Perturbation of Right Dorsolateral Prefrontal Cortex (rDLPFC) Makes Power-Holders Less Resistant to Tempting Bribes

Yang Hu, Remi Philippe, Valentin Guigon, S Zhao, Edmund A. Derrington, Brice Corgnet, James Bonaiuto, Jean-Claude Dreher

Perturbation of Right Dorsolateral Prefrontal Cortex (rDLPFC) Makes Power-Holders Less Resistant to Tempting Bribes

Yang Hu[1,2], Rémi Phillipe[2,3†], Valentin Guigon[2,3†], Sasa Zhao[2,3†], Edmund Derrington[2,3], Brice Corgnet[4], James Bonaiuto[2,3], Jean-Claude Dreher[2,3*]

[1]School of Psychology and Cognitive Science, East China Normal University, Shanghai, China
[2]Neuroeconomics, Reward and Decision Making Laboratory, Institut des Sciences Cognitives Marc Jeannerod, CNRS, France
[3]Université Claude Bernard Lyon 1, Lyon, France
[4]EmLyon, Ecully, France

[*]Please address correspondence to:

Jean-Claude Dreher

CNRS UMR 5229,
Neuroeconomics, Reward and Decision Making Laboratory,
Institut des Sciences Cognitives Marc Jeannerod,
67 Bd Pinel, 69675 Bron, France
Tel: 00 334 37 91 12 38
Email: dreher@isc.cnrs.fr

[†]These authors equally contributed to this study.

Abstract

Bribery is a common form of corruption that takes place when a briber suborns a power-holder to achieve an advantageous outcome at a cost of moral transgression. While bribery has been extensively investigated in behavioral sciences, its underlying neurobiological basis remains poorly understood. Here we employed transcranial direct current stimulation (tDCS) in combination with a novel paradigm to investigate whether disruption of right dorsolateral prefrontal cortex (rDLPFC) causally changed bribe-taking decisions of power-holders. Perturbing rDLPFC via tDCS specifically made participants more willing to take bribes when the offer proportion ramped up. This tDCS-induced effect could not be explained by changes in other measures. Model-based analyses further revealed that such neural modulation alters the concern for profiting oneself via taking bribes and reshapes that for the distribution inequity between oneself and the briber, thereby influencing the subsequent decisions. These findings reveal a causal role of rDLPFC in modulating corrupt behavior.

Statement of Relevance

Bribery often occurs in interpersonal contexts when bribers suborn power-holders who can act in the bribers' interest, which brings mutual gains but violates the moral principle. How does a power-holder decide whether to take the bribe or not? What are the computational and neurobiological roots underlying bribery behaviors? Combining transcranial direct current stimulation (tDCS) with a novel task, we examined the causal role of the right dorsolateral prefrontal cortex (rDLPFC) in modulating the bribe-taking behaviors of power-holders and the underlying computational process. In particular, disrupting rDLPFC via tDCS specifically made power-holders more willing to accept tempting bribes, putatively through modulating the bribery-elicited moral cost on concern for personal gains and the distribution inequity between oneself and the briber. These findings provide insights for the neurobiological roots of corruption and suggest interventions to modify corrupt behaviors using non-invasive brain stimulation techniques.

Introduction

As one of the most common forms of corruption, bribery pervasively exists in governments, enterprises, and other organizations all over the world (Dreher, Kotsogiannis, & McCorriston, 2007). In real life, bribes usually occur in interpersonal contexts where there is an asymmetry in power between the parties involved, such as a power-holder who can exert an impact in the briber's interest (Köbis, van Prooijen, Righetti, & Van Lange, 2016). Hence, bribes often result in mutual benefits via collaboration between the two parties involved, but transgress moral principles and legal rules. Despite that bribery-related issues have been widely investigated in social sciences (Abbink, 2006; Mauro, 1995; Serra & Wantchekon, 2012), the neurobiological roots of bribery and their underlying computations remain largely elusive.

How does a power-holder decide whether to take or refuse a bribe? Bribery-related decision-making is supposed to follow the general framework of value-based decision-making (Rangel, Camerer, & Montague, 2008) and the account of social preference (Fehr & Krajbich, 2014). In a simplified situation, a power-holder makes a choice based on a relative subjective value (SV) between accepting and rejecting the bribe, calculated by pitting personal profits against the other-regarding interests. Moreover, accepting a bribe often involves the transgression of the moral principle and brings in moral costs, which affects the SV computation (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014). A recent study has identified the moral cost of colluding with a fraud committed by the briber, incurred by the power-holder, which depreciates the decision weights on personal gains from the bribe and thus decreases the acceptance rates (Hu et al., 2021). Notably, the moral cost of taking the bribe is critically distinguished from the psychological cost of dishonesty (Fischbacher & Föllmi-Heusi, 2013; Gneezy, Kajackaite, & Sobel, 2018; Mazar, Amir, & Ariely, 2008). In these studies, the moral cost occurs if an individual cheats for personal profits, whereas in the bribery scenario the moral cost for a power-holder is elicited due to colluding with a briber to obtain morally-tainted benefits via taking a bribe.

It is well-established that the right dorsolateral prefrontal cortex (rDLPFC) is critically involved in modulating human social/moral behaviors. Specifically, previous studies using an ultimatum game (UG) consistently showed that decreasing the neural excitability of rDLPFC either by low-frequency repetitive transcranial magnetic stimulation (TMS) or by cathodal transcranial direct current stimulation (tDCS) makes the respondents more likely to accept disadvantageous offers (Knoch, Nitsche, Fischbacher, Eisenegger, & Fehr, 2008; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Speitel, Traut-Mattausch, & Jonas, 2019). In the moral domain, inhibiting rDLPFC and related anterior prefrontal areas with cathodal tDCS improves deceptive behaviors by reducing the reaction time to tell lies and increasing skillful lies (Karim et al., 2010). Using a different task, a brain-lesion study has illustrated that patients with DLPFC lesions selectively increased self-serving cheating behaviors (Zhu et al., 2014).

Concerning the anodal tDCS effect over rDLPFC on social/moral behaviors, the current evidence is less clear. There is no evidence supporting an increase of intolerance of inequity for the responder in the UG task via anodal tDCS (Speitel et al., 2019). Regarding the moral behaviors, participants receiving anodal tDCS are more likely to behave honestly (Maréchal, Cohn, Ugazio, & Ruff, 2017). Yet, there is also evidence that anodal tDCS over DLPFC speeds up dishonest decisions, suggesting an opposite effect (Mameli et al., 2010). Moreover, a recent fMRI study indicates that the DLPFC guides anti-corrupt behaviors contextually and selectively modulates bribery-specific computations across individuals (Hu et al., 2021). Together, these results suggest that the rDLPFC should play a pivotal role in bribery-related decision making, but how its disruption specifically impacts corrupt acts and the underlying computations remains unclear.

Here, to examine whether rDLPFC exerts a causal influence in determining whether a power-holder would accept a bribe or not, we manipulated the neural excitability of rDLPFC *via* tDCS and measured corrupt behaviors of power-holders using a novel paradigm. In particular, a total of 120 healthy participants were randomly assigned to one of three tDCS groups to causally modulate (anodal or cathodal tDCS), or maintain (sham tDCS) the neural excitability of rDLPFC (see **Figure 1**; also see **Figure S1** in the **SOM**). Participants played the role of a power-holder, who decides whether another (fictitious) person in a separate game would earn a given amount of money or not in a fraudulent (the Bribe condition) or morally proper manner (the Control condition). To achieve this, this person, denoted as a proposer, proposed an offer to influence the power-holder's decision. The task for the participants was to decide whether to accept or reject the offer made by the proposer. If accepted, both the proposer and the participant would profit from the offer, whereas neither would earn any money if the participant rejected the offer (see **Figure 2**). Since deciding in the Bribe (vs. Control) condition additionally brings in the ethical concern of colluding with a briber, this design allows us to uncover the specific role of the rDLPFC in bribery-related decision-making.

Based on our recent study on corruption and previous literature that revealed a role of moral cost on ethical decision-making, we hypothesized that participants would be generally less willing to accept the offers in the Bribe (vs. Control) condition. More importantly, according to the tDCS literature mentioned above, we expected that participants receiving cathodal (vs. sham) tDCS over the rDLPFC would be more likely to accept offers in the Bribe (vs. Control) condition, especially when larger offers were proposed. In contrast, we did not form a specific hypothesis about how anodal tDCS affects corrupt behaviors due to its mixed effect on social and moral behaviors. Moreover, we tested several computational models and identified the one that best characterized actual behaviors for all tDCS groups, which warrants us to delineate how rDLPFC specifically contributes to the computations underlying corrupt acts.

Methods

Participants

One-hundred and twenty French-speaking students from University of Lyon I and local residents (54 females; mean age: 22.4 ± 4.4 years) were recruited via online advertisements. The sample size was adopted based on previous tDCS studies in similar topics (Maréchal et al., 2017; Ruff, Ugazio, & Fehr, 2013), which are standard in the field. All participants were psychiatrically and neurologically healthy and were not taking any medication, as confirmed by a standardized clinical screening. The tDCS study was approved by the local ethics committees. All experimental protocols and procedures were conducted in accordance with the IRB guidelines for experimental testing and were in compliance with the latest revision of the Declaration of Helsinki (BMJ 1991; 302: 1194).

Task and Design

Participants were randomly assigned to one of the three tDCS treatment conditions with 40 persons in each: (i) anodal stimulation (18 females; mean age: 22.6 ± 5.5 years), (ii) cathodal stimulation over the rDLPFC (17 females; mean age: 21.9 ± 2.6 years), or (iii) sham stimulation (19 females; mean age: 22.6 ± 4.8 years), which were unbeknownst to them (see **SOM** for tDCS protocol).

The main experiment included a computerized incentive task and a follow-up paper-and-pencil rating task, which lasted around 30 min in total (see **SOM** for procedure details). In the computerized task, participants were assigned the role of the power-holder who decides to accept or reject financial offers (see **Figure 2A**). In a cover story, they were informed that they would be presented with a series of choices from an independent group, whose data were collected previously by the experimenter. Specifically, participants were led to believe that this independent group of online attendants (denoted as proposers hereafter) played a “Game of Chance”. This independent group did not actually exist and the choices made by this group were pre-determined by the task software. Each proposer was presented with two options that would earn them different payoffs. The larger payoff ranged from 60 to 130 (in €; see details below) and the smaller payoff was fixed at 5. One of the two payoffs was randomly indicated by the computer as the one to be received. According to the rules of the game, the proposer should report the payoff indicated by the computer, which determined his final payoff (i.e., the Control condition). However, the response of the proposer was never checked by the experimenters. This allowed the proposer to lie by reporting the alternative payoff that had not been indicated by the computer when this would earn them more profit (i.e., the Bribe condition). In other words, the only difference between the two conditions is that in the Bribe condition the proposer cheated for a larger payoff by reporting the non-chosen larger payoff, whereas the proposer honestly reports the chosen larger payoff in the Control condition. Importantly, participants were told that each proposer had been informed that whether or not they obtained the payoff of the reported option crucially depended on the decisions of a power-holder (i.e., the participants themselves). To obtain the profits in the reported option, the proposer could "share" a portion of the money from their potential gain (i.e., the reported larger payoff) to influence the power-holder's decision. The task for the power-holder was to decide whether to accept or reject the offer based on the information above. If the power-holder accepted the offer, both of them would benefit from the payoff. If the power-holder rejected the offer, neither of them earned anything. Participants were informed that one of their decisions would be randomly selected for payment in that trial at the end of the experiment.

Several aspects of this task merit additional notes. First, participants were informed that each decision was independent and was matched with different proposers to avoid possible learning effects or strategic responses. Second, each participant was actually always paid €30 at the end, as required by the ethics approval board. Finally, we designed the task such that the proposer always reported the option with a larger payoff, and his/her personal profits after "sharing" with the power-holder were always more than the €5 option. This ensured that selfish motivation was the only source that drove the proposer to cheat for a higher payoff, and ruled out other motivations perceived by participants that might influence their subsequent behaviors.

We implemented a 3×2 mixed design by manipulating the *tDCS treatment* (a between-subject factor) and the *task condition* (a within-subject factor). Crucially, we operationally defined corrupt behaviors as the acceptance of offers proposed by the proposer only when the proposer lied (the Bribe condition). Compared with the Control condition, accepting offers in the Bribe condition incurred the moral cost of colluding with the proposer's dishonesty. We also manipulated the *offer proportion*, which was defined as the proportion of the amount the proposer decided to share with the power-holder from the payoff the proposer would have earned in the reported option, ranging from 10% to 90% (in steps of 10%; 9 levels). This allowed us to investigate whether and how the degree of temptation of a bribe modulated corrupt behaviors. To further increase the variance of offers, we set potential gains that could be earned by the proposer (i.e., the larger payoff, which ranged from 60 to 130 in steps of 10; 8 levels). As a result, this yielded 72 trials, each involving a unique offer, which appeared once in each condition.

Each trial began with a screen displaying two payoff options in the "Game of Chance", the computer's choice (indicated by a computer icon), the proposer's report (indicated by a blue arrow) together with the identity of the proposer (indicated by initials of the name), and the proposer's offer. Participants were asked to decide whether to accept or reject the offer by pressing relevant buttons with either left or right index finger at their own pace. A yellow bar appeared below the corresponding option for 0.5 s once the decision was made. Each trial ended up with an inter-trial interval of random duration (1 ~ 2 s; see **Figure 2B**) showing a fixation cross. The order of these trials was randomized across participants to reduce the confounding effect of the condition order. Besides, the positions of payoffs were randomized within participants and those of the choice options were counterbalanced across participants. All stimuli were presented using Presentation v14 (Neurobehavioral Systems Inc., *Albany, CA, USA*). After that, participants were asked to perform a follow-up rating task in which they reported their subjective feelings about the task. Then, they filled out a series of task-irrelevant control measures (see **SOM** for details). They were debriefed, paid and thanked at the end of the experiment.

Data Analyses

One participant in the Cathodal group was excluded for having incomplete data recording due to technical issues, thus leaving a total of 119 participants whose data were further analyzed (overall: 54 females; mean age ± SD = 22.4 ± 4.5 years; Anodal group: 18 females; mean age ± SD = 22.6 ± 5.5 years; Cathodal group: 17 females; mean age ± SD = 22.0 ± 2.5 years; Sham group: 19 females; mean age ± SD = 22.6 ± 4.8 years). Overall, participants did not report any uncomfortable feeling after the experiment and were not able to correctly identify the treatment they were assigned ($\chi^2_{(1)}$ = 1.89, p = 0.169). Since no difference in age ($F_{(2, 116)}$ = 0.26, p = 0.775) and gender ($\chi^2_{(2)}$ = 0.13, p = 0.939) was observed between tDCS groups, we did not include these variables as covariates for later analyses. Behavioral analyses were conducted using R (http://www.r-project.org/) and relevant packages (R Core Team, 2014). Model-based analyses were performed using the hierarchical Bayesian approach (HBA) via "hBayesDM" package (Ahn, Haines, & Zhang, 2017). For methods details, see **SOM**.

Results

tDCS over rDLPFC increased the probability of accepting bribes with higher offer proportions

We first tested our main hypothesis regarding choice behavior. Using mixed-effect logistic regression, we observed that participants were less likely to accept an offer in the bribe (vs. control) condition (a main effect of *task condition*: $\chi^2_{(1)} = 126.94$, $p < 0.001$) and more likely to do so when the offer proportion increased (a main effect of offer proportion: $\chi^2_{(1)} = 96.34$, $p < 0.001$). We also detected a significant two-way interaction between *task condition* and *offer proportion* ($\chi^2_{(1)} = 33.05$, $p < 0.001$). *Post-hoc* analyses indicated that compared with the Control condition, participants were more likely to accept offers when the offer proportion increased in the Bribe condition ($z = 5.41$, $p < 0.001$).

More importantly, we found a significant three-way interaction between *tDCS group*, *task condition*, and *offer proportion* with respect to whether the offer was accepted ($\chi^2_{(2)} = 8.04$, $p = 0.018$; see **Figure 3**). To follow up the three-way interaction, we performed *post-hoc* analyses on choice for each tDCS group that incorporated *task condition*, *offer proportion*, and their interaction as fixed-effect predictors. As a result, compared with the Control condition, participants receiving either type of tDCS stimulation were more likely to accept offers when the offer proportion increased in the Bribe condition (anodal: $z = 4.67$, $p < 0001$; cathodal: $z = 4.34$, $p < 0.001$), which was not the case in the Sham group ($z = 0.67$, $p = 0.501$; see **Table S1** in the **SOM** for details).

Notably, we did not observe any tDCS main effect or related interaction on a series of other behavioral measures, including decision time (DT), task-related subjective ratings, and task-irrelevant measures (see **Figure S2** and **Table S2-S4** in the **SOM** for details).

tDCS over rDLPFC modulated the bribery-elicited moral cost on concern for personal gains (β) and fairness (γ)

Bayesian model comparison showed that Model 1 yielded the lowest LOOIC scores and outperformed other competitive models (Model 2-4; see **SOM** for details).

$$SV(P_{PH}, P_P) = \beta P_{PH} + \lambda P_P + \gamma |P_P - P_{PH}|$$

$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \lambda_{Control}, \gamma_{Control}, if\ Control\ condition \\ \beta_{Bribe}, \lambda_{Bribe}, \gamma_{Bribe}, if\ Bribe\ condition \end{cases} \quad \text{Model 1}$$

In this model, SV denotes the subjective value of the choice. P_P and P_{PH} represents the offer's payoff for the proposer and power-holder respectively given different choices (i.e., accept or reject the offer). β and λ measure the decision weights on personal profits and proposer's gain from the offer respectively. γ measures the sensitivity to the absolute payoff inequality between the power-holder and the proposer. The posterior predictive check (PPC) revealed that the proportion of acceptance predicted by this model could capture the proportion of observed acceptance across individuals (both conditions for all groups: *r*s > 0.99, *p*s < 0.001; see **Figure S3-S7** in the **SOM** for the PPC at various levels), which further justified the validity of our model.

To examine how bribery-elicited moral cost affected each parameter and how tDCS treatment modulated such effects, we implemented mixed-effect linear regression on each parameter separately, by including *tDCS group*, *task condition*, and their interactions as the fixed-effect predictors. We also allowed intercepts to vary across participants as the random effects. As a result, we first found a main effect of task condition for all three parameters, namely that participants devalued the personal gains (β: $F_{(1, 116)} = 18.04$, $p < 0.001$, partial-$\eta^2 = 0.092$), the proposer's gains (λ: $F_{(1, 116)} = 172.64$, $p < 0.001$, partial-$\eta^2 = 0.481$), and the absolute payoff differences (γ: $F_{(1, 116)} = 96.33$, $p < 0.001$, partial-$\eta^2 = 0.320$) in the Bribe condition relative to the Control condition. Furthermore, we observed a main effect of tDCS treatment on γ ($F_{(2, 116)} = 20.42$, $p < 0.001$, partial-$\eta^2 = 0.166$). *Post-hoc* analyses showed that participants in the Anodal (vs. Sham) group decreased their concern for the absolute payoff differences ($t_{(116)} = 3.05$, $p_{(\text{FDR-corrected})} = 0.003$, Cohen's d = 0.55, 95% CI = [0.19, 0.92]), which was even further reduced in the Cathodal group (vs. Anodal: $t_{(116)} = 3.35$, $p_{(\text{FDR-corrected})} = 0.002$, Cohen's d = 0.61, 95% CI = [0.24, 0.98]; see **SOM** for details).

More intriguingly, we found an interaction effect between *tDCS group* and *task condition* on decision weights on personal gains (β: $F_{(2, 116)} = 11.71$, $p < 0.001$, partial-$\eta^2 = 0.116$) and absolute payoff differences (γ: $F_{(2, 116)} = 16.14$, $p < 0.001$, partial-$\eta^2 = 0.320$), but not on proposer's gains (λ: $F_{(2, 116)} = 2.35$, $p = 0.100$, partial-$\eta^2 = 0.025$). *Post-hoc* analyses for β showed that participants receiving cathodal (vs. sham) tDCS decreased weights on personal gains in the Control condition ($t_{(213)} = -2.21$, $p_{(\text{FDR-corrected})} = 0.042$, Cohen's d = 0.59, 95% CI = [-1.13, -0.06]) but increased them in the Bribe condition ($t_{(213)} = 2.55$, $p_{(\text{FDR-corrected})} = 0.035$, Cohen's d = 0.68, 95% CI = [0.15, 1.22]). Anodal tDCS induced a similar effect of β in the Control condition ($t_{(213)} = -3.55$, $p_{(\text{FDR-corrected})} = 0.001$, Cohen's d = 0.95, 95% CI = [-1.48, -0.41]), however, the enhancement effect was not statistically significant in the Bribe condition ($t_{(213)} = 1.58$, $p_{(\text{FDR-corrected})} = 0.172$, Cohen's d = 0.42, 95% CI = [-0.11, 0.95]). Regarding γ, *post-hoc* analyses showed that compared with the Sham group, participants in both the Anodal ($t_{(228)} = 5.91$, $p_{(\text{FDR-corrected})} < 0.001$, Cohen's d = 1.42, 95% CI = [0.93, 1.91]) and Cathodal groups ($t_{(228)} = 7.46$, $p_{(\text{FDR-corrected})} < 0.001$, Cohen's d = 1.80, 95% CI = [1.31, 2.29]) were less aversive to absolute payoff differences (i.e., the general inequality) in the Control condition. However, in the Bribe condition, participants in the Cathodal group were less aversive to the absolute payoff inequality compared with either the Sham ($t_{(228)} = 2.15$, $p_{(\text{FDR-corrected})} = 0.049$, Cohen's d

= 0.52, 95% CI = [0.04, 1.00]) or Anodal group ($t_{(228)}$ = 3.45, $p_{(\text{FDR-corrected})}$ = 0.002, Cohen's d = 0.83, 95% CI = [0.35, 1.32]; see **Figure 4** for the descriptive summary for key parameters; see **Figure S8** in the **SOM** for the visualization of the tDCS effect on differential parameters; also see **Table S5-S7** in the **SOM** for details of statistical analyses)

tDCS over rDLPFC modulates bribery-elicited moral cost on the choice behaviors by mediating key parameters of the computation

To further establish the link between the tDCS treatment, the bribery-elicited moral cost on these parameters, and the choice behaviors, we implemented *post-hoc* mediation analyses with tDCS group as the predictor, the differential parameters as the mediator (i.e., $\Delta\beta = \beta_{Bribe} - \beta_{Control}$, $\Delta\gamma = \gamma_{Bribe} - \gamma_{Control}$), and the differential acceptance rate as the dependent variable (i.e., $\Delta Accept = Accept_{Bribe} - Accept_{Control}$). A bootstrapping procedure was applied to the mediation effect (i.e., 5000 bootstraps). We showed that although the tDCS treatment did not directly modify the bribery-specific effect on choice behaviors (i.e., total effect, path c: *p*s > 0.3 for both tDCS effects), the differential parameters mediated the impact of tDCS treatment on the bribery-specific effect on the behaviors (i.e., direct effect, path c': *p*s < 0.001 in both tDCS effects for Δβ and in the anodal tDCS for Δγ, *p* = 0.007 in the cathodal tDCS for Δγ; indirect effect, path a*b: Δβ: anodal: -0.27, 95% CI: [-0.40, -0.15]; cathodal: -0.26, 95% CI: [-0.39, -0.12]; Δγ: anodal: 0.21, 95% CI: [0.13, 0.30]; cathodal: 0.18, 95% CI: [0.07, 0.28]; see **Figure 5;** also see **Table S8** in the **SOM** for detailed regression outputs).

Discussion

In the present study, we combined tDCS with a novel task that captures the essence of real-life bribery to examine whether rDLPFC causally influences the corrupt behaviors of a power-holder. As predicted, participants are less likely to accept a bribe compared with a standard offer, especially when the bribe becomes more tempting. These results cohere with other studies on moral decision-making (Crockett et al., 2014; Mazar et al., 2008; Qu, Hu, Tang, Derrington, & Dreher, 2020) and confirm the role of moral cost for a power-holder when deciding whether to take a bribe. Model-based analyses further reveal how the underlying computations are influenced during bribery-related decision making. Specifically, participants depreciate personal gains (β) earned by taking the bribes, which replicates the findings of our recent fMRI study on corruption (Hu et al., 2021). In addition, we also observed stronger negative weights for both the proposer's gains (λ) and absolute differences between their payoffs (γ) in the Bribe (vs. Control) condition. This aligns with previous findings showing contextual modulation of subjective valuation to a partner (Bhanji & Delgado, 2014; Delgado, Frank, & Phelps, 2005) or to fairness concern (Gao et al., 2018; Hu et al., 2018). Together, the present study reveals that such bribery-elicited moral cost reshapes not only the valuation of self-profits but also other-regarding interests and thus helps to prevent the power-holder from being corrupted.

More interestingly, the disruption of rDLPFC (i.e., both Anodal and Cathodal groups) made participants, as power-holders, more likely to accept bribes (vs. standard offers) as the size of the prospective payoff increased. Importantly, this tDCS effect over rDLPFC did not influence other measures (e.g., DT, subjective ratings), suggesting that general cognitive or affective processes are less likely to constitute the underlying mechanism. Taking a model-based approach, we further showed that disrupting rDLPFC also alters the computations that contribute to bribery decisions. Specifically, the cathodal tDCS over rDLPFC mitigates the effect of the moral cost on personal gains due to bribe-taking ($\Delta\beta$). This finding coheres with a previous brain-lesion study in which patients with lesions of DLPFC selectively reduced the moral cost to personal profits (Zhu et al., 2014). Moreover, altering the rDLPFC excitability via cathodal tDCS enhanced the effect of the bribery-elicited moral cost on fairness concern ($\Delta\gamma$). As noted previously, studies using a standard UG consistently showed that inhibiting the rDLPFC by low-frequency repetitive TMS (Knoch et al., 2006) or cathodal tDCS (Knoch et al., 2008; Speitel et al., 2019) increases the tolerance of unfairness. While we replicated these findings by showing a less negative γ for the Cathodal (vs. Sham) group in the Control condition, we found that participants in the Cathodal group become more aversive to the inequity between themselves and the proposer. Collectively, these results in the Cathodal group indicate a dual role of rDLPFC during bribery-related decision making: it not only overrides selfish motivation when it conflicts with moral principles (Carlson & Crockett, 2018) but also integrates the moral cost in modulating fairness concern. This account is further endorsed by the mediation analyses, which establish the link between rDLPFC, computations underlying bribery-related decision making and final behaviors.

It is worth noting that the excitation of rDLPFC via anodal tDCS has a similar effect as cathodal tDCS in modulating bribe-taking behaviors and underlying computations. There is no *a priori* reason to believe that anodal and cathodal tDCS should induce opposite behavioral effects in the moral domain. Indeed, previous evidence is mixed concerning the anodal effect on moral behaviors which varies in different paradigms. Although Maréchal *et al*. (2017) showed that anodal tDCS over rDLPFC increased honesty in a die-rolling task, another tDCS study with an instrumental deception paradigm indicated the opposite effect (Mameli et al., 2010). In agreement with this, an fMRI study has also shown that DLPFC is recruited more in dishonest individuals when they have a chance to cheat (Greene & Paxton, 2009). Moreover, the classical polarity-effect of tDCS (i.e., anodal-excitation and cathodal-inhibition) has been shown to be much less common in the cognitive domain than in the motor domain (Jacobson, Koslowsky, & Lavidor, 2012). A systematic review has revealed highly variable effects of tDCS over the DLPFC on cognitive functions such as working memory (Tremblay, Lepage, Latulipe-Loiselle, Fregni, & Théoret, 2014). Such inconsistent effects also exist in the social domain. For example, although inhibiting rDLPFC with cathodal tDCS consistently enhances the tolerance to unfairness (Knoch et al., 2008; Speitel et al., 2019), no evidence suggests that anodal tDCS increases fairness concern (Speitel et al., 2019). Lastly, there are large individual variations in tDCS effects on modulating behaviors (López-Alonso, Cheeran, Río-Rodríguez, & Fernández-del-Olmo, 2014; Wiethoff, Hamada, & Rothwell, 2014) and in the relationship between DLPFC engagement and moral behaviors (Hu et al., 2021; Yin & Weber, 2018). Together, our findings confirm that the classical polarity-effect of tDCS, originally observed in the primary motor cortex, should not be expected to be directly applied to other brain areas and social/moral behaviors such as corruption.

Bribery-elicited moral cost also merits further consideration. In our task, taking bribes is presumed to carry the only moral cost, that of colluding in fraud. In the Control condition no fraud is taking place and therefore the offer is not considered to be a bribe. However, it is likely that an extra moral cost might be involved simply because of the action of accepting bribes.

Due to the present design, it is impossible to isolate this putative moral cost because it always covaries with the other moral cost. Future studies may address this issue.

Overall, the present study provides empirical evidence that perturbing rDLPFC via tDCS causally influences a power-holder's decisions of whether or not to accept a bribe, and modifies the underlying computations. These findings shed light on the neurobiological substrates of corrupt acts and open a new window to investigate corruption using a multi-disciplinary research approach.

Acknowledgments

This research was funded by IDEXLYON from Université de Lyon (project INDEPTH) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005) and of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007) by ANR to JCD, and China Postdoctoral Science Foundation to YH (2019M660007).

Figures

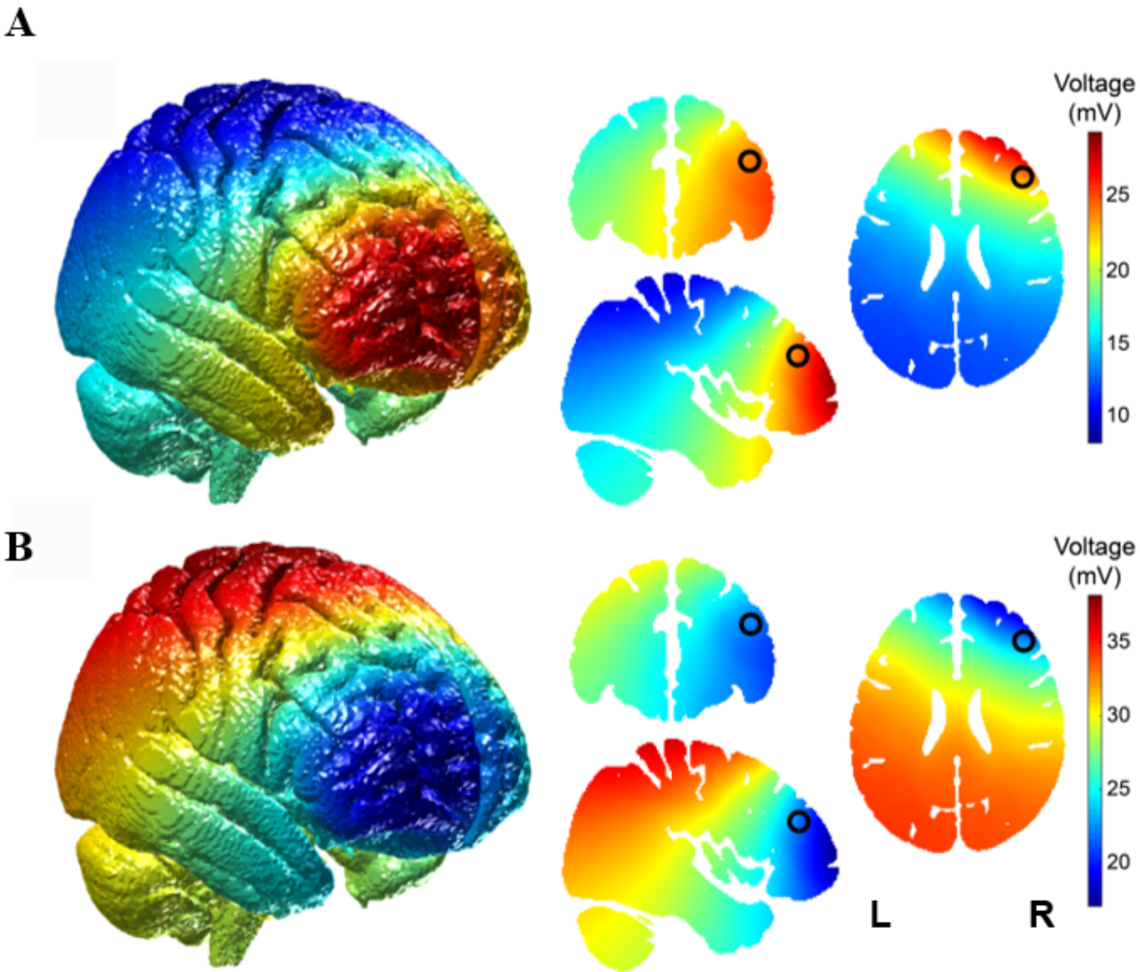

Figure 1. Electric field simulation for (A) anodal and (B) cathodal tDCS stimulation. Based on previous literature closely relevant to the current study (Knoch et al., 2006; Strang et al., 2014), we chose the position centering around the Talaraich coordinate of 39/37/22 as our target site. This location approximately corresponds to the electrode position of AF4 in the 10-20 system of EEG cap (the right panel; marked with a black circle). The vertex was chosen as the reference electrode based on the study by Marechal et al (2017), which corresponds to the electrode position of Cz. Electrodes were simulated as pads, with a 100x100x3mm pad located over Cz and a 70x50x3mm pad located over AF4, using standard 10-10 system locations. Tissue conductivities were set as white matter=0.11 S/m, gray matter=0.21 S/m, CSF=0.53 S/m, bone=0.02 S/m, and skin=0.90 S/m. For the anodal simulation, 1.5mA was set as inward flowing current from the AF4 pad, and -1.5mA outward flowing current from the Cz pad, and vice versa for the cathodal simulation. The simulation was performed via ROAST (Huang, Datta, Bikson, & Parra, 2019; https://github.com/andypotatohy/roast). Abbreviations: L: left; R: right.

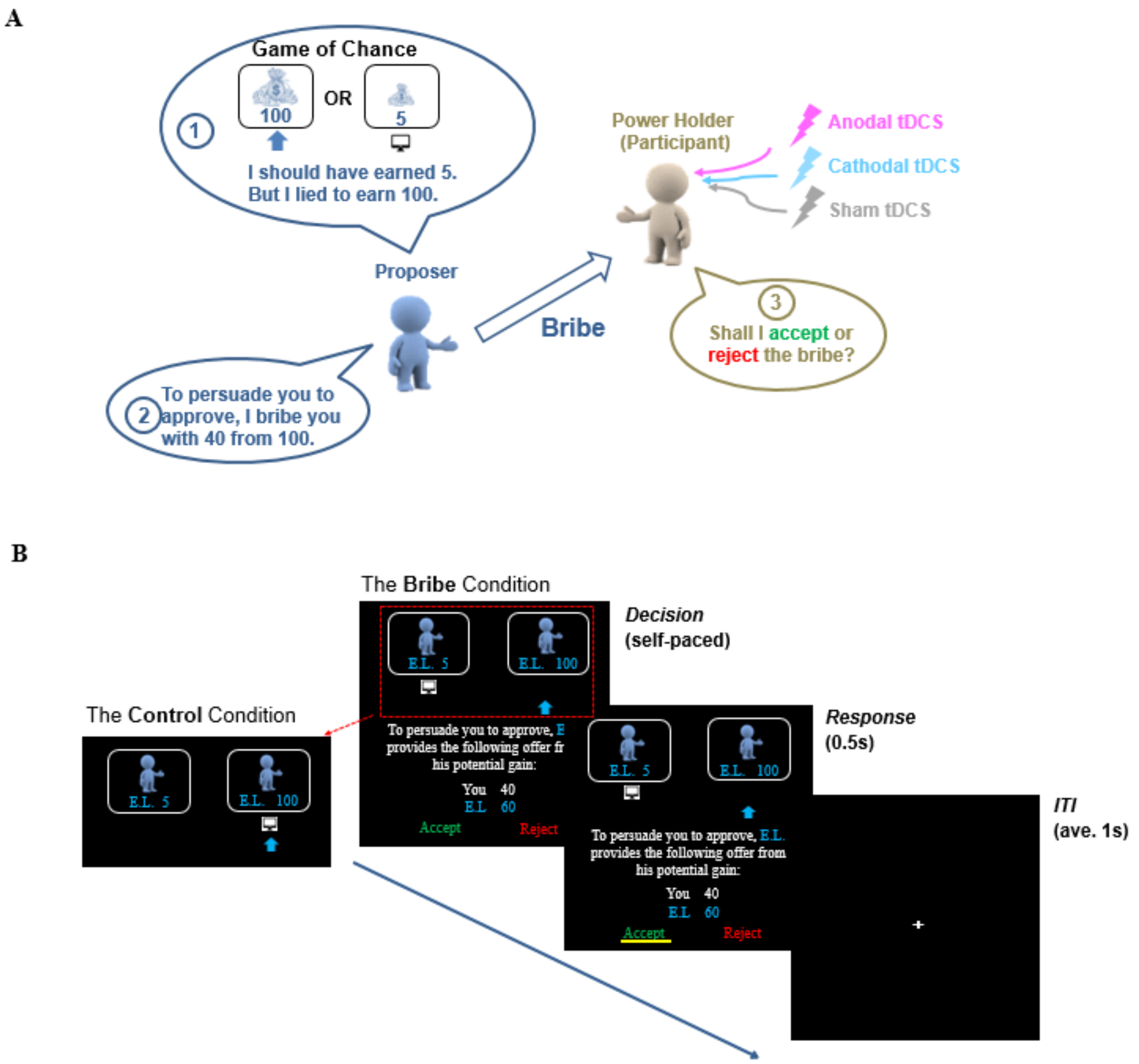

Figure 2 Task design. (A) Schematic illustration of the tDCS manipulation and the behavioral paradigm. All participants were assigned randomly to one of the three tDCS groups (i.e., anodal, cathodal or sham). The task comprised two roles, a proposer (i.e., a fictitious participant in a previous online study where a "Game of Chance" was played) and a power-holder (i.e., the real participants of the current study). In the Control condition, the proposer truthfully reported the larger payoff selected by the computer. In the Bribe condition (as shown here), the proposer lied about the selected larger payoff. In both conditions the proposer offered a certain amount of money to the power-holder, whose task was to decide whether to accept or reject the offer. **(B) Trial procedure.** In this example trial in the Bribe condition, a proposer (E.L.) lied by reporting the non-selected larger payoff (as indicated by the misalignment of the blue arrow and the icon of a computer), and attempted to bribe the power-holder with money from his/her potential gain (i.e., 40 out of 100 Euros). The participant decided whether to accept or reject the offer. Once the decision was made (i.e., accepting the

bribe here), a yellow bar appeared on the corresponding option to highlight the choice for 0.5 s, which was followed by a fixation (i.e., 0.6~1.4 s with a mean of 1s). Trials in the Control condition followed the same procedure except that the proposer truthfully reported the selected larger payoff (as indicated by the alignment of the blue arrow and the icon of a computer).

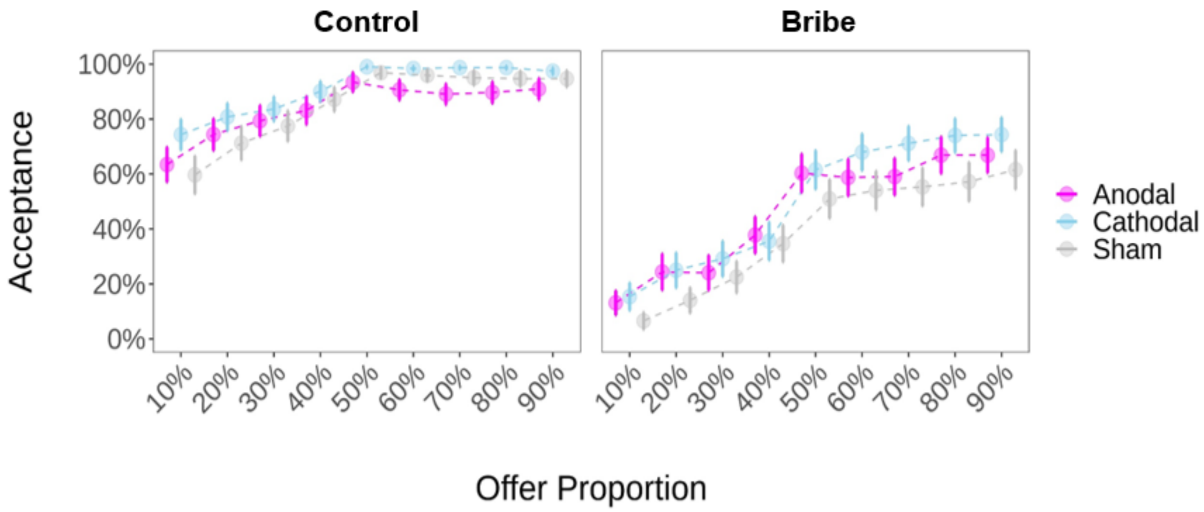

Figure 3. Results of acceptance rate (%). Mean acceptance rate plotted as a function of *tDCS group* (Anodal/Cathodal/Sham), *task condition* (Control/Bribe), and *offer proportion* (10% to 90% in steps of 10%). Error bars represent SEM.

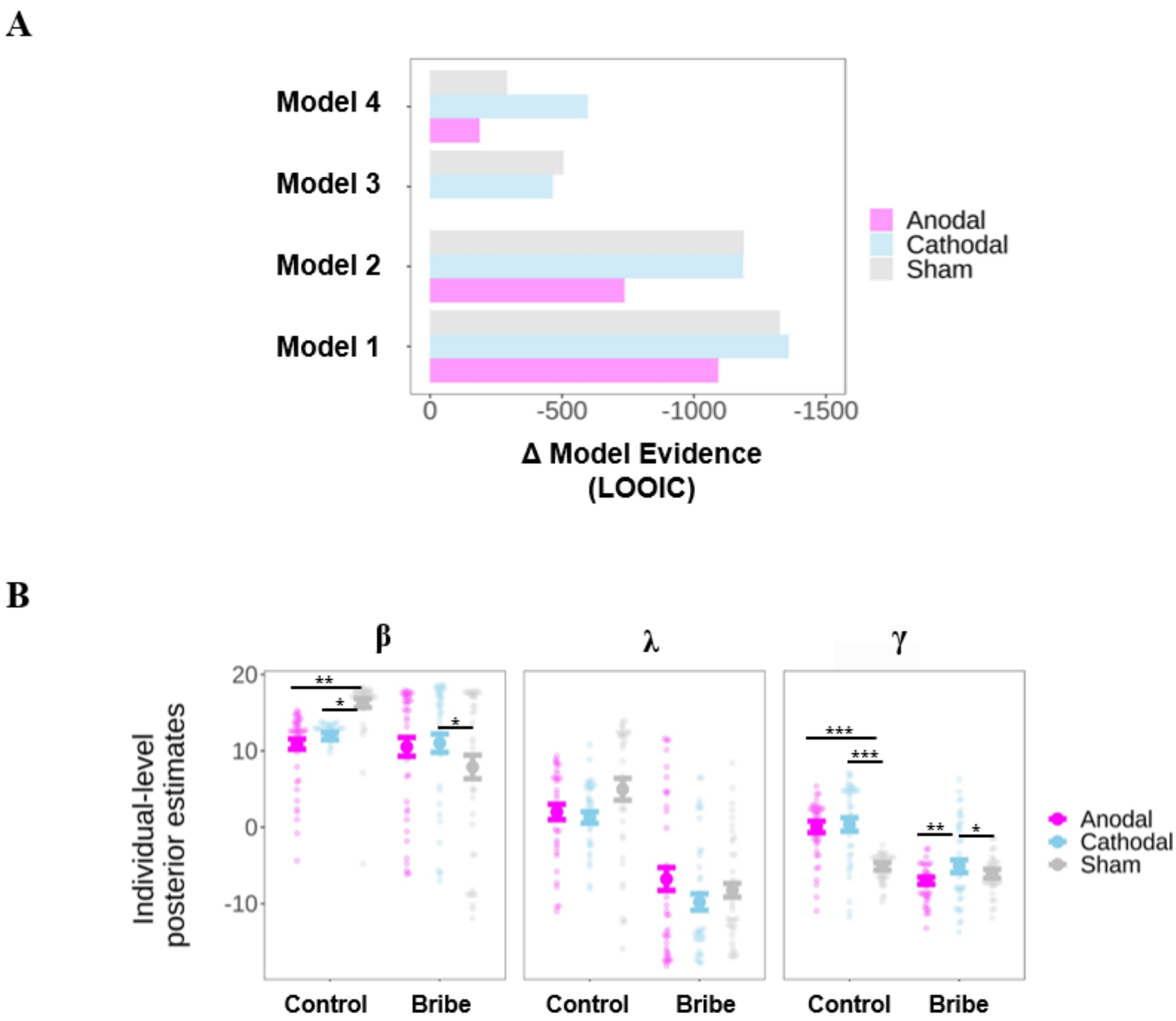

Figure 4. Model-based results. (A) Results of model comparison. Bayesian model evidence for each model was calculated as the difference between its own LOOIC score and that of the model with the worst accuracy of out-of-sample prediction (i.e., Model 2 of the Anodal group in this case). Results clearly favor Model 1 as the winning model across tDCS groups (i.e., more negative difference LOOIC score indicate a better model). Abbreviation: LOOIC = leave-one-out information criterion. **(B) Posterior mean of individual-level key parameters of the new winning model.** β, λ and γ measure the decision weights on personal profits from the proposed offers, the proposer's gain from the offer, and the sensitivity to the absolute payoff inequality between oneself and the proposer respectively. Each large filled dot represents the group-level mean; each smaller filled dot represents the data of a single participant. Error bars represent the SEM; Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, false discovery rate (FDR) corrected.

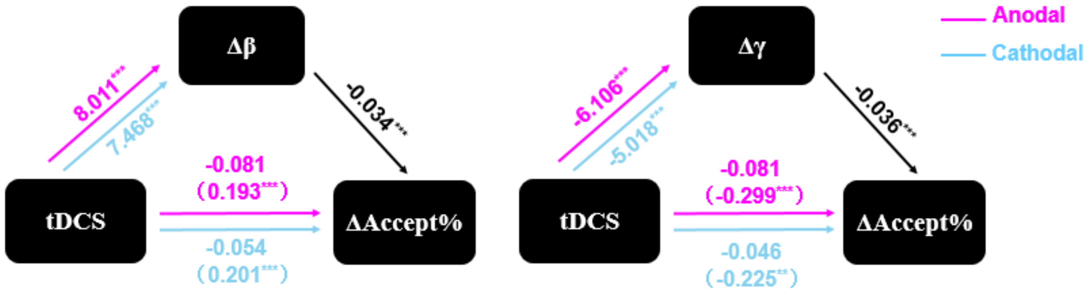

Figure 5. Results of the mediation analysis. Differential parameters (i.e., $\Delta\beta = \beta_{Bribe} - \beta_{Control}$, $\Delta\gamma = \gamma_{Bribe} - \gamma_{Control}$) were found to mediate the impact of the tDCS treatment on the bribery-specific effect on choice behaviors (i.e., $\Delta Accept\% = Accept_{Bribe}\% - Accept_{Control}\%$). Path coefficients are labeled on the arrows. Bootstraps (N = 5,000) were used to test the significance of the indirect effect. Significance: $^{**}p < 0.01$, $^{***}p < 0.001$.

Reference

Abbink, K. (2006). Laboratory experiments on corruption. In S. Rose-Ackerman (Ed.), *International handbook on the economics of corruption* (pp. 418-437).

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neuro-computational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry, 1*, 24-57.

Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: social information processing in the human striatum. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*(1), 61-73.

Carlson, R. W., & Crockett, M. J. (2018). The lateral prefrontal cortex and moral goal pursuit. *Current Opinion in Psychology, 24*, 77-82. doi:10.1016/j.copsyc.2018.09.007

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. *Proceedings of the National Academy of Sciences, 111*(48), 17320-17325.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature neuroscience, 8*(11), 1611.

Dreher, A., Kotsogiannis, C., & McCorriston, S. (2007). Corruption around the world: Evidence from a structural model. *Journal of comparative economics, 35*(3), 443-466.

Fehr, E., & Krajbich, I. (2014). Social preferences and the brain. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 193-218): Elsevier.

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association, 11*(3), 525-547.

Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous-and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences, 115*(33), E7680-E7689.

Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American Economic Review, 108*(2), 419-453.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences, 106*(30), 12506-12511.

Hu, Y., He, L., Zhang, L., Wolk, T., Dreher, J. C., & Weber, B. (2018). Spreading inequality: neural computations underlying paying-it-forward reciprocity. *Soc Cogn Affect Neurosci, 13*(6), 578-589. doi:10.1093/scan/nsy040

Hu, Y., Hu, C., Derrington, E., Corgnet, B., Qu, C., & Dreher, J. C. (2021). Neural basis of corruption in power-holders. *Elife, 10*. doi:10.7554/eLife.63922

Huang, Y., Datta, A., Bikson, M., & Parra, L. C. (2019). Realistic vOlumetric-Approach to Simulate Transcranial Electric Stimulation -- ROAST -- a fully automated open-source pipeline. *Journal of Neural Engineering, 16*(5).

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

Karim, A. A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., & Birbaumer, N. (2010). The Truth about Lying: Inhibition of the Anterior Prefrontal Cortex Improves Deceptive Behavior. *Cerebral cortex, 20*(1), 205-213.

Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., & Fehr, E. (2008). Studying the

Neurobiology of Social Interaction with Transcranial Direct Current Stimulation--The Example of Punishing Unfairness. *Cerebral cortex, 18*(9), 1987-1990.

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.

Köbis, N. C., van Prooijen, J.-W., Righetti, F., & Van Lange, P. A. (2016). Prospection in individual and interpersonal corruption dilemmas. *Review of General Psychology, 20*(1), 71.

López-Alonso, V., Cheeran, B., Río-Rodríguez, D., & Fernández-del-Olmo, M. (2014). Inter-individual Variability in Response to Non-invasive Brain Stimulation Paradigms. *Brain Stimulation, 7*(3), 372-380.

Mameli, F., Mrakic-Sposta, S., Vergari, M., Fumagalli, M., Macis, M., Ferrucci, R., . . . Priori, A. (2010). Dorsolateral prefrontal cortex specifically processes general - but not personal - knowledge deception: Multiple brain networks for lying. *Behavioural Brain Research, 211*(2), 164-168.

Maréchal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with noninvasive brain stimulation. *Proceedings of the National Academy of Sciences, 114*(17), 4360-4364.

Mauro, P. (1995). Corruption and growth. *The quarterly journal of economics, 110*(3), 681-712.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633-644.

Qu, C., Hu, Y., Tang, Z., Derrington, E., & Dreher, J.-C. (2020). Neurocomputational mechanisms underlying immoral decisions benefiting self or others. *Social Cognitive & Affective Neuroence, 15*(2), 135-149.

R Core Team. (2014). R: A language and environment for statistical computing.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9*(7), 545.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science, 342*(6157), 482-484.

Serra, D., & Wantchekon, L. (2012). *New advances in experimental research on corruption* (Vol. 15): Emerald Group Publishing.

Speitel, C., Traut-Mattausch, E., & Jonas, E. (2019). Functions of the right DLPFC and right TPJ in proposers and responders in the ultimatum game. *Soc Cogn Affect Neurosci, 14*(3), 263-270. doi:10.1093/scan/nsz005

Tremblay, S., Lepage, J. F., Latulipe-Loiselle, A., Fregni, F., & Théoret, H. (2014). The Uncertain Outcome of Prefrontal tDCS. *Brain Stimulation, 7*(6), 773-783. doi:10.1016/j.brs.2014.10.003

Wiethoff, S., Hamada, M., & Rothwell, J. C. (2014). Variability in Response to Transcranial Direct Current Stimulation of the Motor Cortex. *Brain Stimulation, 7*(3), 468-475.

Yin, L., & Weber, B. (2018). I lie, why don't you: Neural mechanisms of individual differences in self-serving lying. *Human brain mapping, 40*(4), 1-13. doi:10.1002/hbm.24432

Zhu, L., Jenkins, A. C., Set, E., Scabini, D., Knight, R. T., Chiu, P. H., . . . Hsu, M. (2014). Damage to dorsolateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nature neuroscience, 17*(10), 1319-1321.

Supplementary Materials (SOM) for

Perturbation of Right Dorsolateral Prefrontal Cortex (rDLPFC) Makes Power-Holders Less Resistant to Tempting Bribes

Yang Hu[1,2], Rémi Phillipe[2,3†], Valentin Guigon[2,3†], Sasa Zhao[2,3†], Edmund Derrington[2,3], Brice Corgnet[4], James Bonaiuto[2,3], Jean-Claude Dreher[2,3*]
[1]School of Psychology and Cognitive Science, East China Normal University, Shanghai, China
[2]Neuroeconomics, Reward and Decision Making Laboratory, Institut des Sciences Cognitives Marc Jeannerod, CNRS, France
[3]Université Claude Bernard Lyon 1, Lyon, France
[4]EmLyon, Ecully, France

[*]Correspondence to: dreher@isc.cnrs.fr
[†]These authors equally contributed to this study.

This PDF file includes:

Supplementary Methods
Supplementary Results
Figures S1 to S8
Tables S1 to S8

Supplementary Methods

tDCS Protocol

tDCS was administered using a multichannel stimulator (NeuroConn, Munich) and pairs of standard electrodes covered with conductive paste. Sites of stimulation were fixed through a 10-20 EEG system cap and noted with a marker on the participant's scalp. According to the fairness-related activation foci reported by previous studies (i.e., Talaraich x/y/z: 39/37/22; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Strang et al., 2014), we placed one of the electrodes (5 cm × 7 cm) over AF4 on the 10-20 EEG system for stimulation of the right dorsolateral prefrontal cortex (rDLPFC; see **Figure S1**). The other electrode (10 cm × 10 cm) was placed over Cz (i.e., vertex), based on previous tDCS studies on social decision-making (Maréchal, Cohn, Ugazio, & Ruff, 2017). Following well-established technical guidelines for tDCS studies (Woods et al., 2016), during the experiment we applied stimulation at an intensity of 1.5 mA for up to 30 min in the Anodal and Cathodal groups. To verify that the chosen electrode montage targeted the rDLPFC, we performed current flow simulations using ROAST (Huang, Datta, Bikson, & Parra, 2019) with the MNI152 template brain (see **Figure 1**). For the Sham group, stimulation at the same intensity was set to emit for 1s per minute to simulate the tingling sensations. To minimize the sensations at stimulation onset, the current was linearly ramped up (at the start) and down (at the end) over a period of 20 s.

Procedure

Participants were invited to group sessions with up to 4 in each. Prior to the experiment, participants signed a written informed consent form according to the Declaration of Helsinki. Next, they underwent a clinical screen performed by an experienced neurological doctor in the university hospital, and answered questions from standard health screening questionnaires. Having been confirmed to meet the inclusion criteria for the experiment, they were led to the tDCS room and were randomly placed at seats (desktops), which were separated from each other by shelves. They were then provided with the general instructions and completed the Multidimensional Mood Questionnaire (MDMQ) to report their baseline emotion state. Then, they were given the task instructions, and answered a series of comprehension questions to ensure that they fully understood the task. Meanwhile, two experimenters fitted the participants with the tDCS electrodes. Before the main experiment, participants also practiced a few example trials to get familiar with the paradigm and the response button.

The main experiment included a computerized incentive task (see Task and Design for details) and a follow-up paper-and-pencil rating task, which lasted about 30 min in total. The rating task was aimed to measure the subjective feelings about the task and evaluations of behaviors of either proposers or themselves by means of a Likert scale (0 indicated none, 100 indicated very much). In particular, they indicated the degree of 1) moral inappropriateness of the proposers' behaviors and their decisions (had they accepted offers), 2) moral conflict during the decision period, 3) the guilt they felt (had they accepted offers) in each condition. They also reported the degree to which they had a power advantage over proposers and whether they perceived offers from the proposers as bribes.

Once all participants in the session were prepared, the experimenter started the tDCS stimulation for 45s and then commenced the incentive task. To further protect their privacy, curtains behind the participants' seats were drawn during the whole experiment. The tDCS was maintained until participants in the session finished the main experiment. After that, they took a short break and then filled out a battery of questionnaires for control measures. In particular, they indicated whether they felt comfortable after the stimulation, declared their belief about treatment (stimulation, placebo, or unknown), reported their emotional state again by filling out the Multidimensional Mood Questionnaire (Steyer, 2014), and finished a Cognitive Reflection Test as a measure of their cognitive reflection ability (Frederick, 2005). Finally, participants were debriefed on all task-relevant information, and informed about their final payoffs.

Data Analyses

Model-free analyses

All analyses and visualization were conducted using R (http://www.r-project.org/) and relevant packages (R Core Team, 2014). All reported p values are two-tailed and $p < 0.05$ was considered statistically significant. For choice data, we performed repeated measures mixed-effect logistic regression on the decision of choosing the "accept" option, using the *glmer* function in the "lme4" package (Bates, Maechler, & Bolker, 2013), with *tDCS group* (dummy variable; reference level: Sham), *task condition* (dummy variable; reference level: Control), *offer proportion* (continuous variable), and their interactions as fixed-effects of interest. The effect of the larger payoff the proposer would earn in the reported option (continuous variable; z-scored) was also incorporated as a fixed-effect covariate. The random-effects were established using a "maximal" principle such that we allowed intercepts and slopes (i.e., task condition, offer proportion and their interaction) to vary across participants (Barr, Levy, Scheepers, & Tily, 2013). For statistical inference on each fixed effect, we performed a Type II Wald chi-square test on the model fits by using the *Anova* function in the "car" package (Fox et al., 2016).

For decision time (DT), we first log-transformed the data, because of its non-normal distribution (i.e., Anderson-Darling normality test: A = 1411.1, p < 0.001) and then performed a mixed-effect linear regression on the log-transformed DT using the *lmer* function in the "lme4" package. Random-effect predictors were specified in the same way as above. When a model failed to converge, we dropped one or more of the random slopes until the estimation converged. We followed the procedure recommended by Luke (2017) to obtain the statistics of each predictor by applying the Satterthwaite approximations on the restricted maximum likelihood model (REML) fit via the "lmerTest" package (Luke, 2017). We performed post-hoc analyses of interaction effects using *emtrends* function of the "emmeans" package. For subjective rating, we used mixed analysis of variance (ANOVA) or simple linear regression analyses depending on specific items (see Results for details). Furthermore, we reported the odds ratio as an index of effect size of each predictor on choice. For decision time and other continuous dependent measures (e.g., rating, parameter estimates), we computed the standardized coefficient (b_z) as an index of effect size using the "lm.beta" package (https://cran.r-project.org/web/packages/lm.beta/). We also used *partial* η^2 via the "sjstats" package (https://cran.r-project.org/web/packages/sjstats/) to indicate the effect size of main effects or interactions in ANOVA or mixed-effect regression analyses when applicable.

Computational Modelling

We adopted a basic social preference model that has been used in a modified Dictator Game, i.e., a task of splitting money between oneself and a partner (Tusche & Hutcherson, 2018). Specifically, this model assumes that the participant, in the role of the power-holder, is supposed to pit the personal profit against the proposer's gain as well as their payoff inequity. In our task, the only difference between the Bribe and Control condition is whether a moral transgression of colluding with a fraudulent proposer is involved in the decision-making process. Hence, bribery-related decision making would additionally bring in a moral cost that might prevent the power-holder from taking the bribe. Based on our previous fMRI study using a similar paradigm (Hu et al., 2021), we clearly hypothesized that there would be a moral cost on the personal profit from the bribe. In addition, we explored whether such moral cost also impacts the other components (i.e., the proposer's payoff and the absolute payoff inequality) involved in the trade-off during bribery-related decision-making, which remains an open question. Thus, the utility function can be written as follows:

$$SV(P_{PH}, P_P) = \beta P_{PH} + \lambda P_P + \gamma |P_P - P_{PH}|$$

$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \lambda_{Control}, \gamma_{Control}, if\ Control\ condition \\ \beta_{Bribe}, \lambda_{Bribe}, \gamma_{Bribe}, if\ Bribe\ condition \end{cases} \quad \text{Model 1}$$

In this model, SV denotes the subjective value of the choice, P_P and P_{PH} represent the offer's payoff (i.e., monetary gain) for the proposer and power-holder given the different choices (i.e., accepting or rejecting the offer; same below). Regarding the free parameters, β measures the decision weights on personal profits from the offer, λ measures the decision weights on the proposer's gain from the offer, and γ measures the sensitivity to the absolute payoff inequality between oneself and the proposer ($-20 \leq \beta, \lambda, \gamma \leq 20$). All these parameters were expected to vary across the two conditions.

To examine whether this model fits the data best, we also established several candidate models. Model 2 and Model 3 are similar to Model 1, except that participants take into account neither the absolute payoff inequality nor the proposer's gain respectively.

$$SV(P_{PH}, P_P) = \beta P_{PH} + \lambda P_P$$

$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \lambda_{Control}, if\ Control\ condition \\ \beta_{Bribe}, \lambda_{Bribe}, if\ Bribe\ condition \end{cases} \quad \text{Model 2}$$

$$SV(P_{PH}, P_P) = \beta P_{PH} + \gamma |P_P - P_{PH}|$$

$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \gamma_{Control}, if\ Control\ condition \\ \beta_{Bribe}, \gamma_{Bribe}, if\ Bribe\ condition \end{cases} \quad \text{Model 3}$$

In addition, we also adopted the Fehr-Schmidt model which assumes disparate degrees of inequity aversion depending on whether one person earns more or less than the other, defined as follows:

$$SV(P_{PH}, P_P) = P_{PH} - \alpha \max(P_P - P_{PH}, 0) - \beta \max(P_{PH} - P_P, 0)$$

$$\alpha, \beta = \begin{cases} \alpha_{Control}, \beta_{Control}, if\ Control\ condition \\ \alpha_{Control}, \beta_{Bribe}, if\ Bribe\ condition \end{cases} \quad \text{Model 4}$$

α and β measure the degree of aversion to payoff inequality in disadvantageous and advantageous situations respectively. In other words, these parameters capture how much a participant dislikes the offer when they earn less (measured by α) or more (measured by β) than the proposer in two conditions respectively ($0 \leq \alpha, \beta \leq 20$).

The probability of accepting the offer was determined by the *softmax* function:

$$p(accept) = \frac{e^{\tau SV_{accept}}}{e^{\tau SV_{accept}} + e^{\tau SV_{reject}}} = \frac{1}{1 + e^{-\tau(SV_{accept} - SV_{reject})}}$$

where SV denotes the subjective value (of accepting or rejecting the offer), calculated by the model mentioned earlier. τ is the inverse softmax temperature parameter ($0 \leq \tau \leq 10$) denoting the sensitivity of an individual's decision to the difference in SV between the choice of accepting versus rejecting the offer.

The above model was fit using a hierarchical Bayesian approach (HBA) via the "hBayesDM" package (Ahn, Haines, & Zhang, 2017), which adopts a Markov Chain Monte Carlo (MCMC) sampling scheme to perform full Bayesian inference. We chose HBA because it has been shown to provide much more stable and accurate estimates than other estimation approaches (e. g., maximum likelihood estimation; Ahn, Krawitz, Kim, Busemeyer, & Brown, 2011). Convergence of the MCMC chains was assessed through the Gelman-Rubin R-hat Statistics (Gelman & Rubin, 1992). Here, R-hat values of all estimated parameters of each tDCS group for all models were smaller than 1.02, indicating adequate convergence of the MCMC chains.

For model comparisons, we adopted the leave-one-out information criterion (LOOIC) as the index for model evidence. Compared with other point estimate information criteria (e.g., Akaike information criterion, AIC), LOOIC score can be more reliable by providing the estimate of out-of-sample predictive accuracy in a fully Bayesian way (Vehtari, Gelman, & Gabry, 2017). Conventionally, the lower LOOIC score indicates better out-of-sample prediction accuracy of the candidate model. A difference score of 10 on the information criterion scale is considered decisive (Burnham & Anderson, 2004). We selected the model with the lowest LOOIC for all tDCS groups as the winning model for subsequent analysis of key parameters. We also performed the posterior predictive check (PPC) both at the individual and group level following the procedure suggested by Zhang *et al* (2020) and used by our previous studies (Hu et al., 2021; Qu, Hu, Tang, Derrington, & Dreher, 2020) to examine whether the prediction of the model could capture the features of real behaviors of participants.

For each individual, we obtained the posterior mean of individual-level key parameters of the winning model for each condition (i.e., β, λ, γ of Model 1). To examine how bribery-elicited moral cost affect each parameter and how tDCS treatment modulated such effects, we implemented mixed-effect linear regression on each parameter separately, by including *tDCS group*, *task condition*, and their interactions as the fixed-effect predictors. We also allowed intercepts to vary across participants as the random effects. For further analyses and illustration purpose, the individual-level differential parameters between the Bribe and Control condition were also calculated to characterize the bribery-specific effect (i.e., $\Delta\beta = \beta_{Bribe} - \beta_{Control}$, $\Delta\lambda = \lambda_{Bribe} - \lambda_{Control}$, $\Delta\gamma = \gamma_{Bribe} - \gamma_{Control}$; same below; see **Figure S8**). To further establish the link between the tDCS treatment, the bribery-elicited moral cost on these parameters, and the choice behaviors, we implemented post-hoc mediation analyses using the *bootM* package with tDCS group as the predictor, the differential parameters as the mediator, and the differential acceptance rate (i.e., ΔAccept = $\text{Accept}_{\text{Bribe}}$ – $\text{Accept}_{\text{Control}}$) as the dependent variable. Statistical inference was confirmed by using a bootstrapping procedure to test the mediation effect (i.e., 5000 bootstraps).

Supplementary Results

No tDCS effect was observed in other behavioral measures

We investigated whether a similar effect of tDCS over rDLPFC existed in other behavioral measures. Analyses on log-transformed DT revealed that participants responded slightly slower in the Bribe condition (vs. Control; a main effect of task condition: $F_{(1,325)} = 5.97$, $p < 0.001$, partial-$\eta^2 < 0.001$) and more quickly when the offer proportion increased (a main effect of offer proportion: $F_{(1,17012)} = 67.03$, $p < 0.001$, partial-$\eta^2 = 0.004$). In addition, we observed a two-way interaction between *task condition* and *offer proportion* ($F_{(1,16937)} = 16.59$, $p < 0.001$, partial-$\eta^2 = 0.001$; see **Figure S2**). *Post-hoc* analyses indicated that participants responded faster when the offer proportion increased in both conditions (zs < -3.15, ps < 0.002) but the slope was less steep in the Bribe condition (vs. Control; $z = 4.07$, $p < 0.001$; see **Table S5** for details of the regression output).

In addition, we also examined whether tDCS over rDLPFC affected subjective ratings, in order to rule out alternative accounts that might explain the effect of tDCS on bribe-taking behaviors. First, compared with the Control condition, participants in the Bribe condition felt a higher level of moral conflict during the decision period ($F_{(1,116)} = 103.50$, $p < 0.001$, *partial-*η^2 = 0.157). They thought that the proposer's offering act ($F_{(1,116)} = 21.65$, $p < 0.001$, *partial-*η^2 = 0.472) and their hypothetical acceptance were more morally inappropriate ($F_{(1,115)} = 157.73$, $p < 0.001$, *partial-*η^2 = 0.578). They also felt more guilty for their hypothetical acceptances of offers provided by the proposer ($F_{(1,115)} = 101.64$, $p < 0.001$, *partial-*η^2 = 0.469). However, none of these measures were modulated by tDCS (Fs < 1.01, ps > 0.36, *partial-*η^2s < 0.02) nor its interaction with task conditions (Fs < 1.34, ps > 0.26, *partial-*η^2s < 0.03). Second, participants from the three tDCS groups reported similar levels of the sense of power over the proposer ($F_{(2,116)} = 0.52$, $p = 0.597$, *partial-*η^2 = 0.009) and the sense of being bribed ($F_{(2,116)} = 1.04$, $p = 0.357$, *partial-*η^2 = 0.018).

Regarding task-irrelevant measures, no difference between the three tDCS groups was found in emotional state, as measured by the Multidimensional Mood Questionnaire (MDMQ) (Steyer, 2014), reported before the main task (the awake-tired [AT] subscale: $F_{(2,115)} = 0.85$, $p = 0.429$, *partial-*η^2 = 0.015; the calm-nervous [CN] subscale: $F_{(2,114)} = 0.22$, $p = 0.804$, *partial-*η^2 = 0.004; the good-bad [GB] subscale: $F_{(2,115)} = 0.44$, $p = 0.645$, *partial-*η^2 = 0.008) or after (AT: $F_{(2,116)} = 0.39$, $p = 0.677$, *partial-*η^2 = 0.007; CN: $F_{(2,116)} = 1.18$, $p = 0.312$, *partial-*η^2 = 0.020; GB: $F_{(2,116)} = 0.95$, $p = 0.389$, *partial-*η^2 = 0.016). Cognitive reflection ability, as measured by the Cognitive Reflection Test (Frederick, 2005), was unaffected by the tDCS manipulation ($\chi^2_{(4)} = 5.28$, $p = 0.260$; see **Table S6** and **S7** for a descriptive summary of these measures).

Inverse temperature did not influence the tDCS effect on choice behavior and key parameters in the winning model

As the inverse temperature parameter (τ) varied between tDCS groups ($F_{(2, 116)}$ = 4.67, p = 0.019, *partial*-η^2 = 0.08; see **Table S4** for the descriptive summary), we performed control analyses on the choice behavior and key parameters (i.e., β and γ) by including τ as a between-group covariate to rule out the confounding effect of τ,. Results showed that the main findings related with the tDCS effect on behaviors (tDCS Group ×Condition × Offer Proportion three-way interaction: $\chi^2_{(2)}$ = 7.93, p = 0.019) and key parameters (tDCS Group ×Condition two-way interaction: β: $F_{(2, 116)}$ = 11.71, p < 0.001, *partial*-η^2 = 0.12; γ: $F_{(2, 116)}$ = 16.14, p < 0.001, *partial*-η^2 = 0.14) still held after we took the effect of τ into account (see **Table R4** for complete regression outputs).These findings indicated that the inverse temperature might not well explained the tDCS effect on behaviors and its underlying computations.

Supplementary Figures

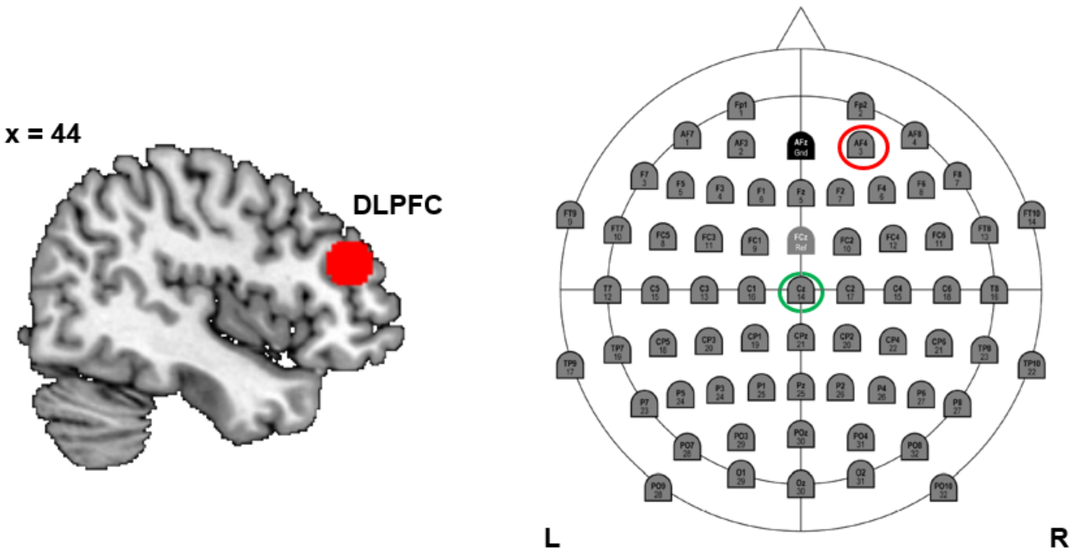

Figure S1. Display of the tDCS electrode localization. Based on previous literature highly relevant to the current study (Knoch *et al.*, 2006; Strang *et al.*, 2014), we chose the position centering around the MNI coordinate of 39/37/22 as our target site (the left panel; a sphere of a 10mm radius was used for visualization). This location approximately corresponds to the electrode position of AF4 in the 10-20 system of 64-channel EEG cap (the right panel; marked with a red circle). The vertex was chosen as the reference electrode based on the study by Marechal *et al* (2017), which corresponds to the electrode position of Cz (the right panel; marked with a green circle).

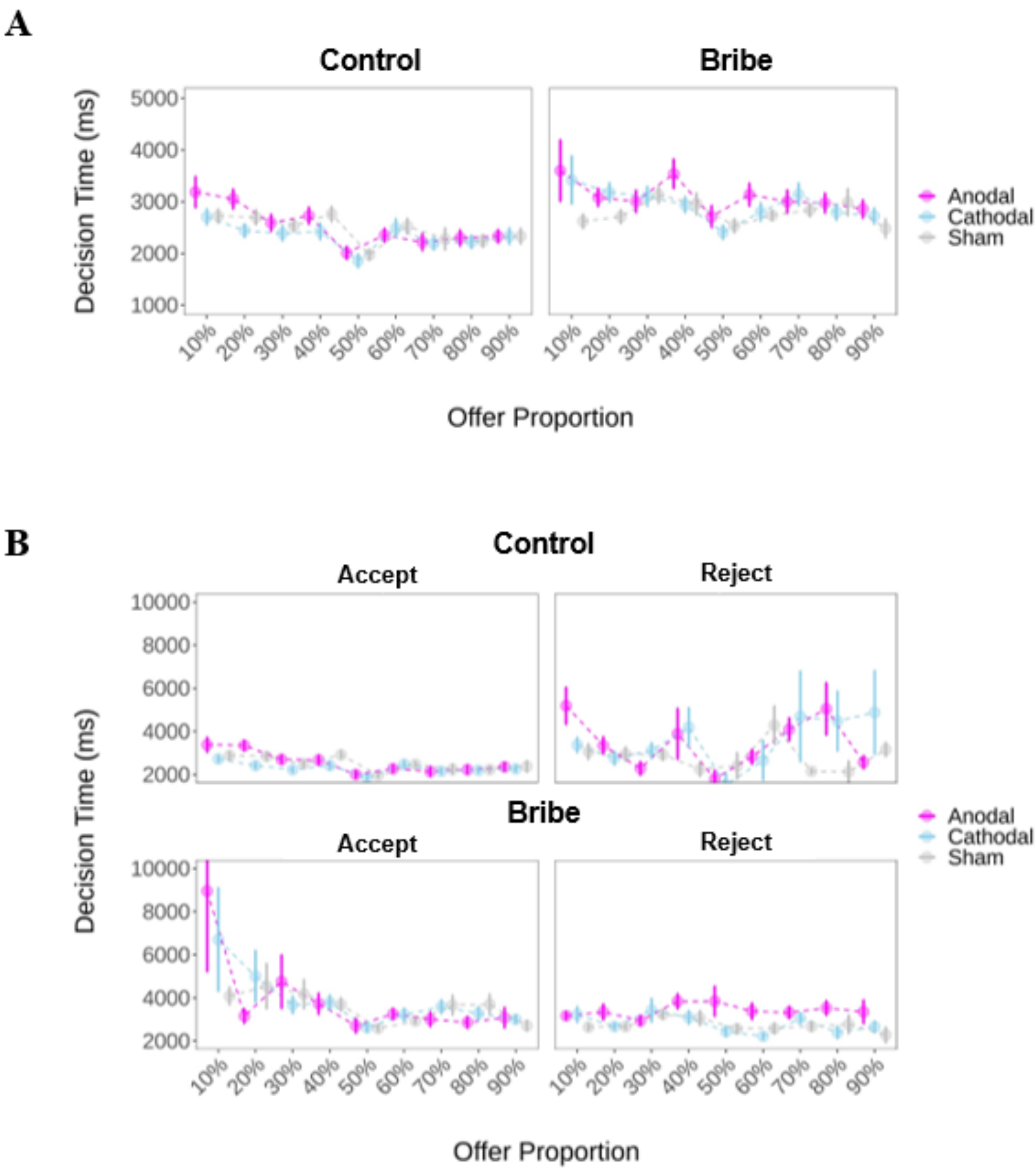

Figure S2. Results of decision time (DT; ms). (A) Mean DT are plotted as a function of tDCS group (Anodal/Cathodal/Sham), task condition (Control/Bribe), and offer proportion (10% to 90% in a step of 10%). (B) Mean DT are plotted as a function of these independent variables for acceptance trials and rejections trials respectively. Error bars represent SEM.

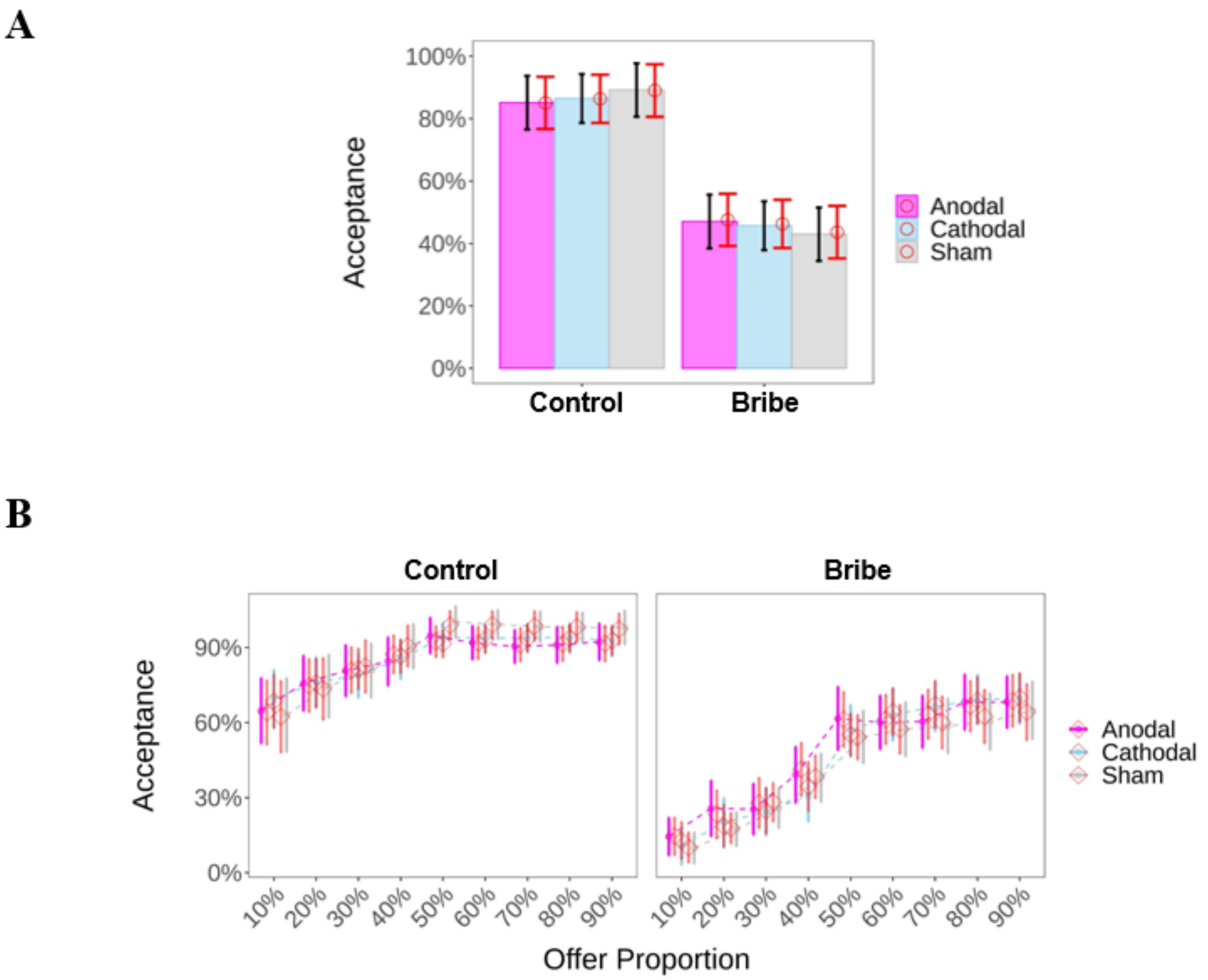

Figure S3. Posterior predictive check at the group level. (A) Mean predicted (red circles) and actual acceptance rates (histogram bars) plotted as a function of tDCS treatment, and task condition. (B) Mean predicted (red circles) and actual acceptance rates (filled dots; connected by dashed lines) plotted as a function of tDCS treatment, task condition, and offer proportion. Error bars represent 95% CI.

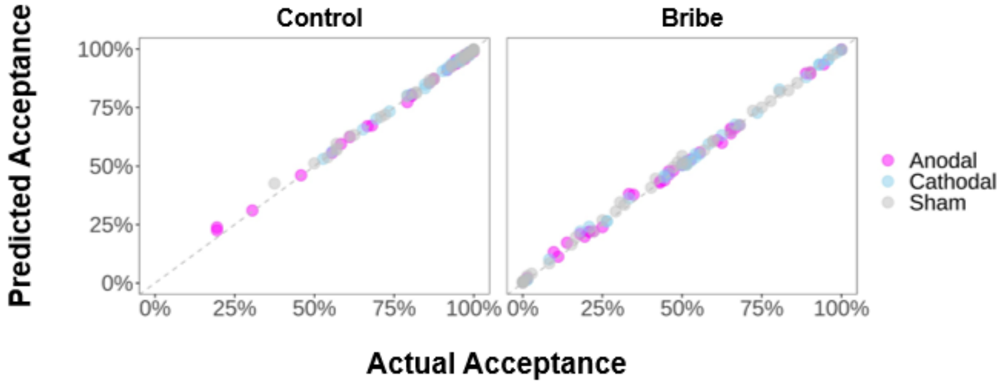

Figure S4. Posterior predictive check at the individual level. Relationship between predicted acceptance rates and actual acceptance rates across individuals. Filled dots represent individual data. Error bars represent 95% CI.

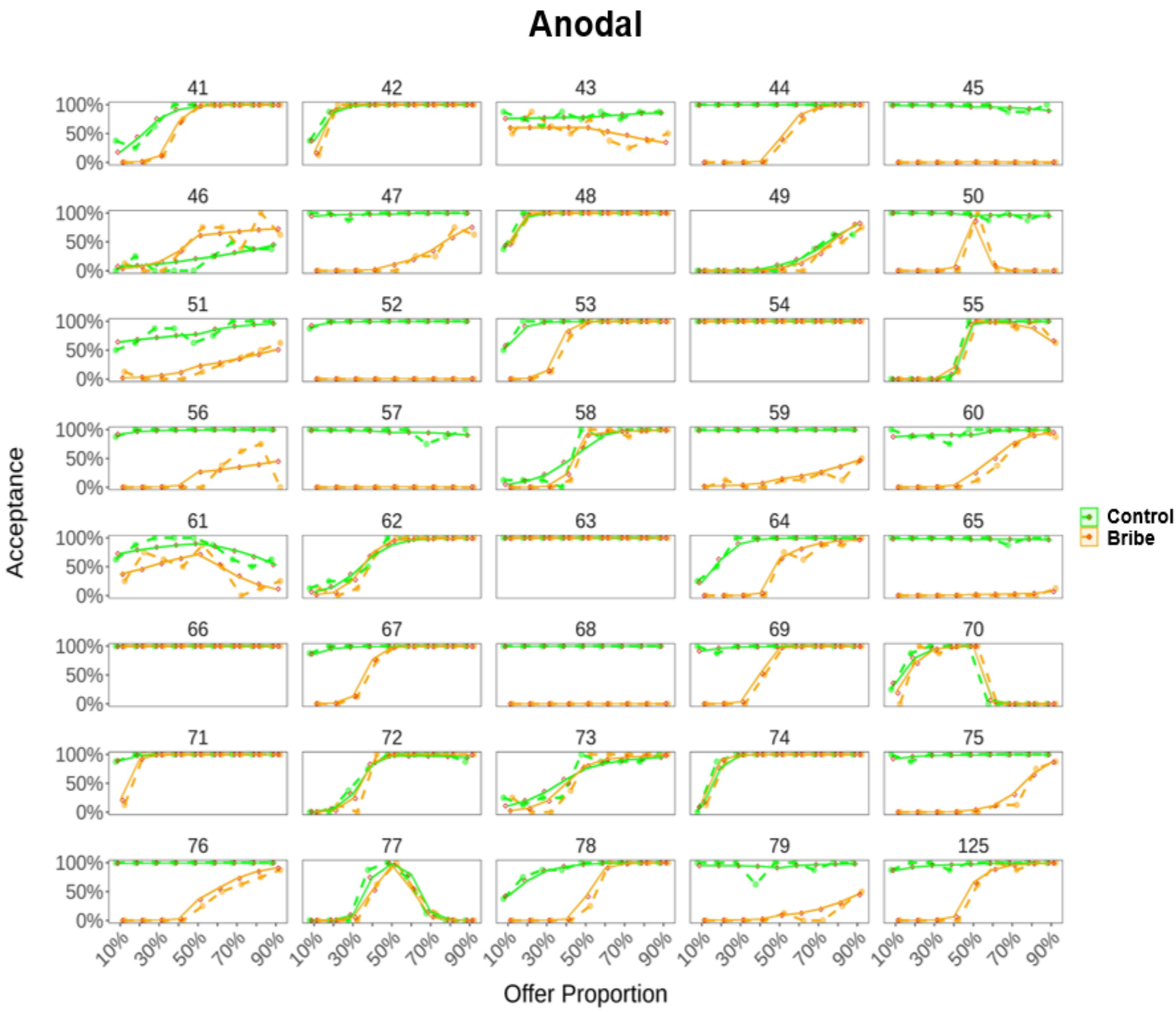

Figure S5. Posterior predictive check at the individual level for the Anodal group. Mean predicted (red circles; connected by solid lines) and actual acceptance rates (filled dots; connected by dashed lines) plotted as a function of task condition and offer proportion across individuals in the Anodal group. Numbers refer to subject ID. Solid lines that are actually shaded areas represent 95% CI based on 4000 posterior samples.

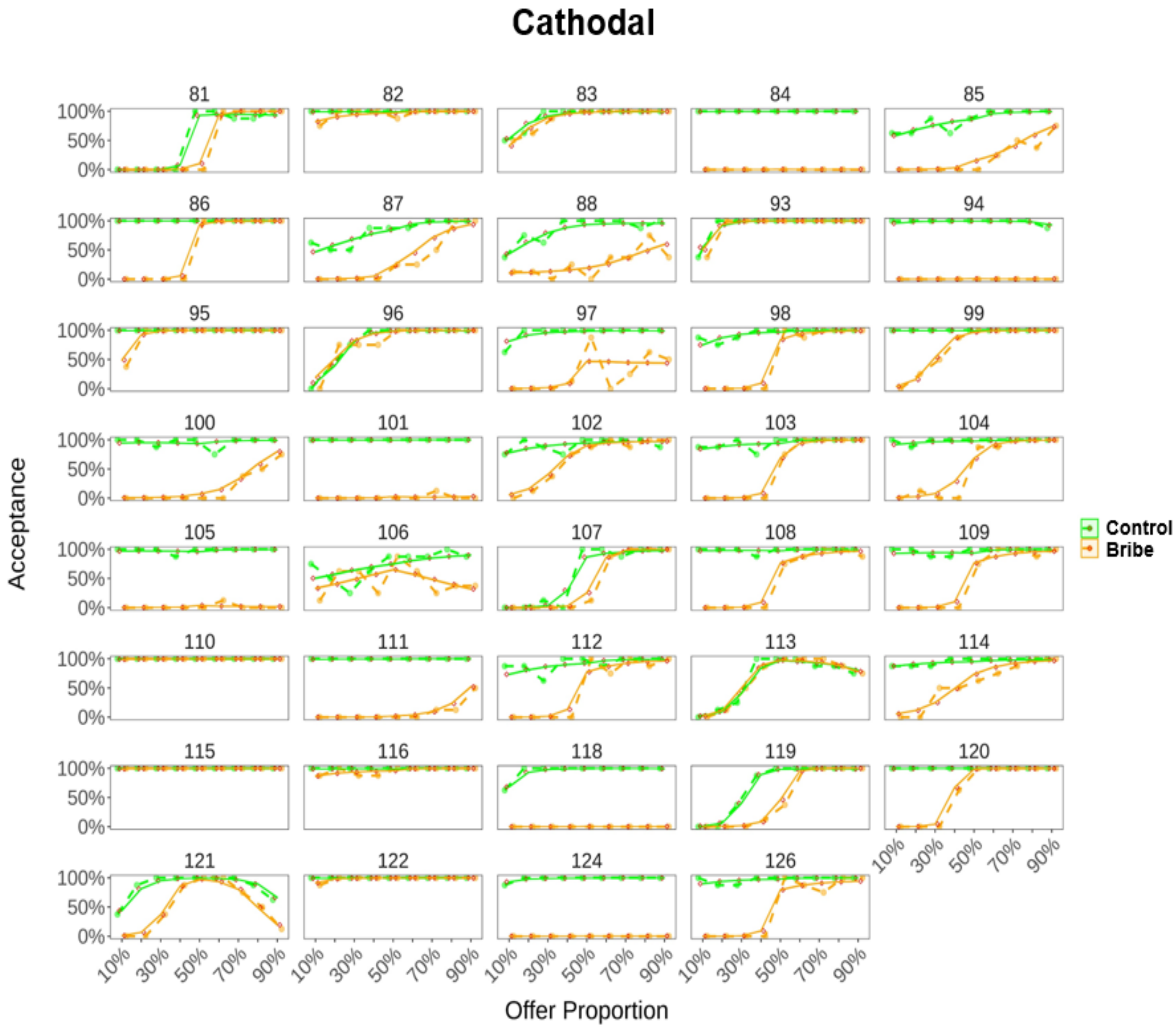

Figure S6. Posterior predictive check at the individual level for the Cathodal group. Mean predicted (red circles; connected by solid lines) and actual acceptance rates (filled dots; connected by dashed lines) plotted as a function of task condition and offer proportion across individuals in the Cathodal group. Numbers refer to subject ID. Solid lines that are actually shaded areas represent 95% CI based on 4000 posterior samples.

825

826 **Figure S7. Posterior predictive check at the individual level for the Sham group.** Mean
827 predicted (red circles; connected by solid lines) and actual acceptance rates (filled dots;
828 connected by dashed lines) plotted as a function of task condition and offer proportion
829 across individuals in the Sham group. Numbers refer to subject ID. Solid lines that are
830 actually shaded areas represent 95% CI based on 4000 posterior samples.

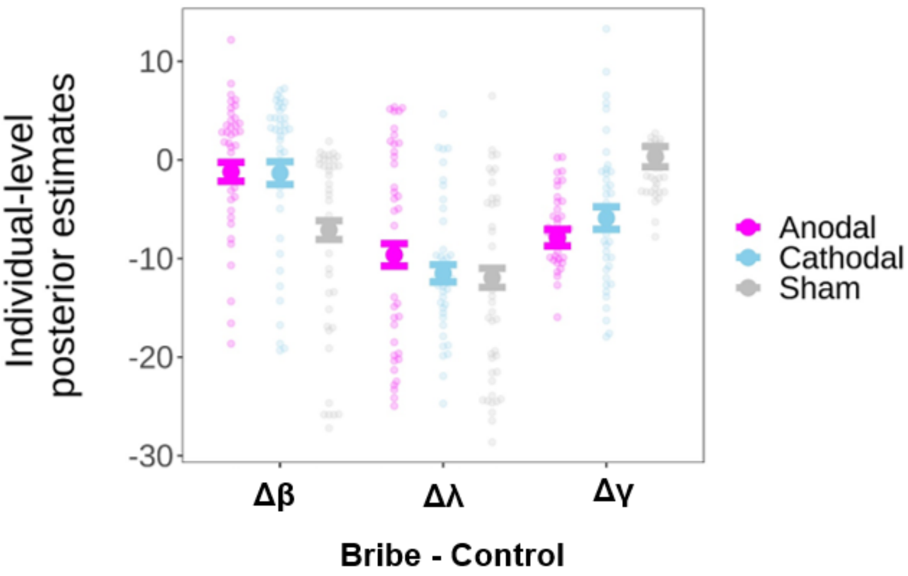

Figure S8. The tDCS effect on differential parameters of the winning model. This is another way to illustrate the interaction effect on key parameters. Differential parameters are calculated as follows: $\Delta\beta = \beta_{Bribe} - \beta_{Control}$, $\Delta\lambda = \lambda_{Bribe} - \lambda_{Control}$, $\Delta\gamma = \gamma_{Bribe} - \gamma_{Control}$. Each large filled dot represents the group-level mean; each smaller filled dot represents the data of a single participant. Error bars represent the SEM; Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Supplementary Tables

Table S1 Results of mixed-effect logistic regressions predicting acceptance

	All	Control	Bribe
	b (SE)	*b* (SE)	*b* (SE)
Intercept	0.25 (0.80)	0.23 (0.88)	-6.58*** (0.83)
tDCS (Anodal)	0.72 (1.12)	0.67 (1.20)	0.44 (1.17)
tDCS (Cathodal)	1.49 (1.14)	1.64 (1.23)	0.14 (1.18)
Condition	-6.79*** (1.03)		
Offer Proportion	10.47*** (1.58)	10.26*** (1.78)	11.51*** (1.87)
tDCS (Anodal) × Condition	-0.23 (1.43)		
tDCS (Cathodal) × Condition	-1.29 (1.45)		
tDCS (Anodal) × Offer Proportion	-3.22 (2.17)	-3.19 (2.25)	1.90 (2.65)
tDCS (Cathodal) × Offer Proportion	-2.86 (2.22)	-3.11 (2.30)	2.37 (2.66)
Condition × Offer Proportion	1.06 (1.57)		
tDCS (Anodal) × Condition × Offer Proportion	5.33* (2.08)		
tDCS (Cathodal) × Condition × Offer Proportion	5.20* (2.13)		
Larger payoff for proposer in the reported option[a]	0.29*** (0.03)	0.18*** (0.05)	0.37*** (0.04)
AIC	7400.6	3211.6	4243.8
BIC	7578.8	3282.2	4314.4
N (Observation)	17136	8568	8568
N (Participant)	119	119	119

Note: [a] This variable was standardized before the analyses. Reference levels in dummy variables were set as follows: tDCS Group = Sham, Condition = Control. Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table S2 Results of mixed-effect linear regressions predicting decision time (DT)

	All	Control[b]	Bribe[b]
	b (SE)	*b* (SE)	*b* (SE)
Intercept	7.58*** (0.08)	7.56*** (0.08)	7.69*** (0.09)
tDCS (Anodal)	0.03 (0.12)	-0.005 (0.11)	0.06 (0.12)
tDCS (Cathodal)	-0.04 (0.12)	-0.03 (0.11)	0.07 (0.12)
Condition	0.04 (0.06)		
Offer Proportion	-0.22*** (0.05)	-0.21*** (0.03)	-0.15*** (0.03)
Decision	0.03 (0.02)	0.14*** (0.02)	-0.05* (0.02)
tDCS (Anoda) × Condition	0.01 (0.08)		
tDCS (Cathodal) × Condition	0.11 (0.08)		
tDCS (Anodal) × Offer Proportion	-0.07 (0.06)		
tDCS (Cathodal) × Offer Proportion	-0.01 (0.06)		
Condition × Offer Proportion	0.11† (0.06)		
tDCS (Anodal) × Condition × Offer Proportion	0.11 (0.09)		
tDCS (Cathodal) × Condition × Offer Proportion	0.01 (0.09)		
Larger payoff for proposer in the reported option[a]	-0.01** (0.005)	-0.01 (0.007)	-0.02** (0.007)
AIC	33637.4	16653.2	17095.3
BIC	33776.9	16709.6	17151.7
N (Observation)	17136	8568	8568
N (Participant)	119	119	119

Note: [a] This variable was standardized before the analyses.

[b] We did not incorporate interactions between tDCS Group and offer proportion, as none of these effects was significant in the regression using all trials. Reference levels in dummy variables were set as follows: tDCS Group = Sham, Condition = Control, Decision = acceptance. Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table S3 Descriptive statistics of task-relevant subjective rating

		Anodal (N = 40)	Cathodal (N = 39)	Sham (N = 40)
Perceived as bribe		68.6 ± 31.4	67.6 ± 27.4	76.1 ± 27.4
Sense of Power		71.6 ± 30.9	77.9 ± 27.2	72.8 ± 29.1
Moral conflict	Bribe	42.2 ± 29.0	41.1 ± 31.8	36.9 ± 31.3
	Control	14.5 ± 22.1	6.3 ± 13.2	13.3 ± 24.0
Guilt[a]	Bribe	44.2 ± 32.8	48.0 ± 36.7	48.2 ± 37.7
	Control	14.2 ± 22.8	8.7 ± 17.3	11.8 ± 22.4
Moral Inappropriateness: Self[a]	Bribe	56.7 ± 33.8	54.7 ± 34.6	60.8 ± 33.4
	Control	11.6 ± 21.0	13.9 ± 23.0	16.5 ± 25.8
Moral Inappropriateness: Proposer	Bribe	56.4 ± 34.0	51.3 ± 33.2	54.0 ± 33.6
	Control	25.0 ± 31.9	30.6 ± 36.6	39.5 ± 33.5

Note: [a] Ratings of these items in the Bribe condition from one participants in the Cathodal group was missing. Thus we dropped this participant for analyses on these two items.

Table S4 Descriptive statistics of other measures

		Anodal (N = 40)	Cathodal (N = 39)	Sham (N = 40)
MDMQ: pre-task	AT[a]	35.2 ± 6.6	33.8 ± 6.5	35.5 ± 5.7
	CN[a,b]	39.4 ± 6.9	39.3 ± 6.7	40.2 ± 5.8
	GB[a]	39.0 ± 5.0	40.4 ± 8.9	39.8 ± 4.9
MDMQ: post-task	AT	31.9 ± 7.5	30.4 ± 6.3	31.4 ± 7.8
	CN	37.3 ± 7.5	38.1 ± 6.1	39.5 ± 5.9
	GB	36.4 ± 5.9	37.0 ± 5.6	38.1 ± 5.7
CRT		0.9 ± 0.8	1.1 ± 0.9	0.8 ± 0.8

Note: [a]Data of the pre-task MDMQ measures from one participant in the Cathodal group was missing [b]Data of pre-task MDMQ measures (only in CN subscale) from one participant in the Sham group was missing.

Abbreviations: MDMQ: multidimensional mood questionnaire; subscales: AT: awake-tired, CN: calm-nervous, GB: good-bad; CRT: cognitive reflection ability.

Table S5 Descriptive statistics of posterior mean of individual-level key parameters in the winning model

		Anodal (N = 40)	Cathodal (N = 39)	Sham (N = 40)
β (mean ± SD)	Control	10.50 ± 4.93	12.56 ± 0.91	16.04 ± 3.99
	Bribe	10.13 ± 8.25	11.66 ± 8.27	7.66 ± 10.67
λ (mean ± SD)	Control	1.61 ± 5.72	1.92 ± 4.36	4.75 ± 8.60
	Bribe	-7.17 ± 9.95	-9.15 ± 7.73	-8.47 ± 6.92
γ (mean ± SD)	Control	-0.35 ± 3.84	1.01 ± 5.28	-5.35 ± 1.81
	Bribe	-7.40 ± 2.44	-4.46 ± 5.43	-6.29 ± 2.31
τ (mean ± SD)		0.013 ± 0.008	0.010 ± 0.004	0.010 ± 0.004

Table S6 Results of linear regressions predicting parameters in the winning model

	β	λ	γ
	b (SE)	*b* (SE)	*b* (SE)
Intercept	16.04*** (1.10)	4.75*** (1.18)	-5.35*** (0.60)
tDCS (Anodal)	-5.54*** (1.56)	-3.15 (1.67)	5.00*** (0.85)
tDCS (Cathodal)	-3.47* (1.57)	-2.84 (1.68)	6.36*** (0.85)
Condition	-8.38*** (1.31)	-13.22***(1.45)	-0.94 (0.79)
tDCS (Anodal) × Condition	8.01*** (1.85)	4.44* (2.05)	-6.11*** (1.11)
tDCS (Cathodal) × Condition	7.47*** (1.86)	2.15 (2.06)	-4.52*** (1.12)
AIC	1586.9	1621.2	1312.1
BIC	1614.7	1649.0	1339.9
N (Observation)	238	238	238
N (Participant)	119	119	119

Note: Reference levels in dummy variables were set as follows: tDCS Group = Sham, Condition = Control. Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table S7 Results of regressions predicting acceptance and key parameters after controlling for the effect of inverse temperature (τ)

	Acceptance *b* (SE)	β *b* (SE)	γ *b* (SE)
Intercept	-0.99 (0.93)	16.85*** (1.39)	-4.42*** (0.73)
tDCS (Anodal)	0.19 (1.15)	-5.23** (1.59)	5.36*** (0.85)
tDCS (Cathodal)	1.43 (1.16)	-3.44* (1.57)	6.40*** (0.84)
Condition	-6.84*** (1.03)	-8.38***(1.31)	-0.94(0.79)
Offer Proportion	10.28*** (1.59)		
tDCS (Anodal) × Condition	-0.25 (1.43)	8.01*** (1.85)	-6.11*** (1.11)
tDCS (Cathodal) × Condition	-1.27 (1.46)	7.47*** (1.86)	-4.52*** (1.12)
tDCS (Anodal) × Offer Proportion	-3.16 (2.17)		
tDCS (Cathodal) × Offer Proportion	-2.84 (2.22)		
Condition × Offer Proportion	1.22 (1.57)		
tDCS (Anodal) × Condition × Offer Proportion	5.32* (2.08)		
tDCS (Cathodal) × Condition × Offer Proportion	5.11* (2.13)		
Larger payoff for proposer in the reported option[a]	0.29*** (0.03)		
Inverse Temperature (τ)	139.06** (47.55)	-85.65(89.23)	-98.46*(44.48)
AIC	7394.4	1577.1	1299.8
BIC	7580.4	1608.4	1331.1
N (Observation)	17136	238	238
N (Participant)	119	119	119

Note: [a]This variable was standardized before the analyses. We did not implement the same analysis for Δλ because no tDCS effect or related interaction on λ was observed in the regression analysis. Reference levels in dummy variables were set as follows: tDCS Group = Sham, Condition = Control. Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table S8 Results of regressions used for the mediation analyses

	Path c (Total Effect)	Path a	Path a*b and c' (Direct and Indirect Effect)
	ΔAccept% b (SE)	Δβ b (SE)	ΔAccept% b (SE)
Intercept	0.46*** (0.06)	-8.38*** (1.31)	0.18*** (0.04)
tDCS (Anodal)	-0.08 (0.08)	8.01*** (1.85)	0.19*** (0.06)
tDCS (Cathodal)	-0.05 (0.08)	7.47*** (1.86)	0.20*** (0.06)
Δβ			-0.03*** (0.003)
R^2	0.01	0.17	0.60
	ΔAccept% b (SE)	Δγ b (SE)	ΔAccept% b (SE)
Intercept	0.46*** (0.06)	-0.94 (0.74)	0.43*** (0.05)
tDCS (Anodal)	-0.08 (0.08)	-6.11*** (1.05)	-0.30*** (0.08)
tDCS (Cathodal)	-0.05 (0.08)	-5.02*** (1.06)	-0.22** (0.08)
Δγ			-0.04*** (0.01)
R^2	0.01	0.25	0.33

Note: Reference levels in dummy variables were set as follows: tDCS Group = Sham. We did not implement the same analysis for Δλ because no tDCS effect or related interactions on λ was observed in the regression analysis. Table also shows goodness-of-fit statistics. Significance: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

References

Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neuro-computational mechanisms of reinforcement learning and decision-making with the hBayesDM package. *Computational Psychiatry, 1*, 24-57.

Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2011). A model-based fMRI analysis with hierarchical Bayesian parameter estimation. *Journal of neuroscience, psychology, and economics, 4*(2), 95.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. 2012. *URL:* http://CRAN*. R-project. org/package= lme4*.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research, 33*(2), 261-304.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., . . . Graves, S. (2016). Package 'car'.

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19*(4), 25-42.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457-472.

Hu, Y., Hu, C., Derrington, E., Corgnet, B., Qu, C., & Dreher, J. C. (2021). Neural basis of corruption in power-holders. *Elife, 10*. doi:10.7554/eLife.63922

Huang, Y., Datta, A., Bikson, M., & Parra, L. C. (2019). Realistic vOlumetric-Approach to Simulate Transcranial Electric Stimulation -- ROAST -- a fully automated open-source pipeline. *Journal of Neural Engineering, 16*(5).

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.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494-1502.

Maréchal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with noninvasive brain stimulation. *Proceedings of the National Academy of Sciences, 114*(17), 4360-4364.

Qu, C., Hu, Y., Tang, Z., Derrington, E., & Dreher, J. C. (2020). Neurocomputational mechanisms underlying immoral decisions benefiting self or others. *Social Cognitive and Affective Neuroscience, nsaa029*.

R Core Team. (2014). R: A language and environment for statistical computing.

Steyer, R. (2014). MDMQ Questionnaire (English Version of Mdbf) [Online] Jena: Friedrich-Schiller-Universität Jena, Institut für Psychologie, Lehrstuhl für Methodenlehre und Evaluationsforschung. Available online at: https://www.metheval.uni-jena.de/mdbf.php (Accessed April 4, 2016).

Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A. (2014). Be nice if you have to-The neurobiological roots of strategic fairness. *Social Cognitive and Affective Neuroscience*, nsu114.

Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits.

Elife, 7, e31185.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413-1432.

Zhang, L., Langersdorff, L., Mikus, N., Glaescher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: frameworks, pitfalls, and suggestions of best practices. *Social Cognitive & Affective Neuroscience, 15*(6), 695-707. doi:10.1093/scan/nsaa089