers in psychology, 7, 1437.

Zhong, S., Chark, R., Hsu, M., & Chew, S. H. (2016). Computational substrates of social norm enforcement by unaffected third parties. NeuroImage, 129, 95-104.

Zinchenko, O., & Klucharev, V. (2017). Commentary: The Emerging Neuroscience of Third-Party Punishment. Frontiers in human neuroscience, 11, 512.

Zinchenko, O., & Arsalidou, M. (2018). Brain responses to social norms: Meta- analyses of f MRI studies. Human brain mapping, 39(2), 955-970.