NeuroImage 200 (2019) 501–510

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

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

Emanuele Lo Gerfo [a,b,h,i,*,1], Alessia Gallucci [c,h,1], Rosalba Morese [e], Alessandra Vergallito [c,h], Stefania Ottone [a,h,i], Ferruccio Ponzano [g,i], Gaia Locatelli [c], Francesca Bosco [d,f], Leonor Josefina Romero Lauro [c,h,i]

[a] *Department of Economic, Management and Statistics, University of Milano Bicocca, Italy*
[b] *Clinical Psychology Service of Mediterranean Institute for Transplantation and Advanced Specialized Therapies (IRCSS IsMeTT), Italy*
[c] *Department of Psychology, University of Milano Bicocca, Italy*
[d] *Department of Psychology, University of Turin, Turin, Italy*
[e] *Faculty of Communication Sciences, University della Svizzera italiana, Lugano, Switzerland*
[f] *Neuroscience Institute of Turin, University of Turin, Italy*
[g] *University of Piemonte Orientale, Department of Law and Political, Economic and Social Sciences, Italy*
[h] *NeuroMi - Milan Center for Neuroscience, Italy*
[i] *CISEPS, University of Milano Bicocca, Italy*

ARTICLE INFO

Keywords:
tDCS
VMPFC
TPJ
Altruistic punishment
Antisocial punishment

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 brain areas belonging to the reward system (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 right TPJ (rTPJ), (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 they would punish the 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 rTPJ 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.

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 and Rockenbach, 2004; Helbing et al., 2010). Indeed, respecting spoken or unspoken shared social rules promotes cooperation and leads human social behavior (Bendor and 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

[*] Corresponding author. Economic, Management and Statistics, University of Milano Bicocca, Piazza dell’Ateneo Nuovo 1, Milan, Italy.
E-mail address: emanuele.logerfo@unimib.it (E. Lo Gerfo).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.neuroimage.2019.06.047
Received 12 December 2018; Received in revised form 13 May 2019; Accepted 19 June 2019
Available online 21 June 2019

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 behaviors are altruistic and antisocial punishment. Both of them represent a costly form of punishment, since the outcome 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 and Gachter, 2000; Fehr and 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 and 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 and 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 and Wilson, 2006; Henrich et al., 2006; Hennig-Schmidt et al., 2008) showed that antisocial punishment could also be considered as a form of do-gooder derogation. In these studies, people reject fair and hyperfair proposes. According to the authors, people might be suspicious of others who appear too generous.

Economic games are useful to explore humans' punishment behaviors, finding one of their main application as behavioral tasks in neuroimaging studies interested in shedding light on the neural substrates of punishment behaviors (Boyd et al., 2010; Güth et al., 1982; Strobel et al., 2011). In particular, the TPP game has been effectively used in a broader range of research. In a typical TPP game an impartial bystander (the third party, player C) can decide, spending part of his endowment, to punish a player (the dictator, player A) who allocates fair or unfair amount of money to a dummy player (the receiver, player B) (Fehr and Fischbacher, 2004; Ottone et al., 2015).

1.1. Neural correlates of sanctioning behavior

A growing body 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). Specifically, the implicated networks include: the *salience network*, which detects 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 and 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, involving the medial prefrontal cortex; the *mentalizing network*, involved in the inference of mental states or durable characteristics, comprising the dorsomedial prefrontal cortex and TPJ (Feng et al., 2016; Güroğlu et al., 2011; Krueger and Hoffman, 2016; Bosco et al., 2017); the *central-executive network*, which transforms signals coming from the default mode network into punishment behaviors, relying upon the posterior parietal cortex and the dorsolateral prefrontal cortex (DLPFC) (Krueger and Hoffman, 2016; Zinchenko and Arsalidou, 2018); the *reward network*, involving the nucleus accumbens and VMPFC (De Quervain et al., 2004; Hu et al., 2015). All the above mentioned systems seem to play a general role in both norms' representation and violation processing (Zinchenko and Arsalidou, 2018).

Concerning membership influence in economic games, differences in the neural responses were reported when the group status was manipulated. 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 increased activity of the 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.

Among the neural networks involved, two brain regions seem to be crucial hubs of the punishing behavior, although with different roles: TPJ and VMPFC.

Concerning TPJ role, some authors speculated an antagonistic relationship between this region and the DLPFC during TPP (Krueger and Hoffman, 2016). Indeed, the prefrontal cortex 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. More specifically, the right portion of TPJ was generally found to have a high specialization for mentalizing (Saxe and Powell, 2006; Saxe et al., 2009; Young et al., 2010), a processing type that is crucial to interact in a social environment. A study by Lombardo et al. (2011) showed that, compared with healthy controls, rTPJ responses of autism spectrum patients were similar for both mentalizing and physical judgments, with anomalous rTPJ activations of patients correlating with the degree of their social impairment. Focusing on sanctioning behavior, several neuroimaging studies support the specific role of rTPJ 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 the activity of rTPJ and the subjective value of sanctioning (Zhong et al., 2016) as well as a causal relationship between rTPJ activity and parochial punishment (Baumgartner et al., 2013) in TPP.

Regarding VMPFC, a recent study unveiled a specific role of this brain region in the antisocial punishment, (Morese et al., 2016). Hence, together with the rTPJ, VMPFC seems to be involved in 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).

Despite in the last years increasingly attempts to explore the neural correlates of punishing behaviors have been made, research in this field is still at the beginning and more evidence is required to explain the causal relationship between a certain brain network and its behavioral counterpart.

With the present study, we aimed to fill this gap by investigating, through the application of the tDCS, the causal role of the reward and mentalizing networks, targeting respectively VMPFC and rTPJ, in the altruistic and the antisocial punishment behaviors.

TDCS has already been shown to be effective 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). However, to the best of our knowledge, no previous study investigated tDCS effects on punishment behaviors, in particular antisocial punishment mechanisms, in the context of TPP.

Building on previous neuroimaging evidence, and particularly on results of a recent study by Morese et al. (2016), we expect that anodal tDCS would modulate altruistic punishment when applied to both VMPFC and rTPJ, whereas only tDCS over VMPFC would modulate antisocial punishment. Moreover, we expected tDCS effects to interact with membership manipulation. In this sense, we expected participants to punish more frequently an outgroup member as dictator making unfair offers to an ingroup member as receiver in the sham condition and that stimulation over rTPJ 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 and 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) in order to measure individual prosociality, empathy and prejudice and their impact on punishment behaviors.

2. Material and methods

2.1. Participants

Sixty healthy Italian students participated to the study (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 written informed consent prior to study procedures. The study 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.

2.2. Experimental design and procedures

The experimental procedure was divided in two different sessions.

2.2.1. 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 couples were made, i.e. Caucasian/negative stimuli and Asian/positive stimuli (incongruent trials).

In order to control the impact of prosociality, we used the SVO slider measure which allows to measure prosociality vs individualism on the basis of participants' choices during a series of dictator games with another person. In particular, in order to have a continuous variable, we computed the SVO angle. An angle of around zero indicates proself persons, who maximize their own outcomes without considering outcomes of others; positive values correspond to more prosocial behavior, indicative of people who gain positive utility from outcomes of others.

After completing IAT and SVO, participants filled the IRI, a 28-items questionnaire that investigates empathy skills using a five-points Likert scale from 1 (it does not describe my behavior) to 5 (it totally describes my behavior). IRI items can be divided into four subscales which investigate perspective-taking, fantasy, empathic concern personal distress.

2.2.2. Session two: tDCS procedure and third party punishment game

The second session took place after a week.First, tDCS stimulation was delivered, followed by the execution of the TPP game.

2.2.3. tDCS procedures

Participants were randomly assigned to one out of three experimental conditions (20 participants for each group): 1) anodal tDCS over rTPJ (10 males and 10 females); 2) anodal tDCS over VMPFC (9 males and 11 females); 3) sham condition (6 males, 14 females): in this condition, half of participants received the placebo stimulation over the rTPJ and the other half over VMPFC, a common procedure used when more brain regions are targeted in the same experiment (e.g. Mattavelli et al., 2019; Vergallito et al., 2018). The sample size was estimated using G*Power3 (Faul et al., 2007). A sample size of 19 participants per group (for a total of 57 for the entire study) was indicated as necessary to detect an effect size of 0.20 with 90% power and an alfa of 0.05.

We performed a double-blind study design, therefore both participants and experimenters were blinded about the experimental conditions.

TDCS was delivered through a Brain Stim stimulator (Newronica, Milan, Italy). Electrodes' position was established through the EEG 10–20 International System. Specifically, for the stimulation of VMPFC, the center of the anode was positioned in a middle point between Fp1 and Fp2 (Chib et al., 2013), while for the rTPJ it was centered on Cp6 (Santiesteban et al., 2012). In both conditions the cathode was vertically positioned, with its center over O1 (see Fig. 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 × 5 cm anode and a 10 × 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 min. In the sham tDCS, instead, the stimulator turned off automatically after 30 s, 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). In order to standardize the procedure, participants watched a cartoon video during tDCS delivering (Giustolisi et al., 2018).

2.2.4. Third party punishment game (TPP)

Immediately after the stimulation, participants were involved in a TPP game (Fehr and Fischbacher, 2004) comprising 160 trials (Fig. 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 (see Fig. 2).

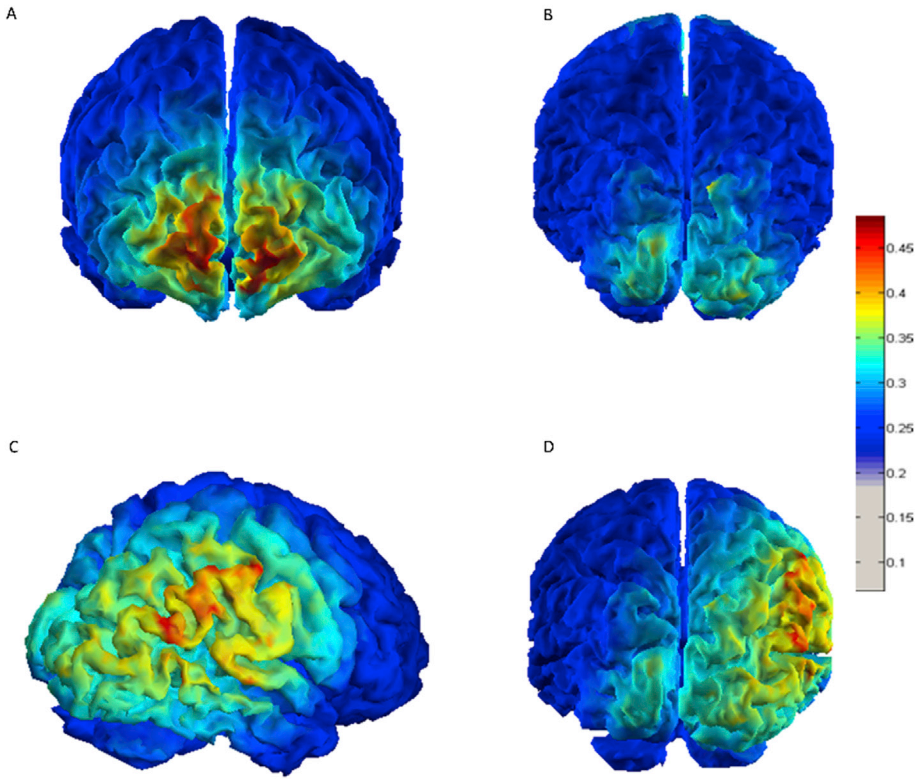

Fig. 1. Computational model of tDCS-induced electric field. A simulation of the electrical field induced by the tDCS protocol used in the study was computed using Comets (Jung et al., 2013). Following EEG international 10–20 system, in the anodal VMPFC condition (upper figure A & B), the anode (25 cm^2) was placed in the middle point between Fp1 and Fp2 (panel A) and the cathode (50 cm^2) 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 color indicates the strongest electrical field over the VMPFC and rTPJ.

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 and 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 trials 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 to ideally transfer to player B. Then subjects were debriefed about tDCS and TPP game real procedures and experimental aims.

Participants received 2 € for the participation plus an additional payment based on the amount of money earned during the SVO and the TTP game (mean = 12.5 €, SD = 2.7).

2.3. Statistical analysis

TPP trials were classified as fair and unfair based on participants' subjective fairness level, as assessed by the questionnaire administered at the end of the second session (mean of subjective fairness was equal to 7.13 tokens, SD = ± 2.9 median and mode were equal to 5).

Specifically, we classified 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. For each participant the decision to punish (equal to 1 when player C subtracted tokens to player A, 0 otherwise) and the amount of punishment (which values were censored between 0 and 10) were analyzed as dependent variables.

We applied two typical regression models: random-effect probit and a random-effect tobit regressions (McDonald and Moffitt, 1980; Gibbons and Hedeker, 1994) for analyzing the decision to punish and the amount of punishment, respectively. Considering the high correlation between SVO and IRI scales ($p < .001$), we only introduced SVO angle as predictor in our regression models.

2.4. Preliminary results

Gender. To control for gender distribution across the three conditions (rTPJ, VMPFC, sham) we run a Chi-square, indicating no significant difference (X (3) = 2.16, p = .33).

Individual differences. A series of one-way ANOVA showed no significant differences among the three conditions (rTPJ, VMPFC, sham) for age [F (2, 60) = 0.36 p = .70], IAT-D score [F (2, 60) = 1.25 p = .30], IRI total score [F (2, 60) = 0.26 p = .77] and SVO angle [F (2, 60) = 0.21 p = .81]. These results indicate that the randomization was successful. Differences among the three tDCS conditions (VMPFC, rTPJ, sham) were analyzed also for the level of fairness. A one-way ANOVA among the three groups (rTPJ, VMPFC, sham) did not show differences in fairness reference point [F (2, 60) = 1.49 p = .233)], showing that the level of fairness was not influenced by the stimulation.

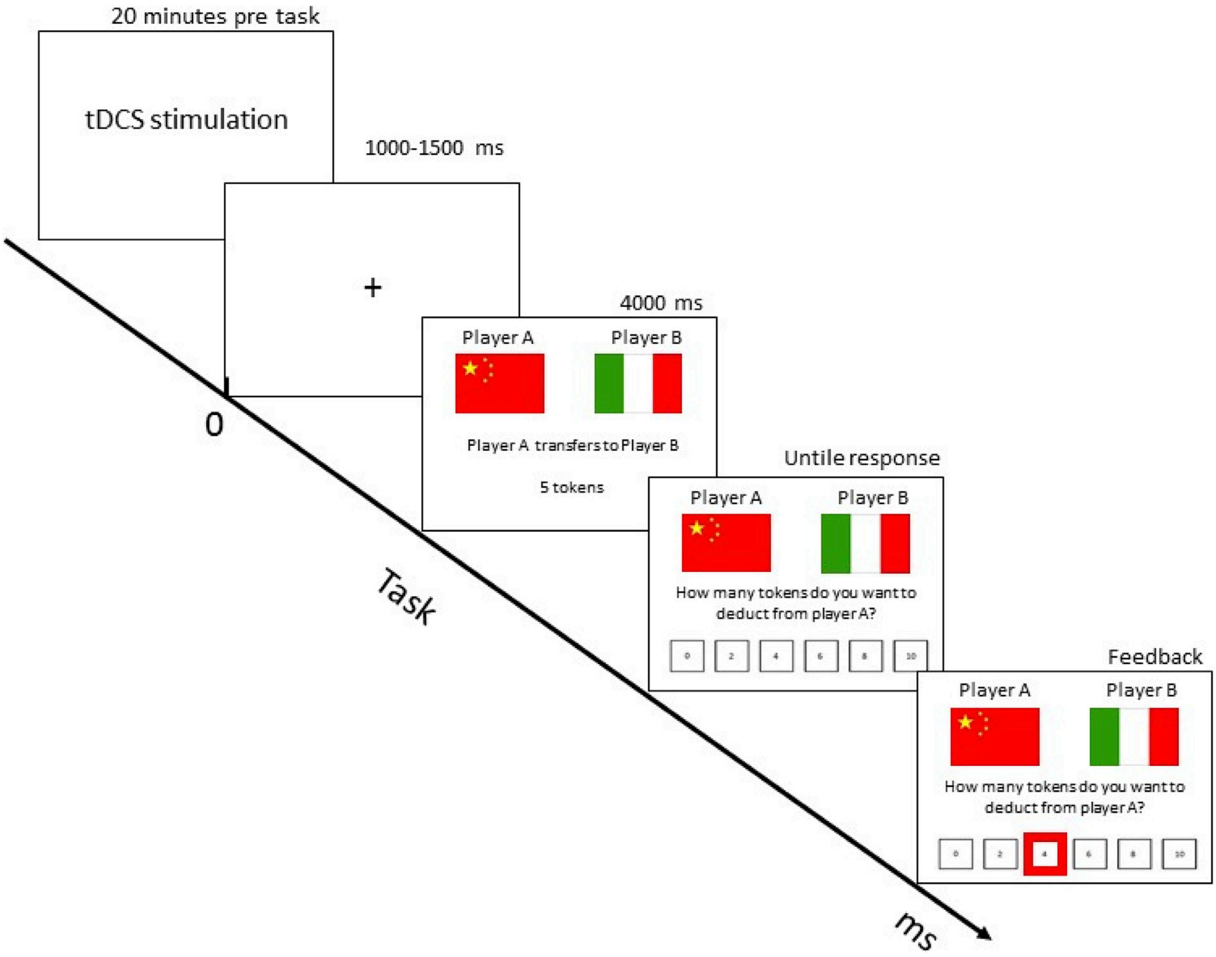

Fig. 2. Schematic experimental procedure. TPP started after 20 min of sham or real tDCS. Each trial started with a fixation cross followed by a first screen where player A could decide to transfer part of his endowment (from 0 to 5 tokens) to player B. In a second screen, player C, observing player A's 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.

3. Results

In a first analysis, we considered the three stimulation conditions separately. We entered in both probit and tobit regressions (in which the dependent variables were the *decision to punish* and the *amount of punishment,* respectively) the following predictors: fairness (a variable computed as the difference between Player C's fairness reference point and the actual transfer); age; gender (a dummy variable equal to 1 if player C was a woman, 0 otherwise; the SVO angle (the higher the SVO angle value, the higher the player C's concern for the others); the IAT index.

In both regressions, the variable fairness had a positive and significant effect (see Table 1 with the regression results over the dichotomous variable *decision to punish* and Table 2 for the regression over the dependent variable *amount of punishment*): participants punish more frequently and intensely the more player A's transfers were considered unfair. This difference was observed across the three tDCS conditions.

From our first analysis, another significant result emerged: higher IAT values lead to a lower probability to punish (p = .043) and to a lower level of punishment (p = .039), but only in the sham condition.

In order to study this point more in details, we divided fair and unfair trials and checked whether personality traits interacted with stimulation conditions on the decision to punish and how much. More specifically, we run six random-effect probit and six random-effect tobit regressions,

Table 1
The table shows the random – effect probit regression values over the dichotomous variable "decision to punish" (1 = punish behavior, 0 = no punish behavior). ***1% significance **5%significance *10% significance.

	SHAM	TPJ	VMPFC
Fairness	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

Table 2
The table shows the random – effect tobit regression values over the censored dependent variable "amount of punishment" (from 0 to 10 tokens). ***1% significance **5%significance *10% significance.

	SHAM	TPJ	VMPFC
Fairness	0.732***	0.698***	1.129***
Age	−0.158	0.665	−0.005
Gender	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

namely a regression for each tDCS condition (anodal VMPFC, anodal rTPJ, sham) for fair and unfair transfers.

The dependent variables were again the decision to punish and the level of punishment. Regressors in both fair and unfair analysis were fairness/unfairness, age, gender, SVO angle, IAT index and the interaction between the level of fairness/unfairness and the IAT index. Crucially, from here we refer to fairness for fair trial's regression and unfairness for unfair trial's regression.

Fairness was computed as the difference between the actual transfer and C's fairness reference point, whereas unfairness was computed considering the difference between C's fairness reference point and the actual transfer. It turns out is that the IAT index interacted with the level of fairness/unfairness, affecting punishment decision both in the sham tDCS and in the VMPFC tDCS conditions. In particular, in both sham and VMPFC conditions, higher IAT D score increased punishment as the level of fairness/unfairness increased (under fair and unfair trials respectively).

These results were further investigated in a second analysis, which aimed to isolate and study *antisocial* (within fair trials) and *altruistic* (within unfair trials) punishment.

We run a series of random-effect probit and tobit regressions with the following predictors: unfairness/fairness; membership (a dummy variable equal to 1 if both player A and player B were Italian, 0 otherwise); VMPFC and rTPJ (two dummy variables for tDCS-VMPFC and tDCS-TPJ condition respectively; the control condition was sham tDCS); age; gender; SVO angle; IAT index; the interaction between VMPFC and IAT index; the interaction between fairness/unfairness and the IAT, the interaction among fairness, IAT and VMPFC.

For the *fair trials* (see the left column of Table 3), analysis on the dichotomous variable decision to punish showed a significant effect of fairness (p < .001), indicating that participants tended to punish less frequently when Player A fairness was higher. We also found a main effect of rTPJ stimulation (p = .031), showing that participants punished more frequently after this region stimulation as compared to sham (Fig. 3). IAT index was also significant (p < .05): the higher was the IAT, the less participants decided to punish.

We found a significant interaction between fairness and IAT index (p < .001): participants decided to punish more frequently when IAT and fairness increased.

Results on the censored variable amount on punishment (see the left column of Table 4) were on the same direction as the decision to punish. Also in this case we found a main effect of fairness, showing that participants punished subtracting less tokens for higher levels of fairness (p < .001) and when the IAT index was higher (p = .04). rTPJ stimulation increased the amount of tokens subtracted to Player A (p = .024) as

Table 3
The table shows the random – effect probit regression values separately for fair and unfair trials over the dichotomous variable “decision to punish” (1 = punish behavior, 0 = no punish behavior). ***1% significance **5%significance *10% trend toward significance.

	Fair trials	Unfair Trials
Level of unfairness		0.128***
Level of fairness	−0.344***	
rTPJ	1.416**	0.506
VMPFC	0.07	1.554
Age	−0.104	0.091
Gender	−0.256	−0.195
SVO angle	−0.007	0.028**
IAT index	−2.372**	−0.369
VMPFC*IAT	1.358	−1.982
Unfair * IAT		0.024
Unfair*IAT*VMPFC		0.138***
Fairness*IAT	0.333***	
Fairness*IAT*VMPFC	−0.029	
Membership	−0.069	−0.004
Costant	1.936	−3.335*

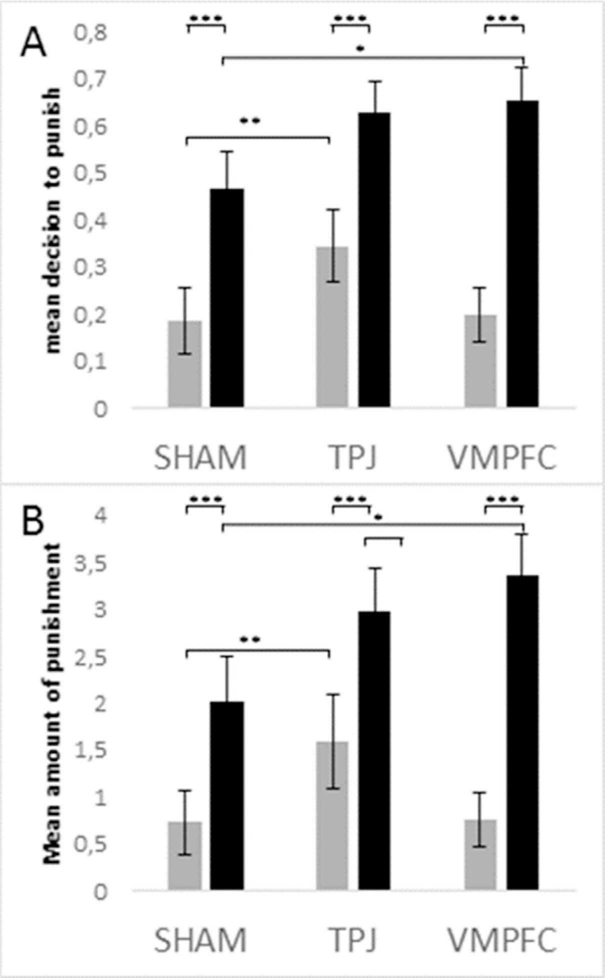

Fig. 3. A, random-effect probit with decision to punish as dependent variable. **B**, random-effect tobit regressions with amount of punishment as dependent variable. Bars represent the standard error. Gray and Black indicate Fair and Unfair trials.

Table 4
The table shows the random – effect tobit regression values separately for fair and unfair trials over the censored dependent variable “amount of punishment” (from 0 to 10 tokens). ***1% significance **5%significance *10% trend toward significance.

	Fair trials	Unfair Trials
Level of unfairness		0.695***
Level of fairness	−0.748***	
rTPJ	5.12**	1.662
VMPFC	0.183	4.512
Age	−0.382	0.216
Gender	−1.422	−1.267
SVO angle	−0.026	0.103**
IAT index	−7.72**	−1.779
VMPFC*IAT	5.6	−4.882
Ufair*IAT		0.036
Unfair*IAT*VMPFC		0.435***
Fairness*IAT	0.857***	
Fairness*IAT*VMPFC	−0.319*	
Membership	−0.146	−0.019
Costant	6.394	−9.767
Left-censored obs	2831	2615
Uncensored obs	938	2761
Right-cesored obs	107	348

compared to the sham condition (Fig. 3).

Also in this case the interaction between fairness and IAT index was significant (p < .001), highlighting more intense amount of punishment for higher levels of IAT and fairness.

Analysis of *unfair trials* on the dichotomous variable decision to punish (see the right column of Table 3) highlighted a main effect of unfairness (p < .001), suggesting that participants punished more frequently when Player A transfers were higher. We also found a main effect of SVO (p = .029): participants with higher SVO angle, which indicates higher concern for others, punished more frequently Player A's unfair transfers. Moreover, we found an interaction among unfairness, IAT index and VMPFC stimulation (p < .001) (Fig. 3), suggesting that the probability of punishing increased with increasing the level of unfairness and IAT scores.

Also in the case of unfair trials, the intensity of punishment was in the same direction as the decision to punish (see the right column of Table 3): the amount of tokens subtracted to Player A was higher when he/she behaved more unfairly (p < .001), when the SVO angle was higher (p = .031). Moreover, the three-way interaction among unfairness, IAT index and VMPFC stimulation was significant (p < .001) (Fig. 3), suggesting that the intensity of the punishment increased with the level of unfairness and IAT index.

4. Discussion

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

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

At the behavioral level, results showed the feasibility of applying a TPP game to trigger both altruistic punishment and antisocial punishment. Indeed, unfair trials increased decisions to punish and the amounts of punishment decided by players C, compared with fair trials. This finding is in line with previous literature, which showed that third parties tend to spend own resources to punish unfair behaviors even when they are not directly involved in the unfair economic exchanges (Ciaramidaro et al., 2018; Fehr and Fischbacher, 2004; Jensen et al., 2010; Morese et al., 2016; Rabellino et al., 2016).

Secondly, our data show that during the game antisocial punishment, namely the decision to punish fair transfers, was traceable even if less frequent than altruistic punishment. This result converges with the evidence, described by Rabellino et al. (2016) and Morese et al. (2016), that participants spend small amounts of money to punish fair behaviors.

Regarding the effect of tDCS applied to VMPFC and rTPJ in punishing behaviors, our data reveal for the first time that reward and mentalizing networks differently modulate altruistic and antisocial punishment. Indeed, while anodal tDCS over VMPFC interacted with IAT index in increasing altruistic punishment, anodal tDCS over rTPJ increased antisocial punishment choices. This latter result represents a novelty for the research on neural correlates of punishment: indeed, anodal tDCS over rTPJ significantly triggered, during fair trials, an increase of punishment behaviors. Recent neuroimaging studies demonstrated a critical role of rTPJ and mentalizing system in TPP (Baumgartner et al., 2012; Bellucci et al., 2017; Zinchenko and Klucharev, 2017). The ability to attribute mental states to others and ourselves is an important aspect in social cognition. This ability is often referred to as “mentalizing”, “mindreading” or “theory of mind” (Frith and Frith, 2006; Saxe, 2006) and plays a crucial role in altruistic decision making because it allows understanding 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 left 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 and Barclay, 2018). Indeed, according to the biological markets theory (Noë and 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 and Roberts, 2010). Pleasant and 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 rTPJ could increase the capability to infer Player A's mental state. Consequently, Players C could punish first party's altruistic behavior because they could interpret it as a way of player A to develop his reputation for a possible next competition.

As previously mentioned, analyses showed that anodal stimulation of VMPFC interacted with IAT index, increasing both the decision to punish and the amount of punishment in unfair trials, confirming the role of this region in mediating the altruistic punishment. Our findings 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 damage 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 Morese study the VMPFC was also activated during the punishment of player A fair transfers, suggesting that the VMPFC might have a key role in both altruistic and antisocial punishment. The interaction with the IAT score was an unexpected result. Higher D scores at IAT corresponded to stronger implicit coupling of positive/in-group and negative/out-group. This result is even more surprisingly since we did not find a membership effect on the probability to punish and its intensity. We don't have a clear explanation for this result. We can speculate that increasing activity of VMPFC, which is part of the reward system, induced participants with higher implicit in-group membership to sanctioning more unfair trials. Interestingly, indeed, in our study no significant effect of membership emerged, suggesting that in-group membership exerts and effects only at an implicit but not at an explicit level. Neuroimaging and stimulation studies converging in highlighting the involvement of the mPFC in implicit measurement. For example, effects on IAT were traceable for food evaluation (Mattavelli et al., 2015), self and others discrimination (Mitchell et al., 2006), prejudice detection (Sellaro et al., 2015). This idea, however, should be better investigated by future research.

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 rTPJ differently affected punishment behaviors, with anodal tDCS over VMPFC increasing altruistic punishment and anodal tDCS over rTPJ increasing antisocial punishment. We can speculate that the mentalizing and reward networks, which includes TPJ e VMPFC respectively, are crucially involved when social interactions require punishing (Baumgartner et al., 2012; Morese et al., 2016), such as in the case of TPP, and that these two regions have specific and discernible 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 and Hoffman, 2016; Nihonsugi et al., 2015; Zinchenko and 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, 2007).

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 and Van Lange, 2001; Stouten et al., 2005; Van Dijk et al., 2004; Van Lange, 1999) which showed that prosocials (people with higher SVO angle) endorsed a true norm of fairness and therefore they should reject unfair proposes or punish unfair behavior more likely than Proselfs (people with lower SVO angle).

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 due 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 over VMPFC and the rTPJ differently modulates the choice to punish 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 behaviors. 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 behaviors.

5. Limitation of the present study

The present study holds several limitations. First of all, despite the sample size was based on a preliminary power analysis, the results need to be corroborated by additional and independent data collection, such to enlarge the sample.

Another limitation regards the well-known low spatial resolution of tDCS. Although in the present study the computational model of current flow showed a confined electrical field within the target sites, previous studies suggest caution on the spatial resolution of tDCS usually because the electric fields may not be limited underneath the stimulating electrodes rather influencing other regions (Moretto et al., 2013; Romero Lauro et al., 2014, 2016 but see also Pisoni et al., 2017). Thus, we cannot rule out the possibility that the stimulation affected also other areas, for instance within the prefrontal cortex, such as the DLPFC. Further research with more focal techniques (e.g., TMS) could specify our results.

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, L.J. R. L A.G, A.V jointly produced the final draft.

Funding

Funding for this research was provided by University of Piemonte Orientale, Italy; Department of Psychology, University Milano Bicocca, Italy; Department of Economic, Management and Statistics, University of Milano Bicocca, Italy; PRIN 2015 PER-FORMARE IL SOCIALE. Formazione, cura e inclusione sociale attraverso il teatro, PI Prof. Bernardi.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.06.047.

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 Stimul. 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. J. Econ. Behav. Organ. 60 (1), 37–54.
Barclay, P., 2013. Strategies for cooperation in biological markets, especially for humans. Evol. Hum. Behav. 34 (3), 164–175.
Barclay, P., 2016. Biological markets and the effects of partner choice on cooperation and friendship. Curr. Opin. Psychol. 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. Hum. Brain Mapp. 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. Soc. Cognit. Affect Neurosci. 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. Soc. Neurosci. 12 (2), 124–134.
Bendor, J., Swistak, P., 2001. The evolution of norms. Am. J. Sociol. 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. J. Behav. Decis. Mak. 30 (2), 569–581.
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.
Boyd, R., Gintis, H., Bowles, S., 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. Science 328 (5978), 617–620.
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. Transl. 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. Sci. Rep. 8 (1), 6822.
Civai, C., Miniussi, C., Rumiati, R.I., 2014. Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. Soc. Cognit. Affect Neurosci. 10 (8), 1054–1060.
David, B., Hu, Y., Krüger, F., Weber, B., 2017. Other-regarding attention focus modulates third- party altruistic choice: an fMRI study. Sci. Rep. 7, 43024.
Davis, M.H., 1980. Interpersonal Reactivity Index. Edwin Mellen Press.
De Cremer, D., Van Lange, P.A., 2001. Why prosocials exhibit greater cooperation than proselfs: the roles of social responsibility and reciprocity. Eur. J. Personal. 15 (S1), S5–S18.
De Dreu, C.K., Greer, L.L., Handgraaf, M.J., Shalvi, S., Van Kleef, G.A., Baas, M., Feith, S.W., et al., 2010. The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. Science 328 (5984), 1408–1411.
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. J. Econ. Perspect. 3 (4), 99–117.
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. Behav. Res. Methods 39 (2), 175–191.

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. Evol. Hum. Behav. 25 (2), 63–87.
Fehr, E., Gachter, S., 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90 (4), 980–994.
Fehr, E., Rockenbach, B., 2004. Human altruism: economic, neural, and evolutionary perspectives. Curr. Opin. Neurobiol. 14 (6), 784–790.
Fehr, E., Fischbacher, U., Gächter, S., 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. Hum. Nat. 13 (1), 1–25.
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. Hum. Brain Mapp. 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.
Gandiga, P.C., Hummel, F.C., Cohen, L.G., 2006. Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. Clin. Neurophysiol. 117 (4), 845–850.
Gibbons, R.D., Hedeker, D., 1994. Application of random-effects probit regression models. J. Consult. Clin. Psychol. 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 Lang. 176, 36–41.
Goette, L., Huffman, D., Meier, S., Sutter, M., 2012. Competition between organizational groups: its impact on altruistic and antisocial motivations. Manag. Sci. 58 (5), 948–960.
Greenwald, A.G., McGhee, D.E., Schwartz, J.L., 1998. Measuring individual differences in implicit cognition: the implicit association test. J. Personal. Soc. Psychol. 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. J. Personal. Soc. Psychol. 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. J. Econ. Behav. Organ. 3 (4), 367–388.
Halevy, N., Weisel, O., Bornstein, G., 2012. "In- group love" and "out- group hate" in repeated interaction between groups. J. Behav. Decis. Mak. 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. Sci. Rep. 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 Comput. Biol. 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. J. Econ. Behav. Organ. 65 (2), 373–384.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., , et al.Henrich, N.S., 2005. "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. Behav. Brain Sci. 28 (6), 795–815.
Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Lesorogol, C., et al., 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. Front. Behav. Neurosci. 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. Sci. Rep. 6, 20236.
Jensen, K., 2010. Punishment and spite, the dark side of cooperation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 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.
Jung, Y.J., Kim, J.H., Im, C.H., 2013. COMETS: a MATLAB toolbox for simulating local electric fields generated by transcranial direct current stimulation (tDCS). Biomed. Eng. Lett. 3 (1), 39–46.
Keel, J.C., Smith, M.J., Wassermann, E.M., 2001. A safety screening questionnaire for transcranial magnetic stimulation. Clin. Neurophysiol. 112 (4), 720.
Keeser, D., Meindl, T., Bor, J., Palm, U., Pogarell, O., Mulert, C., Padberg, F., et al., 2011. Prefrontal transcranial direct current stimulation changes connectivity of resting-state networks during fMRI. J. Neurosci. 31 (43), 15284–15293.
Kirman, A., Teschl, M., 2010. Selfish or selfless? The role of empathy in economics. Philos. Trans. R. Soc. Lond. B Biol. Sci. 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. Cerebr. 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. Proc. Natl. Acad. Sci. U.S.A., 0911619106 pnas.
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. J. Neurosci. 29 (7), 2188–2192.
Krueger, F., Hoffman, M., 2016. The emerging neuroscience of third-party punishment. Trends Neurosci. 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. Pers. Soc. Psychol. Bull. 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., Zuglian, P., Dabroi, E., Gaslini, G., Clerici, M., Papagno, C., 2015. Transcranial magnetic stimulation of medial prefrontal cortex modulates implicit attitudes towards food. Appetite 89, 70–76.
Mattavelli, G., Gallucci, A., Schiena, G., D'Agostino, A., Sassetti, T., Bonora, S., Sassaroli, S., et al., 2019. Transcranial direct current stimulation modulates implicit attitudes towards food in eating disorders. Int. J. Eat. Disord. 52 (5), 576–581.
McDonald, J.F., Moffitt, R.A., 1980. The uses of Tobit analysis. Rev. Econ. Stat. 318–321.
Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50 (4), 655–663.
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. Front. Hum. Neurosci. 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. J. Neurosci. 35 (8), 3412–3419.
Nikiforakis, N., 2008. Punishment and counter-punishment in public good games: can we really govern ourselves? J. Public Econ. 92 (1–2), 91–112.
Nitsche, M.A., Cohen, L.G., Wassermann, E.M., Priori, A., Lang, N., Antal, A., Pascual-Leone, A., et al., 2008. Transcranial direct current stimulation: state of the art 2008. Brain Stimul. 1 (3), 206–223.
Noë, R., Hammerstein, P., 1994. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. Behav. Ecol. Sociobiol. 35 (1), 1–11.
Noë, R., Hammerstein, P., 1995. Biological markets. Trends Ecol. Evol. 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. J. Econ. Perspect. 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. J. Law Econ. Organ. 31 (2), 347–382.
Peña-Gómez, C., Sala-Lonch, R., Junqué, C., Clemente, I.C., Vidal, D., Bargalló, N., Bartrés- Faz, D., et al., 2012. Modulation of large-scale brain networks by transcranial direct current stimulation evidenced by resting-state functional MRI. Brain Stimul. 5 (3), 252–263.
Pisoni, A., Mattavelli, G., Papagno, C., Rosanova, M., Casali, A.G., Romero Lauro, L.J., 2017. Cognitive enhancement induced by anodal tDCS drives circuit-specific cortical plasticity. Cerebr. Cortex 28 (4), 1132–1140.
Pleasant, A., Barclay, P., 2018. Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. Psychol. Sci. 29 (6), 868–876.
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. Nat. Commun. 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. J. Cogn. Psychol. 28 (4), 486–495.
Riedl, K., Jensen, K., Call, J., Tomasello, M., 2012. No third-party punishment in chimpanzees. Proc. Natl. Acad. Sci. U.S.A. 109 (37), 14824–14829.
Lauro, L.J.R., Rosanova, M., Mattavelli, G., Convento, S., Pisoni, A., Opitz, A., Vallar, G., 2014. TDCS increases cortical excitability: direct evidence from TMS–EEG. Cortex 58, 99–111.
Lauro, L.J.R., Pisoni, A., Rosanova, M., Casarotto, S., Mattavelli, G., Bolognini, N., Vallar, G., 2016. Localizing the effects of anodal tDCS at the level of cortical sources: A Reply to Bailey et al., 2015. Cortex 100 (74), 323–328.
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. Clin. Neurophysiol. 120 (12), 2008–2039.

Sääksvuori, L., Mappes, T., Puurtinen, M., 2011. Costly punishment prevails in intergroup conflict. Proc. Biol. Sci. 278 (1723), 3428–3436.

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W., 2004. Left temporoparietal junction is necessary for representing someone else's belief. Nat. Neurosci. 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. Curr. Biol. 22 (23), 2274–2277.

Saxe, R., 2006. Uniquely human social cognition. Curr. Opin. Neurobiol. 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. Psychol. Sci. 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 Dev. 80 (4), 1197–1209.

Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R., 2015. Deconstructing and reconstructing theory of mind. Trends Cognit. Sci. 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. Neurosci. Biobehav. Rev. 42, 9–34.

Sellaro, R., Derks, B., Nitsche, M.A., Hommel, B., van den Wildenberg, W.P., van Dam, K., Colzato, L.S., 2015. Reducing prejudice through brain stimulation. Brain Stimul. 8 (5), 891–897.

Stanley, D.A., Sokol-Hessner, P., Banaji, M.R., Phelps, E.A., 2011. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proc. Natl. Acad. Sci. U.S.A. 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. Eur. J. Soc. Psychol. 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. Biol. Lett., rsbl20100209

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

Van Dijk, E., De Cremer, D., Handgraaf, M.J., 2004. Social value orientations and the strategic use of fairness in ultimatum bargaining. J. Exp. Soc. Psychol. 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. J. Personal. Soc. Psychol. 77 (2), 337.

Vergallito, A., Lauro, L.J.R., Bonandrini, R., Zapparoli, L., Danelli, L., Berlingeri, M., 2018. What is difficult for you can be easy for me. Effects of increasing individual task demand on prefrontal lateralization: a tDCS study. Neuropsychologia 109, 283–294.

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. Sci. Rep. 5, 11017.

Waytz, A., Zaki, J., Mitchell, J.P., 2012. Response of dorsomedial prefrontal cortex predicts altruistic behavior. J. Neurosci. 32 (22), 7646–7650.

Woods, A.J., Antal, A., Bikson, M., Boggio, P.S., Brunoni, A.R., Celnik, P., Knotkova, H., et al., 2016. A technical guide to tDCS, and related non-invasive brain stimulation tools. Clin. Neurophysiol. 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., et al., 2016. Modulating the activity of ventromedial prefrontal cortex by anodal tDCS enhances the trustee's repayment through altruism. Front. Psychol. 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., Arsalidou, M., 2018. Brain responses to social norms: meta- analyses of f MRI studies. Hum. Brain Mapp. 39 (2), 955–970.

Zinchenko, O., Klucharev, V., 2017. Commentary: the emerging neuroscience of third-party punishment. Front. Hum. Neurosci. 11, 512.