European Journal of Neuroscience

FENS Federation of European Neuroscience Societies

European Journal of Neuroscience, pp. 1–8, 2017 doi:10.1111/ejn.13653

Lateral prefrontal/orbitofrontal cortex has different roles in norm compliance in gain and loss domains: a transcranial direct current stimulation study

Yunlu Yin,[1,*] Hongbo Yu,[1,2,*] Zhongbin Su,[1] Yuan Zhang[1,3] and Xiaolin Zhou[1,4,5,6]

[1]School of Psychological and Cognitive Sciences and Center for Brain and Cognitive Sciences, Peking University, Beijing 100871, China
[2]Department of Experimental Psychology, University of Oxford, Oxford, UK
[3]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA
[4]Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China
[5]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China
[6]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

Keywords: gain-loss domain, HD-tDCS, lateral prefrontal and orbitofrontal cortex (LPFC/LOFC), norm compliance, sanction threat

Abstract

Sanction is used by almost all known human societies to enforce fairness norm in resource distribution. Previous studies have consistently shown that the lateral prefrontal cortex (lPFC) and the adjacent orbitofrontal cortex (lOFC) play a causal role in mediating the effect of sanction threat on norm compliance. However, most of these studies were conducted in gain domain in which resources are distributed. Little is known about the mechanisms underlying norm compliance in loss domain in which individual sacrifices are needed. Here we employed a modified version of dictator game (DG) and high-definition transcranial direct current stimulation (HD-tDCS) to investigate to what extent lPFC/lOFC is involved in norm compliance (with and without sanction threat) in both gain- and loss-sharing contexts. Participants allocated a fixed total amount of monetary gain or loss between themselves and an anonymous partner in multiple rounds of the game. A computer program randomly decided whether a given round involved sanction threat for the participants. Results showed that disruption of the right lPFC/lOFC by tDCS increased the voluntary norm compliance in the gain domain, but not in the loss domain; tDCS on lPFC/lOFC had no effect on compliance under sanction threat in either the gain or loss domain. Our findings reveal a context-dependent nature of norm compliance and differential roles of lPFC/lOFC in norm compliance in gain and loss domains.

Introduction

Fairness is the cornerstone of social and political justice across human societies and throughout history (Rawls, 1958; Reeve, 1998). It is about the nature of a socially just allocation of resources in a group or society. Aristotle summarized the principle of fairness as “something *equal* should be to those who are *equal*” (Aristotle, cf. Reeve, 1998). As fairness norm is inevitably in conflict with the selfish interest of certain parties in social exchanges, some sorts of sanction are adopted to enforce the fairness norm (Sober & Wilson, 1998; Fehr & Gachter, 2002; Henrich *et al.*, 2006; Montague & Lohrenz, 2007). In other words, whether conforming to fairness norm can be a cost–benefit trade-off or strategic decision in which an individual takes into account the benefit of violating the fairness norm and the cost of being punished (Güth & Damme, 1998).

Mounting evidence has shown that introducing sanction threat for the sake of fairness norm increases norm compliance behavior in resource distribution (Fehr & Gachter, 2002; Spitzer *et al.*, 2007) and that norm compliance under sanction threat has distinct psychological and neural basis compared with voluntary norm compliance, i.e., compliance without sanction threat (Ruff *et al.*, 2013). For example, using a resource allocation task (i.e., the Dictator Game

Correspondence: Dr Xiaolin Zhou, as above.
E-mail: xz104@pku.edu.cn

*These authors contributed equally.

Received 23 November 2016, revised 8 July 2017, accepted 11 July 2017

Edited by Chandan Vaidya

Reviewed by Adam Green, Georgetown University, USA; and Christian Ruff, University of Zurich, Switzerland

The associated peer review process communications can be found in the online version of this article.

with and without sanction) and functional MRI, Spitzer *et al.* (2007) showed that when asked to allocate a certain amount of monetary reward between themselves and an anonymous co-player, the participants allocated more to the co-player (i.e., closer to fairness norm) when a monetary sanction threat was introduced. In parallel with the increased allocation, neural activations in the lateral prefrontal cortex (lPFC) and lateral orbitofrontal cortex (lOFC) were enhanced relative to the sanction-free condition, indicating the involvement of these frontal structures in mediating norm compliance under sanction threat in the gain context. More recently, Ruff *et al.* (2013) combined such a behavioral task with transcranial direct current stimulation (tDCS) and demonstrated that modulations of the right lPFC function had opposite effects on voluntary norm compliance and norm compliance under sanction threat (see also Strang *et al.*, 2015).

However, these previous studies only focused on norm compliance in gain domain (hereafter, the term 'domain,', 'frame', and 'context' are used interchangeably), leaving aside the situation in which individuals have to share a certain amount of loss or harm. In fact, the latter situations are prevalent in human society, such as sharing compensation for a traffic accident between perpetrators and insurance company. Are people more or less willing to conform to fairness norm in sharing loss than in sharing gain? Is the neural mechanism underlying norm compliance in loss domain the same as that in gain domain? Although to our knowledge no research has directly compared the psychological and neural mechanisms underlying norm compliance in gain and the loss domains, a number of studies, including ours, have consistently shown that people are more willing to enforce fairness norm by costly punishment in loss than in gain domain (Buchan *et al.*, 2005; Leliveld *et al.*, 2009; Zhou & Wu, 2011; Guo *et al.*, 2013; Wu *et al.*, 2014). This finding is in line with the general principle in behavioral economics, i.e., potential losses have a greater impact on people's emotion and decision-making processes than equivalent gains (Tversky & Kahneman, 1981, 1991; Tom *et al.*, 2007; Rutledge *et al.*, 2016). Relatedly, Buhl (1999) showed that in inter-group interaction, in-group favoritism is less severe in tasks involving allocating negative resources than in tasks involving allocating positive resources. Taken together, norm compliance is highly context-dependent and the psychological and neural basis underlying norm compliance may vary across contexts, such as the distinction between gain and loss frames and the presence and absence of sanction threat. In this study, we combined a resource distribution task (i.e., Dictator Game, DG) and tDCS to investigate the context-dependent nature of norm compliance. We attempted to answer three interrelated questions. First, does the loss context in general increase norm compliance compared to the gain context, when the lPFC/lOFC function is intact (i.e., in the sham group)? Second, does gain/loss context affect voluntary norm compliance and compliance under sanction threat to a different degree? Third, does the lPFC/lOFC play a causal role in mediating the influence of context on norm compliance? Answering these questions may help understand the context-dependent nature of norm compliance and its neural mechanisms.

Our rationale for focusing on the right lPFC/lOFC is twofold. On the one hand, a number of neuroimaging and neurophysiology studies have consistently implicated the lOFC in representing socially relevant information (e.g., Watson & Platt, 2012) that guides individual's behavior and decision-making in a socially appropriate manner (Rushworth *et al.*, 2007; Willis *et al.*, 2010). In other words, this area may inform individual about what is socially appropriate behavior in a given context by setting value to the potential behavioral repertoire (Ursu & Carter, 2005). This argument is in line with a number of findings about social conformity and social influence (Campbell-Meiklejohn *et al.*, 2010, 2012), which is conceptually related to norm compliance. On the other hand, the lPFC/lOFC is of specific interest to studies of the influence of sanction threat on norm compliance, where conflicting results have been reported (see Spitzer *et al.*, 2007; Li *et al.*, 2009). Our recent study combining fMRI and tDCS has provided a way to reconcile the contradicting evidence by incorporating the intention behind sanction threat (Zhang *et al.*, 2016), suggesting that the lOFC may integrate various social and non-social information to give rise to a compliance decision. For these reasons, we decided to further explore the context-dependent nature of the function of lPFC/lOFC in mediating norm compliance, hereby introducing gain–loss context.

Materials and methods

Participants

Power calculation was carried out to determine sample size. In a previous study that combined tDCS and a similar experimental design (Zhang *et al.*, 2016), we found that the effect size of tDCS on norm compliance was medium to large (i.e., Cohen's d ranged from 0.4 to 0.5). Based on this knowledge, we estimated the sample size required to obtain a similar effect using G*Power 3 (Faul *et al.*, 2007). The power calculation showed that at least 30 participants were required for each group to achieve the targeted effect size. Sixty-six right-handed participants were recruited. Seven participants were excluded from analysis (three of them always transferred nothing to the partner, and four of them did not believe the experimental setup according to a post-experiment manipulation check and interview), leaving 59 participants in the analysis (tDCS sham group: 17 females and 12 males, mean age 22.3 years, age range 18–25 years; tDCS cathodal group: 23 females and 7 males, mean age 21.5 years, age range 18–25 years). All participants had normal or corrected-to-normal vision; none of them reported a history of neurological or psychiatric disorders. They all gave informed written consent prior to the experiment in accordance with the Declaration of Helsinki. The study was approved by the Ethics Committee of School of Psychological and Cognitive Sciences, Peking University.

HD-tDCS

High-density stimulation was delivered by a multi-channel stimulation adapter (SoterixMedical, 4 × 1 – C3, New York) connected to a battery-driven stimulator (SoterixMedical, Model 1300-A, New York). Five Ag–AgCl-sintered ring electrodes were put into plastic casings which are filled with conductive gel, embedded in a standard EEG cap according to the international 10–20 system, and attached to the adaptor device. Each electrode had ~4 cm^2 contact with the skull. The electrodes were arranged on the skull in a 4 × 1 ring configuration based on previous literature (Edwards *et al.*, 2013; Villamar *et al.*, 2013a). The four return electrodes formed a square and were spaced ~7.5 cm radially around the central electrode according to previous HD-tDCS studies (Villamar *et al.*, 2013a,b). To deliver stimulation on the right lPFC/lOFC, we placed three return electrodes on the locations corresponding to F2, F8, Fp1, one return electrode at lower eyelid, and one central electrode on FP2 (Manuel *et al.*, 2014; Mondino *et al.*, 2015; Willis *et al.*, 2015; Zhang *et al.*, 2016). Current polarity on the target brain area depended on the central electrode. The current distribution under HD-tDCS has been partially validated by empirical data through a MRI-guided finite element model (Datta *et al.*, 2009; Edwards *et al.*, 2013), and recent studies showed that current density of HD-tDCS falls off with increasing cortical depth (Datta *et al.*, 2009). The current intensity was 2.0 mA which created ~0.5 mA/cm^2 peak current density at the central electrode, and ~0.125 mA/cm^2 peak current density at the return electrodes. Stimulation started 8 min before the task, and was delivered during the entire course of the task (~20 min) with an additional 30-s ramp-up at the beginning of stimulation and 30-s ramp-down at the end. The placement of electrodes was the same for the sham and the cathodal stimulation. However, for the sham stimulation, the initial 30 s ramp-up was immediately followed by the 30-s ramp-down, and there was no stimulation for the rest of the session (cf. Gandiga *et al.*, 2006; Douglas *et al.*, 2015). For both the cathodal and sham stimulation conditions, participants felt a little uncomfortable initially, but gradually the feelings associated with stimulation became negligible before the task started, according to our post-experiment interview.

Compared with the conventional bipolar tDCS, HD-tDCS has been shown to have better spatial focality and prolonged effect (Datta *et al.*, 2009; Caparelli-Daquer *et al.*, 2012; Kuo *et al.*, 2013; Shen *et al.*, 2016). Although HD-tDCS is associated with stronger scalp sensations than conventional tDCS, it has been shown to be safe and tolerable with applications of up to 2.0 mA for about 20 min (Minhas *et al.*, 2010; Borckardt *et al.*, 2012; Kuo *et al.*, 2013). It should be noted that the spatial resolution of tDCS is limited compared to transcranial magnetic stimulation (TMS), even in the mode of HD-tDCS (see Fig. 1C for the area estimated to be affected by the tDCS). However, in order to better compare the current findings with the findings from a few previous tDCS studies on similar topics (e.g., Knoch *et al.*, 2008; Ruff *et al.*, 2013; Zhang *et al.*, 2016), we use tDCS to manipulate the activity of the lPFC/lOFC.

Procedure

The experiment had a 2 (stimulation: sham vs. cathodal) by 2 (context: Gain vs. Loss) by 2 (threat: threat-on vs. threat-off) mixed factorial design with stimulation as between-subject factor whereas context and threat as within-subject factors. A modified repeated one-shot Dictator Game was employed (cf. Zhang *et al.*, 2016), in which the participants allocated either a profit or a loss of 20 Chinese *yuan* (about U.S. $ 3.5) between themselves and a randomly paired co-player randomly chosen from three confederates. In each round, before the participant made the allocation, the computer randomly decided to retain or waive the punishment threat (4 *yuan*). If the threat was retained and the amount allocated to the paired co-player was lower (higher) than what the latter had expected in the gain (loss) context, the participant would be penalized by 4 *yuan*, although no feedback was given concerning how much the co-player expected and whether the participants were in fact punished. Voluntary compliance was defined as the amount allocated when no threat was imposed, whereas compliance under sanction threat was defined as the amount allocated when sanction threat was retained. Moreover, threat-induced strategic compliance was defined as the difference in allocation amount between the threat retained and the threat waived conditions.

Upon arrival at the laboratory, the participant and three same-sex strangers (confederates of the experimenter) went through a randomization procedure (i.e., drawing lots) to determine their role in the game. We told the participant that one lot had a letter ‘A’ on it, while the other three had ‘B’. The one who drew the unique lot would be assigned the role of allocator, while the others would be assigned the role of receiver. Unbeknownst to the participant, all the four lots had an “A” on it to ensure that the participant be assigned the role of allocator. The participant believed that he/she would play each round through internet with a randomly paired receiver who was in another room. We told the participant that on each round the paired receiver would indicate the minimum share he/she expected

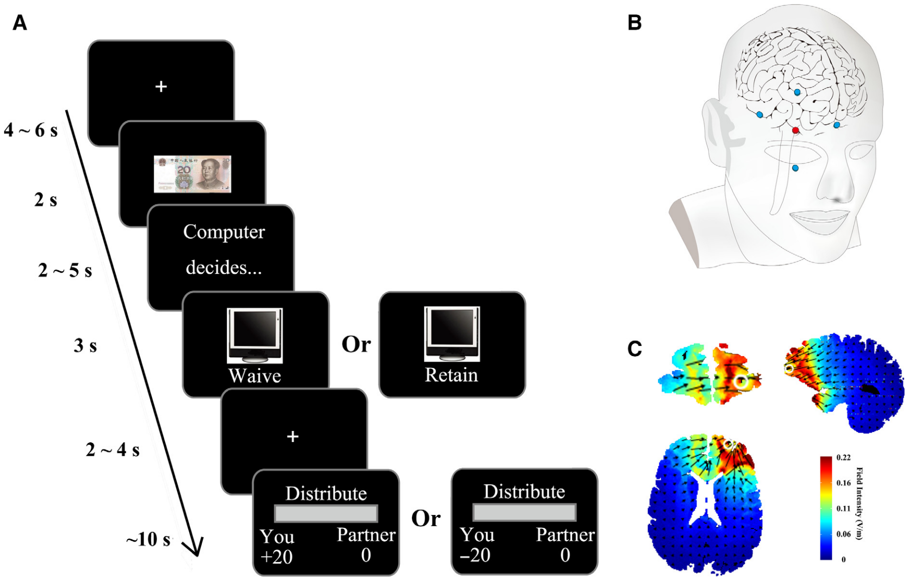

FIG. 1. (A) Procedure and task sequence. The participant allocated 20-*yuan* profit (about $ 3.5) or 20-*yuan* loss between him/herself and a randomly paired partner in each round. The computer randomly decided to retain or waive the punishment threat (4 *yuan*) before the participant made the allocation. (B) Schematic illustration of the HD-tDCS electrodes placement: Right OFC was localized at FP2 in the 10/20 EEG system (red circle). (C) Electric field simulation was performed with the HD-explorer software (SoterixMedical, New York, USA); simulated field intensity was indicated by the color bar. Arrow direction indicated current flow direction and arrow length indicated current flow intensity.

from the allocator. If the amount the allocator (i.e., the participant) allocated to the receiver was less than that minimum amount, a sanction may or may not be imposed on the allocator, depending on a prior decision by the computer (see below). To avoid learning effect, no feedback of earning/loss or sanction was provided. The participant was also told that a gain round and a loss round would be randomly chosen and realized after the experiment; this was to motivate the participant to treat each round equally and independently.

Each round began with the presentation of a white fixation cross against a black background, lasting for 4000 to 6000 ms with a step of 400 ms (Fig. 1). Then a cue of the total allocation amount (a picture of 20 *yuan* bill) was presented for 2000 ms, followed by a sentence indicating that punishment threat would be randomly decided by the computer for this trial. This sentence remained on the screen for 2000–5000 ms (with a step of 400 ms). Then the decision (Waive vs. Retain) together with a picture of computer were presented on the screen for 3000 ms. Specifically, 'Waive' means the computer decides that no sanction will be imposed on the current round, so the participant can allocate as she wishes without worrying about sanction. 'Retain' means the computer decides to keep the sanction threat on the current trial. In that case, if the participant's allocation was less than the minimum expectation given by the receiver, the participant would receive a sanction (although he/she did not know whether he/she was actually sanctioned in a given trial). Finally, after a 2000-to-4000-ms fixation, a distribution screen was presented. The participant was required to make the allocation within 10 s by pressing two buttons to adjust the allocation amount with a step of 2 *yuan* and a third button to confirm the allocation. The allocation was directed to the receiver so that in the gain context the positive points allocated to the receiver would be added to the receiver's account, while in the loss context, the negative points allocated to the receiver would be deducted from the partner's account. Button press was counterbalanced across participants. The initial amount on the side of the participant was either 0 or 20 *yuan* (0 or −20 *yuan* in the loss context) and was counterbalanced across conditions.

The allocation task consisted of a gain block and a loss block, each of which had 32 trials. Overall the task lasted about 20 min. Block sequence was counterbalanced across participants. A regression analysis showed that the sequence did not have any significant influence on participants' allocation decision. Therefore, in our data analysis we collapsed this factor. Each of the four experimental conditions (context × threat) has 16 trials. Presentation order of Waive and Retain conditions was pseudo-randomized and different sequences were created for different participants. To make sure that the participants actually believe our experimental setup, we included in the post-experiment questionnaire a number of questions assessing the participants' thoughts and attitudes about the experiment. These questions are 'To what extent you care about your payoff in the game' (1, not at all; 5, very much), 'To what extent you think you are interacting with a real human partner' (1, not at all; 5, very much), 'Do you have any questions, comments, and concerns about this experiment' (open-ended question). If a participant chose 1 for any of the first two questions or expressed suspicion about the experiment in the third question, we excluded him/her from data analysis.

Participants were randomly assigned to the inhibitory group (i.e., cathodal stimulation) or the control group (i.e., sham stimulation). Before the main task, the participants were familiarized with the task with a practice block of 8 trials. They performed the task while receiving cathodal or sham stimulation. To test whether fairness perception was affected by tDCS, participants indicated, before and after the tDCS stimulation, which of the ten different allocation schemes (from 0 to 20 *yuan* in steps of 2) to the receiver was fair.

Results

To achieve a similar measure of the degree of compliance in both the gain and the loss domains, we computed the distance between the participant's allocation and the least compliance situation in each context. Let us suppose that the participant's allocation in a given trial is x. In the gain context, the degree of compliance, according to our definition, is x–0 = x, which is straightforward. In the loss context, it is x– (−20) = 20+ x. For example, if the allocation is −16 for the partner and −4 for the participant, then the degree of compliance to fairness norm is 20+ (−16) = 4. Thus, in both the gain and the loss contexts, the lowest level of compliance is 0, which means the participant allocates all the 20 points of gain to him/herself in the gain context or all the 20 points of loss to the partner in the loss context. In the fairest situation, the degree of compliance is 10, both for the gain and the loss contexts. Any amount higher than 10 reflects benevolence of the participant. In the following data analysis, we used this measure as dependent variable.

Loss context influences voluntary compliance and compliance under sanction threat differently

To answer our first and second questions, we carried out a two (context: gain vs. loss) by two (threat: Retain vs. Waive) within-subject ANOVA on the degree of compliance only for the sham group. The main effect of context was significant, $F_{1,28} = 5.40$, $P = 0.028$, partial $\eta^2 = 0.16$, such that the participants in general showed higher norm compliance in loss than in gain domain. Moreover, the context-by-threat interaction was significant, $F_{1,28} = 6.34$, $P = 0.018$, partial $\eta^2 = 0.19$. Pairwise comparison showed that relative to the gain context, the voluntary norm compliance (i.e., when the sanction threat was waived) in the loss context was significantly higher, $t(28) = 2.81$, $P = 0.009$ (Bonferroni-corrected for multiple comparison). However, no such difference was observed for the norm compliance under sanction threat (i.e., when the sanction threat was retained; Fig. 2A), $t(28) = 0.93$, $P > 0.1$.

lPFC/lOFC is causally involved in norm compliance only in gain domain

To answer our third question, we carried out a two (context: Gain vs. Loss) by two (threat: Retain vs. Waive) by two (treatment: sham vs. cathodal) mixed-design ANOVA on the degree of compliance. The three-way interaction was significant, $F_{1,57} = 6.08$, $P = 0.017$, partial $\eta^2 = 0.10$ (Fig. 2A). We then carried out two ANOVAs for gain and loss domains separately. For gain domain, the two (treatment: sham vs. cathodal) by two (threat: Retain vs. Waive) interaction was significant, $F_{1,57} = 9.99$, $P = 0.003$, partial $\eta^2 = 0.15$. Pairwise comparison revealed that relative to the sham group, the cathodal group showed significantly higher voluntary compliance (i.e., when the threat was waived), $t(57) = 2.03$, $P = 0.047$ (Bonferroni-corrected for multiple comparison). The norm compliance under sanction threat was not affected by tDCS condition, $t(57) = 1.17$, $P > 0.1$. For loss domain, the main effect of threat was significant, $F_{1,57} = 12.55$, $P = 0.001$, partial $\eta^2 = 0.18$, whereas the interaction between threat and treatment was not, $F_{1,57} = 0.12$, $P > 0.1$,

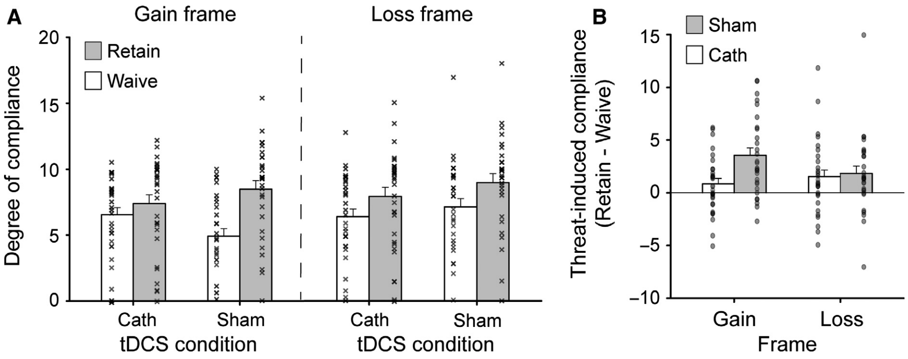

FIG. 2. Degree of compliance and sanction threat-induced (i.e., strategic) compliance as a function of context, sanction threat, and tDCS condition. (A) Degree of compliance, defined as the difference between the points of gain (e.g., +7) or loss (e.g., −16) that the participant allocated to the partner and the least possible compliance situation (0 in the gain context, −20 in the loss context), was affected both by the gain/loss context and the tDCS condition. (B) Disruption of the lPFC/lOFC function reduced the threat-induced compliance in the gain domain, but not in the loss domain. Error bars indicate standard error.

indicating that lPFC/lOFC may not play a direct role in mediating norm compliance in the loss-sharing situation.

To view the data from another perspective, we calculated the degree of threat-induced (or strategic) compliance for each domain and each treatment group by subtracting the degree of voluntary compliance from that of the compliance under sanction threat (Retain-Waive; cf. Ruff *et al.*, 2013). This analysis is not independent of the above analysis for the data in Fig. 2A, but it allows us to make cross-study comparison (e.g., Ruff *et al.*, 2013). It is clear from Fig. 2B that disruption of the lPFC/lOFC function reduced the threat-induced compliance in the gain domain, but not in the loss domain. To compare threat-induced compliance between the current study and that in Ruff *et al.* (2013) (termed "sanction-induced compliance" there), we carried out a two (context: Gain vs. Loss) by two (treatment: sham vs. cathodal) mixed-design ANOVA for threat-induced compliance. The two-way interaction between context and treatment was significant, $F_{1,57} = 6.08$, $P = 0.017$, partial $\eta^2 = 0.10$. Pairwise comparison showed that in the gain domain, cathodal tDCS significantly reduced threat-induced compliance, $t(57) = 3.16$, $P = 0.003$ after Bonferroni-correction for multiple comparison, replicating Ruff *et al.* (2013; see their Fig. 2A). By contrast, the tDCS effect was not significant for the loss domain, $t(57) = 0.34$, $P > 0.1$. Viewed in an alternative way, the difference in threat-induced compliance between gain and loss domains was significant only in the sham group, t (28), $P = 0.018$, not in the cathodal group, $t(29) = 0.98$, $P > 0.1$. This indicated that the flexibility in adjusting one's strategy across contexts relies causally on the function of lPFC/lOFC.

Fairness perception is not affected by gain-loss context or tDCS condition

To test whether participants' perception of fairness norm was affected by gain–loss context and tDCS condition, we carried out a three-way ANOVA with time (before vs. after experiment), context (gain vs. loss), and tDCS treatment (sham vs. cathodal) as independent variable, and the perceived fairness ratings as dependent variable. Due to a technical error, the perceived fairness ratings from 10 participants in the cathodal group and nine participants in the sham group were not available. Nevertheless, we estimated the effects based on the rating data from other 40 participants (20 for the cathodal group, 20 for the sham group). Results revealed neither significant main effects nor interactions, ps > 0.1 (Table 1), which was in line with a number of previous brain stimulation studies on fairness and norm compliance (Knoch *et al.*, 2006, 2008; Ruff *et al.*, 2013; Zhang *et al.*, 2016). These studies consistently showed that brain stimulation changes decision-making while leaves the knowledge about social norm intact. These results implied that the knowledge of fairness norm is not affected by lPFC/lOFC function or gain–loss context.

Discussion

Using HD-tDCS and a modified dictator game, we investigated the context-dependent nature of norm compliance behavior and its neural mechanism. We found that, relative to the gain-sharing context, participants in general conformed more to the fairness norm in the loss-sharing context, especially for voluntary norm compliance. Moreover, disruption of lPFC/lOFC function selectively increased voluntary norm compliance, but not the compliance under sanction threat, in the gain context. In the loss context, neither the voluntary compliance nor the compliance under sanction threat was influenced by lPFC/lOFC function. In other words, these findings showed that disruption of lPFC/lOFC function selectively reduced the strategic compliance in the gain, but not in the loss-sharing context.

How to interpret the higher norm compliance in the loss context? One clue may come from the studies that investigate how gain/loss context influences costly punishment (Leliveld *et al.*, 2009; Zhou & Wu, 2011; Guo *et al.*, 2013; Wu *et al.*, 2014). In these studies, the participant carried out an ultimatum game as a responder, considering whether to accept or reject a monetary division proposed by a co-player (the proposer). If the participant accepts, both get their shares according to the division; if he/she rejects, both get nothing (in the gain context) or lose the entire stake (in the loss context). These studies consistently showed that participants had higher

TABLE 1. Perceived fairness as a function of context, sanction threat, and tDCS condition

	Cathodal		Sham	
Frame	Before	After	Before	After
Gain	9.2 (1.5)	8.7 (1.0)	9.1 (1.5)	9.0 (1.8)
Loss	9.0 (1.0)	9.1 (1.0)	9.8 (1.7)	9.8 (2.0)

Values in brackets are standard deviations.

rejection rates in the loss context than in the gain context, suggesting that they were more willing to suffer personal cost to punish norm violators in the loss context. Using functional MRI, Wu *et al.* (2014) further demonstrate that rejecting unfair offers in the loss domain activate the dorsal striatum, an indication of rewarding and satisfactory experience (see also De Quervain *et al.*, 2004; Crockett *et al.*, 2013). It is thus clear from these studies that people have higher demand for fairness in the loss-sharing context. It is possible that in the current study, the participants were (implicitly or explicitly) aware of the higher demand of norm compliance in the loss domain and behaved accordingly.

Alternatively, although allocating less gain and allocating more loss to the co-player equally deviate from fairness norm, these two types of behaviors may induce different feelings, as incurring loss is more easily appraised as a kind of harm and thus is more likely to elicit the feeling of guilt (cf. Van Beest *et al.*, 2005). Harm aversion (or guilt aversion) theory suggests that when making a social/moral decision, people are motivated to minimize the potential harm incurred to other people and the guilt elicited by such harm (Charness & Dufwenberg, 2006; Chang *et al.*, 2011; Crockett *et al.*, 2014; Nihonsugi *et al.*, 2015; Yu *et al.*, 2015). It is possible that in the loss context the participants are more harm/guilt averse and thus more likely to conform to the fairness norm. These two possible psychological mechanisms are not mutually exclusive and future studies are needed to formally distinguish them.

The lOFC is widely held to be critical for behavioral flexibility in social (e.g., Lee, 2008; Nelson & Guyer, 2011) and non-social decision-making (e.g., Ghahremani *et al.*, 2010; Kehagia *et al.*, 2010). In the gain context, where fairness demand or harm/guilt aversion is not strong enough to dominate decision-making, participants may have more space to adjust their allocation according to specific interactive context (e.g., sanction threat retained vs. waived). Such flexible adjustment seems to be blocked by the disruption of the lOFC function, as the cathodal group made similar allocations in the sanction retained and sanction threat-waived conditions. This interpretation is in line with lOFC's role in mediating task switch ability and reversal learning, where the agent has to inhibit a familiar behavioral rule and adjust to a new one (Ghahremani *et al.*, 2010; Kehagia *et al.*, 2010). It is thus conceivable that disruption of lOFC function impairs such ability and the participants cannot adapt their responses to different interactive contexts. The participants generally conformed more to fairness norm in the loss context and showed less flexible adjustment in the loss context, probably because the fairness demand or guilt/harm aversion motivation was high, and thus the space for strategic adjustment is limited.

It is worth comparing the current findings with previous neuroimaging and a brain stimulation studies concerning the effect of sanction threat on norm compliance (Spitzer *et al.*, 2007; Ruff *et al.*, 2013). At first glance, the setup in our study that a computer program randomly decides which round involves sanction threat is similar to the non-social (computer) condition in Ruff *et al.* (2013) and Spitzer *et al.* (2007). However, this first impression is misleading. In Ruff *et al.* (2013) and Spitzer *et al.* (2007) the computer condition was introduced as a non-social control, where the participants were aware that they did not interact with real human partners and that their allocation decisions did not affect the welfare of any human being. In such context, no social norm prescribes how to share resources with the computer. In contrast, in our study, the computer decided whether sanction threat was present in a given round; the participants believed that they were always interacting with real human partners; their allocation decision thus affected the welfare of another human participant, and they may or may not be punished (randomly determined by computer) if their allocation failed to meet the expectation of that human participant. Therefore, although a computer agent was involved in our study, this should not be mistaken as the non-social control in Ruff *et al.* (2013) and Spitzer *et al.* (2007), but instead should be understood as an unintentional social interaction. We deliberately chose the computer (i.e., unintentional) setting in our study to disentangle the effect of intention and the effect of gain–loss frame, as our previous study has demonstrated the modulatory effect of the intention behind sanction threat on norm compliance (Zhang *et al.*, 2016).

Another critical aspect where our setting is similar to the social, rather than the non-social, condition in Ruff *et al.* (2013) and Spitzer *et al.* (2007) is that in all of these studies, the participants did not know, at the time of the decision, whether the sanction would actually be implemented. In other words, the participants were only aware of and responded to a threat of sanction, rather than the actual sanction *per se*. Moreover, whether a given round should involve such a threat of sanction was not determined by the opponent (or partner), even in the human (i.e., social) condition (Spitzer *et al.*, 2007; Ruff *et al.*, 2013), which is similar to the setting in the current study. The allocation in the gain context in our study replicated the pattern reported in Ruff *et al.* (2013), indicating that the right DLPFC–lOFC complex are consistently involved in mediating the effect of sanction threat on norm compliance in gain-sharing context. New to the existing knowledge is our finding that in loss-sharing context, the dependence of the norm compliance under sanction threat on lPFC/lOFC was abolished, suggesting that other psychological processes (e.g., harm/guilt aversion) may play a role loss-sharing context.

It should be acknowledged that having both cathodal and anodal manipulation could further confirm our conclusion concerning the functional role of lPFC/lOFC in norm compliance in different contexts. It may be argued that the effects of cathodal tDCS are less well-characterized than the effects of anodal tDCS. For example, Jacobson *et al.* (2012) showed that cathodal stimulation produces more variable direction and magnitude of effects from study to study. However, this criticism comes with a qualification: a closer examination of their paper showed that the likelihood to generate Anodal excitation (Ae) effects was significant larger than the likelihood to generate cathodal inhibition (Ci) effects *only* in language studies; in studies of executive function and decision-making, there was no significant difference between the Ae and Ci effects. In fact, cathodal stimulation has been repeatedly suggested to effectively interrupt the activity of specific brain regions and influence individuals' social and non-social decision-making (Knoch *et al.*, 2006, 2008; Ruff *et al.*, 2013; Mengarelli *et al.*, 2015; Shen *et al.*, 2016; Maréchal *et al.*, 2017). Specifically, in our previous work on norm compliance (Zhang *et al.*, 2016), we included both types of stimulation and we demonstrated that opposite effects on norm compliance were produced by cathodal and anodal stimulations. Therefore, we believe that both cathodal and anodal stimulations are effective in producing reliable and opposite effects on brain activity and behaviors, at least in the current social decision-making context.

In conclusion, on the basis of previous research, the current study goes one step further in understanding the context-dependent nature of norm compliance behavior and its underlying brain basis. Individuals are more likely to conform to fairness norm and rely less on the existence of external sanction threat in the loss-sharing context than in the gain-sharing context. The right lPFC/lOFC is causally involved in flexibly adjusting compliance behavior to the interactive context (e.g., threat-on vs. threat-off) in the gain context and such requirement is abolished in the loss context, probably because other motivations (e.g., enhanced fairness demand or harm/guilt aversion) become prominent in loss domain.

Acknowledgement

This work was supported by the Natural Science Foundation of China (Grant 31630034) and the National Basic Research Program of China (973 Program: 2015CB856400). HY was supported by a Newton International Fellowship from the British Academy (NF160700).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

YY, HY, YZ, and XZ contributed to the design of the study and wrote the paper; YY and ZS collected the data; YY and HY analyzed the data.

Data accessibility

The article's supporting data can be accessed through the journal's Figshare page.

References

Borckardt, J.J., Bikson, M., Frohman, H., Reeves, S.T., Datta, A., Bansal, V., Madan, A., Barth, K. *et al.* (2012) A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. *J. Pain*, **13**, 112–120.

Buchan, N., Croson, R., Johnson, E. & Wu, G. (2005) Gain and loss ultimatums. *Adv. Appl. Microecon.*, **13**, 1–23.

Buhl, T. (1999) Positive-negative asymmetry in social discrimination: meta-analytical evidence. *Group Processes Interg.*, **2**, 51–58.

Campbell-Meiklejohn, D.K., Bach, D.R., Roepstorff, A., Dolan, R.J. & Frith, C.D. (2013a) How the opinion of others affects our valuation of objects. *Curr. Biol.*, **20**, 1165–1170.

Campbell-Meiklejohn, D.K., Kanai, R., Bahrami, B., Bach, D.R., Dolan, R.J., Roepstorff, A. & Frith, C.D. (2012) Structure of orbitofrontal cortex predicts social influence. *Curr. Biol.*, **22**, R123–R124.

Caparelli-Daquer, E.M., Zimmermann, T.J., Mooshagian, E., Parra, L.C., Rice, J.K., Datta, A., Bikson, M. & Wassermann, E.M. (2012) A pilot study on effects of 4x1 high-definition tDCS on motor cortex excitability. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, **2012**, 735–738.

Chang, L., Smith, A., Dufwenberg, M. & Sanfey, A. (2011) Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, **70**, 560–572.

Charness, G. & Dufwenberg, M. (2006) Promises and partnership. *Econometrica*, **74**, 1579–1601.

Crockett, M.J., Apergis-Schoute, A., Herrmann, B., Lieberman, M.D., Müller, U., Robbins, T.W. & Clark, L. (2013) Serotonin modulates striatal responses to fairness and retaliation in humans. *J. Neurosci.*, **33**, 3505–3513.

Crockett, M.J., Kurth-Nelson, Z., Siegel, J.Z., Dayan, P. & Dolan, R.J. (2014) Harm to others outweighs harm to self in moral decision making. *Proc. Natl. Acad. Sci.*, **111**, 17320–17325.

Datta, A., Bansal, V., Diaz, J., Patel, J., Reato, D. & Bikson, M. (2009) Gyri-precise head model of transcranial direct current stimulation: improved spatial focality using a ring electrode versus conventional rectangular pad. *Brain Stimul.*, **2**, 201–207.

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

Douglas, Z.H., Maniscalco, B., Hallett, M., Wassermann, E.M. & He, B.J. (2015) Mod- ulating conscious movement intention by noninvasive brain stimulation and the underlying neural mechanisms. *J. Neurosci.*, **35**, 7239–7255.

Edwards, D., Cortes, M., Datta, A., Minhas, P., Wassermann, E.M. & Bikson, M. (2013) Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: a basis for high-definition tDCS. *NeuroImage*, **74**, 266–275.

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**, 175–191.

Fehr, E. & Gachter, S. (2002) Altruistic punishment in humans. *Nature*, **415**, 137–140.

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**, 845–850.

Ghahremani, D.G., Monterosso, J., Jentsch, J.D., Bilder, R.M. & Poldrack, R.A. (2010) Neural components underlying behavioral flexibility in human reversal learning. *Cereb. Cortex*, **20**, 1843–1852.

Guo, X., Li, Z., Lei, Z., Li, J., Wang, Q., Dienes, Z. & Yang, Z. (2013) Increased neural responses to unfairness in a loss context. *NeuroImage*, **77**, 246–253.

Güth, W. & Damme, E.V. (1998) Information, strategic behavior, and fairness in ultimatum bargaining: an experimental study. *J. Math. Psychol.*, **42**, 227–247.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. & Gurven, M. *et al.* (2006) Costly punishment across human societies. *Science*, **312**, 1767–1770.

Jacobson, L., Koslowsky, M. & Lavidor, M. (2012) tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Exp. Brain Res.*, **216**, 1–10.

Kehagia, A.A., Murray, G.K. & Robbins, T.W. (2010) Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Curr. Opin. Neurobiol.*, **20**, 199–204.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, **314**, 829–832.

Knoch, D., Nitsche, M.A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A. & Fehr, E. (2008) Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cereb. Cortex*, **18**, 1987–1990.

Kuo, H.I., Bikson, M., Datta, A., Minhas, P., Paulus, W., Kuo, M.F. & Nitsche, M.A. (2013) Comparing cortical plasticity induced by conventional and high-definition 4 x 1 ring tDCS: a neurophysiological study. *Brain Stimul.*, **6**, 644–648.

Lee, D. (2008) Game theory and neural basis of social decision making. *Nat. Neurosci.*, **11**, 404–409.

Leliveld, M.C., Beest, I., Dijk, E. & Tenbrunsel, A.E. (2009) Understanding the influence of outcome valence in bargaining: a study on fairness accessibility, norms, and behavior. *J. Exp. Soc. Psychol.*, **45**, 505–514.

Li, J., Xiao, E., Houser, D. & Montague, P.R. (2009) Neural responses to sanction threats in two-party economic exchange. *Proc. Natl. Acad. Sci. USA*, **106**, 16835–16840.

Manuel, A., David, A., Bikson, M. & Schnider, A. (2014) Frontal tDCS modulates orbitofrontal reality filtering. *Neuroscience*, **265**, 21–27.

Maréchal, M.A., Cohn, A., Ugazio, G. & Ruff, C.C. (2017) Increasing honesty in humans with noninvasive brain stimulation. *Proc. Natl. Acad. Sci. USA*, **114**, 4360–4364.

Mengarelli, F., Spoglianti, S., Avenanti, A. & Di Pellegrino, G. (2015) Cathodal tDCS over the left prefrontal cortex diminishes choice-induced preference change. *Cereb. Cortex*, **25**, 1219–1227.

Minhas, P., Bansal, V., Patel, J., Ho, J.S., Diaz, J., Datta, A. & Bikson, M. (2010) Electrodes for high-definition transcutaneous DC stimulation for applications in drug delivery and electrotherapy, including tDCS. *J. Neurosci. Meth.*, **190**, 188–197.

Mondino, M., Haesebaert, F., Poulet, E., Saoud, M. & Brunelin, J. (2015) Efficacy of cathodal transcranial direct current stimulation over the left orbitofrontal cortex in a patient with treatment-resistant obsessive-compulsive disorder. *J ECT.*, **31**, 271–272.

Montague, P.R. & Lohrenz, T. (2007) To detect and correct: norm violations and their enforcement. *Neuron*, **56**, 14–18.

Nelson, E.E. & Guyer, A.E. (2011) The development of the ventral prefrontal cortex and social flexibility. *Dev. Cogn. Neurosci.*, **1**, 233–245.

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**, 3412–3419.

Rawls, J. (1958) Justice as fairness. *Philos. Rev.*, **67**, 164–194.

Reeve, C.D.C. (1998). *Aristotle: Politics*. Hackett, Indianapolis, IN.

Ruff, C.C., Ugazio, G. & Fehr, E. (2013) Changing social norm compliance with noninvasive brain stimulation. *Science*, **342**, 482–484.

Rushworth, M.F.S., Behrens, T.E.J., Rudebeck, P.H. & Walton, M.E. (2007) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci.*, **11**, 168–176.

Rutledge, R.B., Smittenaar, P., Zeidman, P., Brown, H.R., Adams, R.A., Lindenberger, U., Dayan, P. & Dolan, R.J. (2016) Risk taking for potential reward decreases across the lifespan. *Curr. Biol.*, **26**, 1634–1639.

Shen, B., Yin, Y., Wang, J., Zhou, X., McClure, S.M. & Li, J. (2016) High-definition tDCS alters impulsivity in a baseline-dependent manner. *NeuroImage*, **143**, 343–352.

Sober, E. & Wilson, D.S. (1998). *Unto Others: the Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge, MA.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G. & Fehr, E. (2007) The neural signature of social norm compliance. *Neuron*, **56**, 185–196.

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B. & Sack, A. (2015) Be nice if you have to – the neurobiological roots of strategic fairness. *Soc. Cogn. Affect. Neur.*, **10**, 790–796.

Tom, S.M., Fox, C.R., Trepel, C. & Poldrack, R.A. (2007) The neural basis of loss aversion in decision-making under risk. *Science*, **315**, 515–518.

Tversky, A. & Kahneman, D. (1981) The framing of decisions and the psychology of choice. *Science*, **211**, 453–458.

Tversky, A. & Kahneman, D. (1991) Loss aversion in riskless choice: a reference-dependent model. *Quart. J. Econ.*, **106**, 1039–1061.

Ursu, S. & Carter, C.S. (2005) Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. *Cognitive Brain Res.*, **23**, 51–60.

Van Beest, I., Van Dijk, E., De Dreu, C.K. & Wilke, H.A. (2005) Do-no-harm in coalition formation: why losses inhibit exclusion and promote fairness cognitions. *J. Exp. Soc. Psychol.*, **41**, 609–617.

Villamar, M.F., Volz, M.S., Bikson, M., Datta, A., DaSilva, A.F. & Fregni, F. (2013a) Technique and considerations in the use of 4x1 ring high-definition transcranial direct current stimulation (HD-tDCS). *J. Vis. Exp.*, **77**, e50309.

Villamar, M.F., Wivatvongvana, P., Patumanond, J., Bikson, M., Truong, D.Q., Datta, A. & Fregni, F. (2013b) Focal modulation of the primary motor cortex in fibromyalgia using 4x1-ring high-definition transcranial direct current stimulation (HD-tDCS): immediate and delayed analgesic effects of cathodal and anodal stimulation. *J. Pain*, **14**, 371–383.

Watson, K.K. & Platt, M.L. (2012) Social signals in primate orbitofrontal cortex. *Curr. Biol.*, **22**, 2268–2273.

Willis, M.L., Palermo, R., Burke, D., McGrillen, K. & Miller, L. (2010) Orbitofrontal cortex lesions result in abnormal social judgements to emotional faces. *Neuropsychologia*, **48**, 2182–2187.

Willis, M.L., Murphy, J.M., Ridley, N.J. & Vercammen, A. (2015) Anodal tDCS targeting the right orbitofrontal cortex enhances facial expression recognition. *Soc. Cogn. Affect. Neur.*, **12**, 1677–1683.

Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., Jiang, Y. & Zhou, X. (2014) Neural basis of increased costly norm enforcement under adversity. *Soc. Cogn. Affect. Neur.*, **9**, 1862–1871.

Yu, H., Shen, B., Yin, Y., Blue, P.R. & Chang, L.J. (2015) Dissociating guilt-and inequity-aversion in cooperation and norm compliance. *J. Neurosci.*, **35**, 8973–8975.

Zhang, Y., Yu, H., Yin, Y. & Zhou, X. (2016) Intention modulates the effect of punishment threat in norm enforcement via the lateral orbitofrontal cortex. *J. Neurosci.*, **36**, 9217–9226.

Zhou, X. & Wu, Y. (2011) Sharing losses and sharing gains: increased demand for fairness under adversity. *J. Exp. Soc. Psychol.*, **47**, 582–588.