



HAL
open science

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

► **To cite this version:**

Yang Hu, Remi Philippe, Valentin Guigon, S Zhao, Edmund A. Derrington, et al.. Perturbation of Right Dorsolateral Prefrontal Cortex (rDLPFC) Makes Power-Holders Less Resistant to Tempting Bribes. *Psychological Science*, In press. hal-03379041

HAL Id: hal-03379041

<https://hal.science/hal-03379041v1>

Submitted on 15 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3
4 Yang Hu^{1,2}, Rémi Phillippe^{2,3†}, Valentin Guigon^{2,3†}, Sasa Zhao^{2,3†}, Edmund Derrington^{2,3},
5 Brice Corgnet⁴, James Bonaiuto^{2,3}, Jean-Claude Dreher^{2,3*}

6
7 ¹School of Psychology and Cognitive Science, East China Normal University, Shanghai,
8 China

9 ²Neuroeconomics, Reward and Decision Making Laboratory, Institut des Sciences
10 Cognitives Marc Jeannerod, CNRS, France

11 ³Université Claude Bernard Lyon 1, Lyon, France

12 ⁴EmLyon, Ecully, France

13
14
15 *Please address correspondence to:

16 Jean-Claude Dreher

17 CNRS UMR 5229,

18 Neuroeconomics, Reward and Decision Making Laboratory,

19 Institut des Sciences Cognitives Marc Jeannerod,

20 67 Bd Pinel, 69675 Bron, France

21 Tel: 00 334 37 91 12 38

22 Email: dreher@isc.cnrs.fr

23
24
25
26 †These authors equally contributed to this study.

27 **Abstract**

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

40
41
42

43 **Statement of Relevance**

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

56 **Introduction**

57 As one of the most common forms of corruption, bribery pervasively exists in governments,
58 enterprises, and other organizations all over the world (Dreher, Kotsogiannis, & McCorrison,
59 2007). In real life, bribes usually occur in interpersonal contexts where there is an asymmetry
60 in power between the parties involved, such as a power-holder who can exert an impact in the
61 briber's interest (Köbis, van Prooijen, Righetti, & Van Lange, 2016). Hence, bribes often result
62 in mutual benefits via collaboration between the two parties involved, but transgress moral
63 principles and legal rules. Despite that bribery-related issues have been widely investigated
64 in social sciences (Abbink, 2006; Mauro, 1995; Serra & Wantchekon, 2012), the
65 neurobiological roots of bribery and their underlying computations remain largely elusive.

66 How does a power-holder decide whether to take or refuse a bribe? Bribery-related
67 decision-making is supposed to follow the general framework of value-based decision-making
68 (Rangel, Camerer, & Montague, 2008) and the account of social preference (Fehr & Krajbich,
69 2014). In a simplified situation, a power-holder makes a choice based on a relative subjective
70 value (SV) between accepting and rejecting the bribe, calculated by pitting personal profits
71 against the other-regarding interests. Moreover, accepting a bribe often involves the
72 transgression of the moral principle and brings in moral costs, which affects the SV
73 computation (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014). A recent study has
74 identified the moral cost of colluding with a fraud committed by the briber, incurred by the
75 power-holder, which depreciates the decision weights on personal gains from the bribe and
76 thus decreases the acceptance rates (Hu et al., 2021). Notably, the moral cost of taking the
77 bribe is critically distinguished from the psychological cost of dishonesty (Fischbacher &
78 Föllmi-Heusi, 2013; Gneezy, Kajackaite, & Sobel, 2018; Mazar, Amir, & Ariely, 2008). In these
79 studies, the moral cost occurs if an individual cheats for personal profits, whereas in the bribery
80 scenario the moral cost for a power-holder is elicited due to colluding with a briber to obtain
81 morally-tainted benefits via taking a bribe.

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

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

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

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

129 **Methods**

130 **Participants**

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

140

141 **Task and Design**

142 Participants were randomly assigned to one of the three tDCS treatment conditions with
143 40 persons in each: (i) anodal stimulation (18 females; mean age: 22.6 ± 5.5 years), (ii)
144 cathodal stimulation over the rDLPFC (17 females; mean age: 21.9 ± 2.6 years), or (iii) sham
145 stimulation (19 females; mean age: 22.6 ± 4.8 years), which were unbeknownst to them (see
146 **SOM** for tDCS protocol).

147 The main experiment included a computerized incentive task and a follow-up paper-and-
148 pencil rating task, which lasted around 30 min in total (see **SOM** for procedure details). In the
149 computerized task, participants were assigned the role of the power-holder who decides to
150 accept or reject financial offers (see **Figure 2A**). In a cover story, they were informed that they
151 would be presented with a series of choices from an independent group, whose data were
152 collected previously by the experimenter. Specifically, participants were led to believe that this
153 independent group of online attendants (denoted as proposers hereafter) played a “Game of
154 Chance”. This independent group did not actually exist and the choices made by this group
155 were pre-determined by the task software. Each proposer was presented with two options that
156 would earn them different payoffs. The larger payoff ranged from 60 to 130 (in €; see details
157 below) and the smaller payoff was fixed at 5. One of the two payoffs was randomly indicated
158 by the computer as the one to be received. According to the rules of the game, the proposer
159 should report the payoff indicated by the computer, which determined his final payoff (i.e., the
160 Control condition). However, the response of the proposer was never checked by the
161 experimenters. This allowed the proposer to lie by reporting the alternative payoff that had not

162 been indicated by the computer when this would earn them more profit (i.e., the Bribe
163 condition). In other words, the only difference between the two conditions is that in the Bribe
164 condition the proposer cheated for a larger payoff by reporting the non-chosen larger payoff,
165 whereas the proposer honestly reports the chosen larger payoff in the Control condition.
166 Importantly, participants were told that each proposer had been informed that whether or not
167 they obtained the payoff of the reported option crucially depended on the decisions of a power-
168 holder (i.e., the participants themselves). To obtain the profits in the reported option, the
169 proposer could “share” a portion of the money from their potential gain (i.e., the reported larger
170 payoff) to influence the power-holder’s decision. The task for the power-holder was to decide
171 whether to accept or reject the offer based on the information above. If the power-holder
172 accepted the offer, both of them would benefit from the payoff. If the power-holder rejected the
173 offer, neither of them earned anything. Participants were informed that one of their decisions
174 would be randomly selected for payment in that trial at the end of the experiment.

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

184 We implemented a 3×2 mixed design by manipulating the *tDCS treatment* (a between-
185 subject factor) and the *task condition* (a within-subject factor). Crucially, we operationally
186 defined corrupt behaviors as the acceptance of offers proposed by the proposer only when
187 the proposer lied (the Bribe condition). Compared with the Control condition, accepting offers
188 in the Bribe condition incurred the moral cost of colluding with the proposer’s dishonesty. We
189 also manipulated the *offer proportion*, which was defined as the proportion of the amount the
190 proposer decided to share with the power-holder from the payoff the proposer would have
191 earned in the reported option, ranging from 10% to 90% (in steps of 10%; 9 levels). This
192 allowed us to investigate whether and how the degree of temptation of a bribe modulated
193 corrupt behaviors. To further increase the variance of offers, we set potential gains that could
194 be earned by the proposer (i.e., the larger payoff, which ranged from 60 to 130 in steps of 10;
195 8 levels). As a result, this yielded 72 trials, each involving a unique offer, which appeared once
196 in each condition.

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

212 **Data Analyses**

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

226 Results

227 tDCS over rDLPFC increased the probability of accepting bribes with higher offer 228 proportions

229 We first tested our main hypothesis regarding choice behavior. Using mixed-effect logistic
230 regression, we observed that participants were less likely to accept an offer in the bribe (vs.
231 control) condition (a main effect of *task condition*: $\chi^2_{(1)} = 126.94$, $p < 0.001$) and more likely to
232 do so when the offer proportion increased (a main effect of offer proportion: $\chi^2_{(1)} = 96.34$, $p <$
233 0.001). We also detected a significant two-way interaction between *task condition* and *offer*
234 *proportion* ($\chi^2_{(1)} = 33.05$, $p < 0.001$). *Post-hoc* analyses indicated that compared with the
235 Control condition, participants were more likely to accept offers when the offer proportion
236 increased in the Bribe condition ($z = 5.41$, $p < 0.001$).

237 More importantly, we found a significant three-way interaction between *tDCS group*, *task*
238 *condition*, and *offer proportion* with respect to whether the offer was accepted ($\chi^2_{(2)} = 8.04$, $p =$
239 0.018 ; see **Figure 3**). To follow up the three-way interaction, we performed *post-hoc* analyses
240 on choice for each tDCS group that incorporated *task condition*, *offer proportion*, and their
241 interaction as fixed-effect predictors. As a result, compared with the Control condition,
242 participants receiving either type of tDCS stimulation were more likely to accept offers when
243 the offer proportion increased in the Bribe condition (anodal: $z = 4.67$, $p < 0.001$; cathodal: $z =$
244 4.34 , $p < 0.001$), which was not the case in the Sham group ($z = 0.67$, $p = 0.501$; see **Table**
245 **S1** in the **SOM** for details).

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

249 tDCS over rDLPFC modulated the bribery-elicited moral cost on concern for personal 250 gains (β) and fairness (γ)

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

$$253 \quad SV(P_{PH}, P_P) = \beta P_{PH} + \lambda P_P + \gamma |P_P - P_{PH}|$$

$$254 \quad \beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \lambda_{Control}, \gamma_{Control}, & \text{if Control condition} \\ \beta_{Bribe}, \lambda_{Bribe}, \gamma_{Bribe}, & \text{if Bribe condition} \end{cases} \quad \text{Model 1}$$

255 In this model, SV denotes the subjective value of the choice. P_P and P_{PH} represents the offer's
256 payoff for the proposer and power-holder respectively given different choices (i.e., accept or
257 reject the offer). β and λ measure the decision weights on personal profits and proposer's gain

258 from the offer respectively. γ measures the sensitivity to the absolute payoff inequality between
259 the power-holder and the proposer. The posterior predictive check (PPC) revealed that the
260 proportion of acceptance predicted by this model could capture the proportion of observed
261 acceptance across individuals (both conditions for all groups: $r_s > 0.99$, $p_s < 0.001$; see **Figure**
262 **S3-S7** in the **SOM** for the PPC at various levels), which further justified the validity of our
263 model.

264

265 To examine how bribery-elicited moral cost affected each parameter and how tDCS
266 treatment modulated such effects, we implemented mixed-effect linear regression on each
267 parameter separately, by including *tDCS group*, *task condition*, and their interactions as the
268 fixed-effect predictors. We also allowed intercepts to vary across participants as the random
269 effects. As a result, we first found a main effect of task condition for all three parameters,
270 namely that participants devalued the personal gains (β : $F_{(1, 116)} = 18.04$, $p < 0.001$, $\text{partial-}\eta^2$
271 $= 0.092$), the proposer's gains (λ : $F_{(1, 116)} = 172.64$, $p < 0.001$, $\text{partial-}\eta^2 = 0.481$), and the
272 absolute payoff differences (γ : $F_{(1, 116)} = 96.33$, $p < 0.001$, $\text{partial-}\eta^2 = 0.320$) in the Bribe
273 condition relative to the Control condition. Furthermore, we observed a main effect of tDCS
274 treatment on γ ($F_{(2, 116)} = 20.42$, $p < 0.001$, $\text{partial-}\eta^2 = 0.166$). *Post-hoc* analyses showed that
275 participants in the Anodal (vs. Sham) group decreased their concern for the absolute payoff
276 differences ($t_{(116)} = 3.05$, $p_{(\text{FDR-corrected})} = 0.003$, Cohen's $d = 0.55$, 95% CI = [0.19, 0.92]), which
277 was even further reduced in the Cathodal group (vs. Anodal: $t_{(116)} = 3.35$, $p_{(\text{FDR-corrected})} = 0.002$,
278 Cohen's $d = 0.61$, 95% CI = [0.24, 0.98]; see **SOM** for details).

279

280 More intriguingly, we found an interaction effect between *tDCS group* and *task condition*
281 on decision weights on personal gains (β : $F_{(2, 116)} = 11.71$, $p < 0.001$, $\text{partial-}\eta^2 = 0.116$) and
282 absolute payoff differences (γ : $F_{(2, 116)} = 16.14$, $p < 0.001$, $\text{partial-}\eta^2 = 0.320$), but not on
283 proposer's gains (λ : $F_{(2, 116)} = 2.35$, $p = 0.100$, $\text{partial-}\eta^2 = 0.025$). *Post-hoc* analyses for β
284 showed that participants receiving cathodal (vs. sham) tDCS decreased weights on personal
285 gains in the Control condition ($t_{(213)} = -2.21$, $p_{(\text{FDR-corrected})} = 0.042$, Cohen's $d = 0.59$, 95% CI =
286 [-1.13, -0.06]) but increased them in the Bribe condition ($t_{(213)} = 2.55$, $p_{(\text{FDR-corrected})} = 0.035$,
287 Cohen's $d = 0.68$, 95% CI = [0.15, 1.22]). Anodal tDCS induced a similar effect of β in the
288 Control condition ($t_{(213)} = -3.55$, $p_{(\text{FDR-corrected})} = 0.001$, Cohen's $d = 0.95$, 95% CI = [-1.48, -0.41]),
289 however, the enhancement effect was not statistically significant in the Bribe condition ($t_{(213)} =$
290 1.58 , $p_{(\text{FDR-corrected})} = 0.172$, Cohen's $d = 0.42$, 95% CI = [-0.11, 0.95]). Regarding γ , *post-hoc*
291 analyses showed that compared with the Sham group, participants in both the Anodal ($t_{(228)} =$
292 5.91 , $p_{(\text{FDR-corrected})} < 0.001$, Cohen's $d = 1.42$, 95% CI = [0.93, 1.91]) and Cathodal groups ($t_{(228)}$
293 $= 7.46$, $p_{(\text{FDR-corrected})} < 0.001$, Cohen's $d = 1.80$, 95% CI = [1.31, 2.29]) were less aversive to
294 absolute payoff differences (i.e., the general inequality) in the Control condition. However, in
295 the Bribe condition, participants in the Cathodal group were less aversive to the absolute
296 payoff inequality compared with either the Sham ($t_{(228)} = 2.15$, $p_{(\text{FDR-corrected})} = 0.049$, Cohen's d

297 = 0.52, 95% CI = [0.04, 1.00]) or Anodal group ($t_{(228)} = 3.45$, $p_{(FDR-corrected)} = 0.002$, Cohen's $d =$
298 0.83, 95% CI = [0.35, 1.32]; see **Figure 4** for the descriptive summary for key parameters; see
299 **Figure S8** in the **SOM** for the visualization of the tDCS effect on differential parameters; also
300 see **Table S5-S7** in the **SOM** for details of statistical analyses)

301

302 **tDCS over rDLPFC modulates bribery-elicited moral cost on the choice behaviors by** 303 **mediating key parameters of the computation**

304 To further establish the link between the tDCS treatment, the bribery-elicited moral cost
305 on these parameters, and the choice behaviors, we implemented *post-hoc* mediation analyses
306 with tDCS group as the predictor, the differential parameters as the mediator (i.e., $\Delta\beta = \beta_{\text{Bribe}}$
307 $-\beta_{\text{Control}}$, $\Delta\gamma = \gamma_{\text{Bribe}} - \gamma_{\text{Control}}$), and the differential acceptance rate as the dependent variable
308 (i.e., $\Delta\text{Accept} = \text{Accept}_{\text{Bribe}} - \text{Accept}_{\text{Control}}$). A bootstrapping procedure was applied to the
309 mediation effect (i.e., 5000 bootstraps). We showed that although the tDCS treatment did not
310 directly modify the bribery-specific effect on choice behaviors (i.e., total effect, path c : $p_s > 0.3$
311 for both tDCS effects), the differential parameters mediated the impact of tDCS treatment on
312 the bribery-specific effect on the behaviors (i.e., direct effect, path c' : $p_s < 0.001$ in both tDCS
313 effects for $\Delta\beta$ and in the anodal tDCS for $\Delta\gamma$, $p = 0.007$ in the cathodal tDCS for $\Delta\gamma$; indirect
314 effect, path $a*b$: $\Delta\beta$: anodal: -0.27, 95% CI: [-0.40, -0.15]; cathodal: -0.26, 95% CI: [-0.39, -
315 0.12]; $\Delta\gamma$: anodal: 0.21, 95% CI: [0.13, 0.30]; cathodal: 0.18, 95% CI: [0.07, 0.28]; see **Figure**
316 **5**; also see **Table S8** in the **SOM** for detailed regression outputs).

317 **Discussion**

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

335
336 More interestingly, the disruption of rDLPFC (i.e., both Anodal and Cathodal groups) made
337 participants, as power-holders, more likely to accept bribes (vs. standard offers) as the size of
338 the prospective payoff increased. Importantly, this tDCS effect over rDLPFC did not influence
339 other measures (e.g., DT, subjective ratings), suggesting that general cognitive or affective
340 processes are less likely to constitute the underlying mechanism. Taking a model-based
341 approach, we further showed that disrupting rDLPFC also alters the computations that
342 contribute to bribery decisions. Specifically, the cathodal tDCS over rDLPFC mitigates the
343 effect of the moral cost on personal gains due to bribe-taking ($\Delta\beta$). This finding coheres with
344 a previous brain-lesion study in which patients with lesions of DLPFC selectively reduced the
345 moral cost to personal profits (Zhu et al., 2014). Moreover, altering the rDLPFC excitability via
346 cathodal tDCS enhanced the effect of the bribery-elicited moral cost on fairness concern ($\Delta\gamma$).
347 As noted previously, studies using a standard UG consistently showed that inhibiting the
348 rDLPFC by low-frequency repetitive TMS (Knoch et al., 2006) or cathodal tDCS (Knoch et al.,
349 2008; Speitel et al., 2019) increases the tolerance of unfairness. While we replicated these
350 findings by showing a less negative γ for the Cathodal (vs. Sham) group in the Control
351 condition, we found that participants in the Cathodal group become more aversive to the

352 inequity between themselves and the proposer. Collectively, these results in the Cathodal
353 group indicate a dual role of rDLPFC during bribery-related decision making: it not only
354 overrides selfish motivation when it conflicts with moral principles (Carlson & Crockett, 2018)
355 but also integrates the moral cost in modulating fairness concern. This account is further
356 endorsed by the mediation analyses, which establish the link between rDLPFC, computations
357 underlying bribery-related decision making and final behaviors.

358

359 It is worth noting that the excitation of rDLPFC via anodal tDCS has a similar effect as
360 cathodal tDCS in modulating bribe-taking behaviors and underlying computations. There is no
361 *a priori* reason to believe that anodal and cathodal tDCS should induce opposite behavioral
362 effects in the moral domain. Indeed, previous evidence is mixed concerning the anodal effect
363 on moral behaviors which varies in different paradigms. Although Maréchal *et al.* (2017)
364 showed that anodal tDCS over rDLPFC increased honesty in a die-rolling task, another tDCS
365 study with an instrumental deception paradigm indicated the opposite effect (Mameli *et al.*,
366 2010). In agreement with this, an fMRI study has also shown that DLPFC is recruited more in
367 dishonest individuals when they have a chance to cheat (Greene & Paxton, 2009). Moreover,
368 the classical polarity-effect of tDCS (i.e., anodal-excitation and cathodal-inhibition) has been
369 shown to be much less common in the cognitive domain than in the motor domain (Jacobson,
370 Koslowsky, & Lavidor, 2012). A systematic review has revealed highly variable effects of tDCS
371 over the DLPFC on cognitive functions such as working memory (Tremblay, Lepage, Latulipe-
372 Loiselle, Fregni, & Théoret, 2014). Such inconsistent effects also exist in the social domain.
373 For example, although inhibiting rDLPFC with cathodal tDCS consistently enhances the
374 tolerance to unfairness (Knoch *et al.*, 2008; Speitel *et al.*, 2019), no evidence suggests that
375 anodal tDCS increases fairness concern (Speitel *et al.*, 2019). Lastly, there are large individual
376 variations in tDCS effects on modulating behaviors (López-Alonso, Cheeran, Río-Rodríguez,
377 & Fernández-del-Olmo, 2014; Wiethoff, Hamada, & Rothwell, 2014) and in the relationship
378 between DLPFC engagement and moral behaviors (Hu *et al.*, 2021; Yin & Weber, 2018).
379 Together, our findings confirm that the classical polarity-effect of tDCS, originally observed in
380 the primary motor cortex, should not be expected to be directly applied to other brain areas
381 and social/moral behaviors such as corruption.

382

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

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

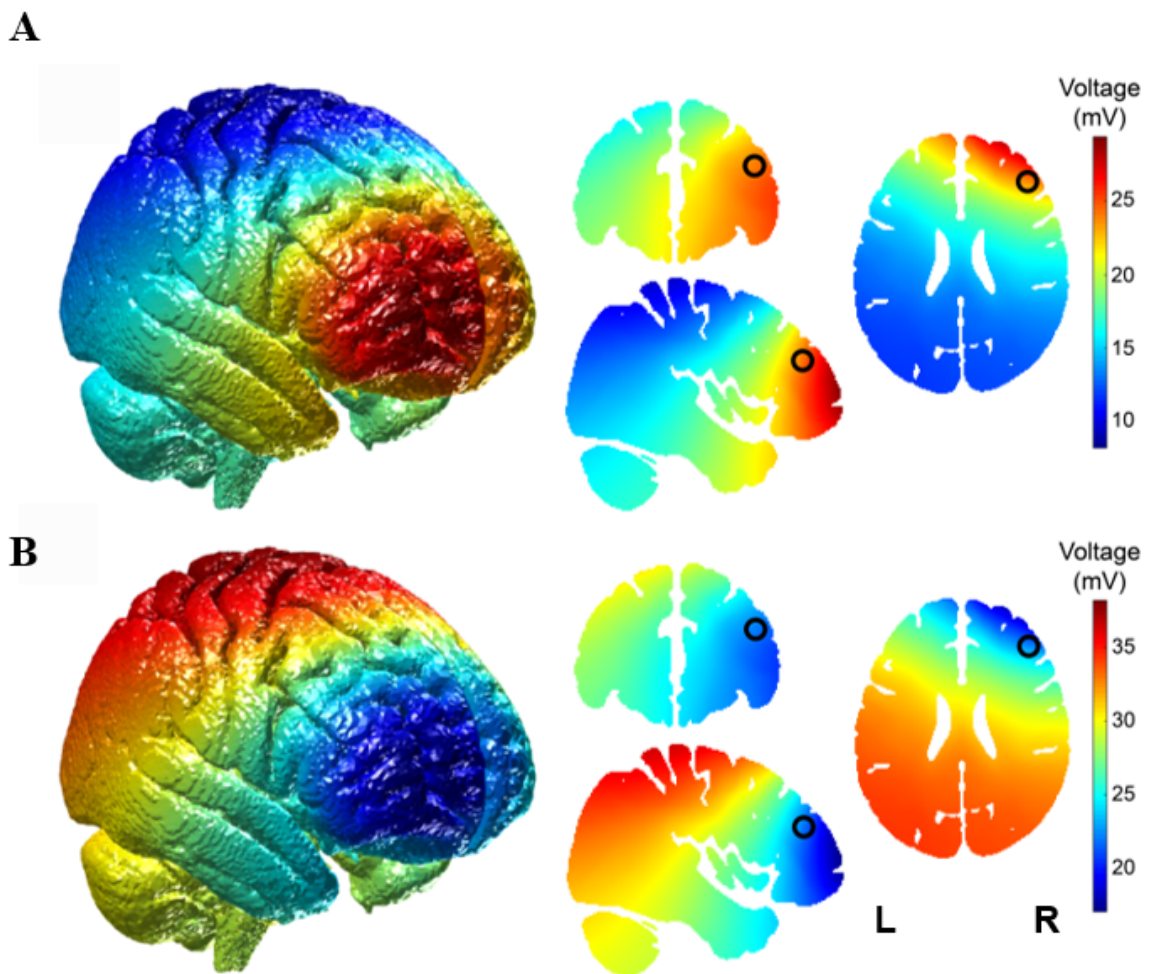
389

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

395 **Acknowledgments**

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

401 **Figures**



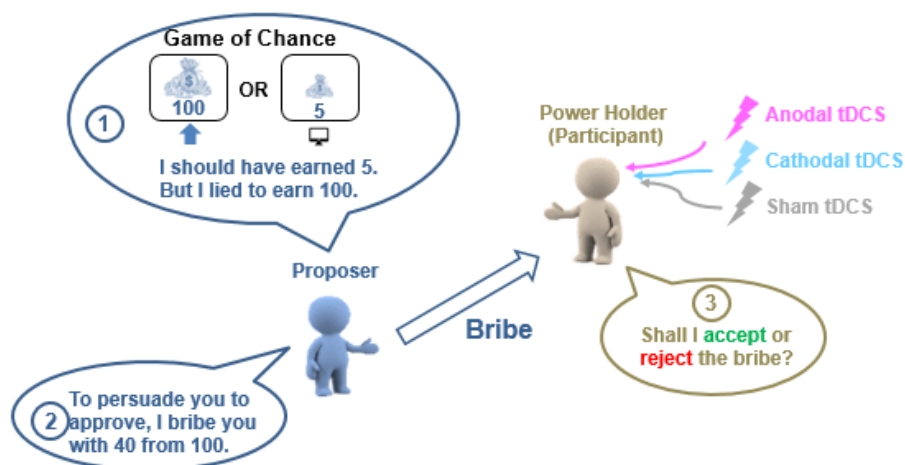
402

403 **Figure 1. Electric field simulation for (A) anodal and (B) cathodal tDCS stimulation.**

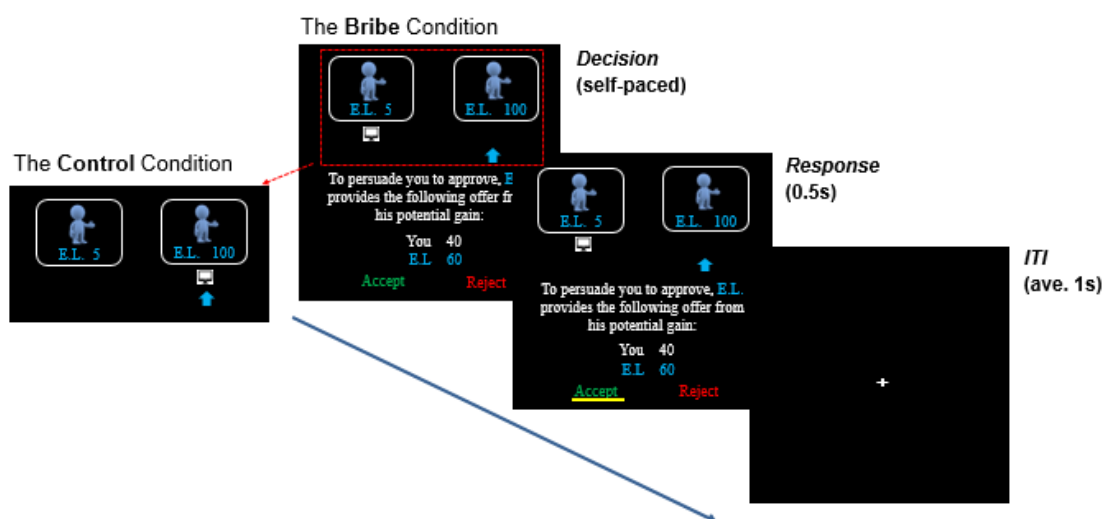
404 Based on previous literature closely relevant to the current study (Knoch et al., 2006; Strang
405 et al., 2014), we chose the position centering around the Talarach coordinate of 39/37/22 as
406 our target site. This location approximately corresponds to the electrode position of AF4 in the
407 10-20 system of EEG cap (the right panel; marked with a black circle). The vertex was chosen
408 as the reference electrode based on the study by Marechal et al (2017), which corresponds to
409 the electrode position of Cz. Electrodes were simulated as pads, with a 100x100x3mm pad
410 located over Cz and a 70x50x3mm pad located over AF4, using standard 10-10 system
411 locations. Tissue conductivities were set as white matter=0.11 S/m, gray matter=0.21 S/m,
412 CSF=0.53 S/m, bone=0.02 S/m, and skin=0.90 S/m. For the anodal simulation, 1.5mA was
413 set as inward flowing current from the AF4 pad, and -1.5mA outward flowing current from the
414 Cz pad, and vice versa for the cathodal simulation. The simulation was performed via ROAST
415 (Huang, Datta, Bikson, & Parra, 2019; <https://github.com/andypotatohy/roast>). Abbreviations:
416 L: left; R: right.

417

A



B

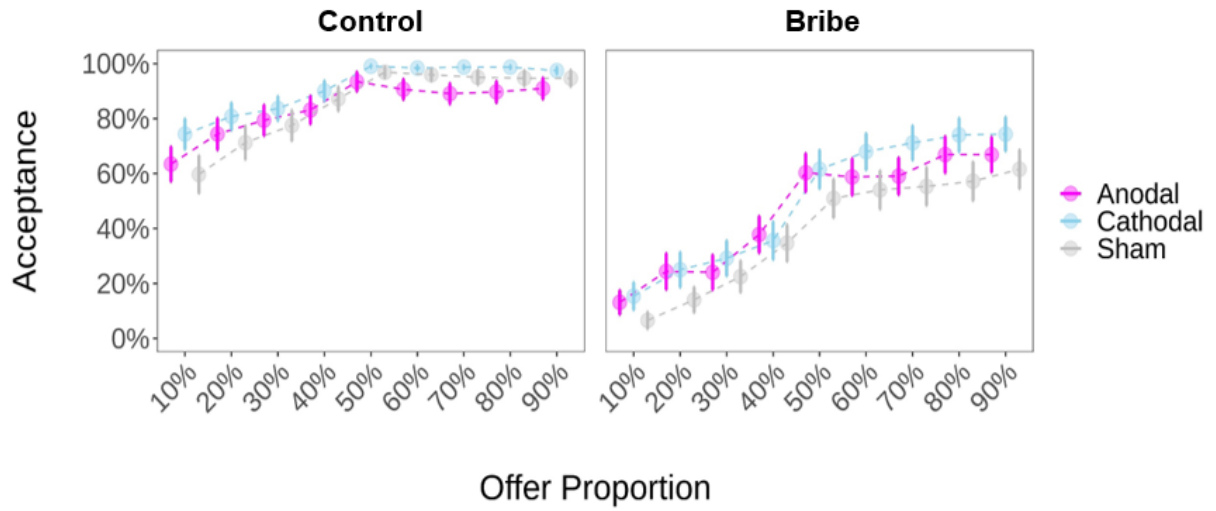


418

419

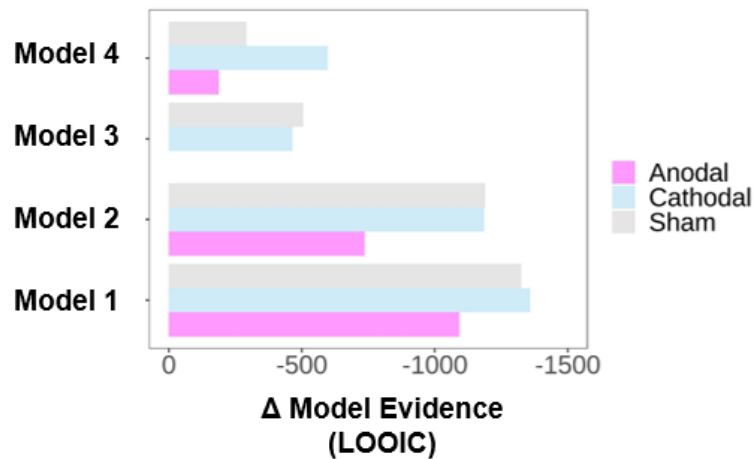
420 **Figure 2 Task design. (A) Schematic illustration of the tDCS manipulation and the**
 421 **behavioral paradigm.** All participants were assigned randomly to one of the three tDCS
 422 groups (i.e., anodal, cathodal or sham). The task comprised two roles, a proposer (i.e., a
 423 fictitious participant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

433 bribe here), a yellow bar appeared on the corresponding option to highlight the choice for 0.5
434 s, which was followed by a fixation (i.e., 0.6~1.4 s with a mean of 1s). Trials in the Control
435 condition followed the same procedure except that the proposer truthfully reported the
436 selected larger payoff (as indicated by the alignment of the blue arrow and the icon of a
437 computer).
438

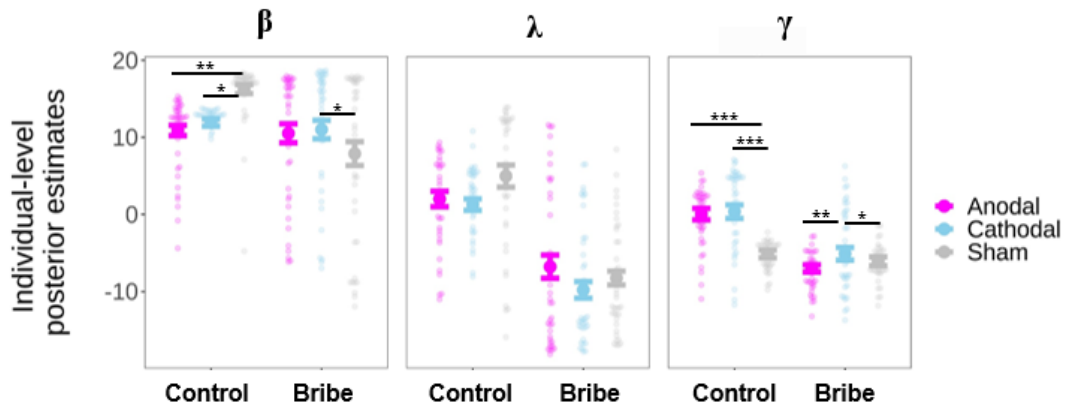


439
 440 **Figure 3. Results of acceptance rate (%).** Mean acceptance rate plotted as a function of
 441 *tDCS group* (Anodal/Cathodal/Sham), *task condition* (Control/Bribe), and *offer proportion* (10%
 442 to 90% in steps of 10%). Error bars represent SEM.
 443
 444

A



B



445

446

447

448

449

450

451

452

453

454

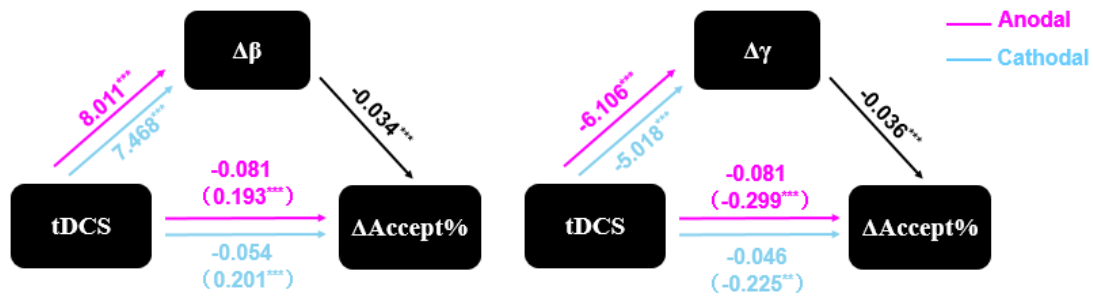
455

456

457

458

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.



459

460

461

462

463

464

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

465 **Reference**

- 466 Abbink, K. (2006). Laboratory experiments on corruption. In S. Rose-Ackerman (Ed.), *International*
467 *handbook on the economics of corruption* (pp. 418-437).
- 468 Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neuro-computational mechanisms of
469 reinforcement learning and decision-making with the hBayesDM package. *Computational*
470 *Psychiatry*, 1, 24-57.
- 471 Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: social information processing
472 in the human striatum. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1), 61-73.
- 473 Carlson, R. W., & Crockett, M. J. (2018). The lateral prefrontal cortex and moral goal pursuit.
474 *Current Opinion in Psychology*, 24, 77-82. doi:10.1016/j.copsy.2018.09.007
- 475 Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others
476 outweighs harm to self in moral decision making. *Proceedings of the National Academy of*
477 *Sciences*, 111(48), 17320-17325.
- 478 Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the
479 neural systems of reward during the trust game. *Nature neuroscience*, 8(11), 1611.
- 480 Dreher, A., Kotsogiannis, C., & McCorrison, S. (2007). Corruption around the world: Evidence from
481 a structural model. *Journal of comparative economics*, 35(3), 443-466.
- 482 Fehr, E., & Krajbich, I. (2014). Social preferences and the brain. In P. W. Glimcher & E. Fehr (Eds.),
483 *Neuroeconomics (Second Edition)* (pp. 193-218): Elsevier.
- 484 Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating.
485 *Journal of the European Economic Association*, 11(3), 525-547.
- 486 Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural
487 correlates of context-dependent advantageous-and disadvantageous-inequity aversion.
488 *Proceedings of the National Academy of Sciences*, 115(33), E7680-E7689.
- 489 Gneezy, U., Kajackaite, A., & Sobel, J. (2018). Lying Aversion and the Size of the Lie. *American*
490 *Economic Review*, 108(2), 419-453.
- 491 Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and
492 dishonest moral decisions. *Proceedings of the National Academy of Sciences*, 106(30),
493 12506-12511.
- 494 Hu, Y., He, L., Zhang, L., Wolk, T., Dreher, J. C., & Weber, B. (2018). Spreading inequality: neural
495 computations underlying paying-it-forward reciprocity. *Soc Cogn Affect Neurosci*, 13(6),
496 578-589. doi:10.1093/scan/nsy040
- 497 Hu, Y., Hu, C., Derrington, E., Corgnet, B., Qu, C., & Dreher, J. C. (2021). Neural basis of corruption
498 in power-holders. *Elife*, 10. doi:10.7554/eLife.63922
- 499 Huang, Y., Datta, A., Bikson, M., & Parra, L. C. (2019). Realistic vOlumetric-Approach to Simulate
500 Transcranial Electric Stimulation -- ROAST -- a fully automated open-source pipeline.
501 *Journal of Neural Engineering*, 16(5).
- 502 Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive
503 domains: a meta-analytical review. *Experimental brain research*, 216(1), 1-10.
- 504 Karim, A. A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., & Birbaumer, N. (2010).
505 The Truth about Lying: Inhibition of the Anterior Prefrontal Cortex Improves Deceptive
506 Behavior. *Cerebral cortex*, 20(1), 205-213.
- 507 Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., & Fehr, E. (2008). Studying the

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

510 Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal
511 fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829-832.

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

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

517 Mameli, F., Mrakic-Sposta, S., Vergari, M., Fumagalli, M., Macis, M., Ferrucci, R., . . . Priori, A.
518 (2010). Dorsolateral prefrontal cortex specifically processes general - but not personal -
519 knowledge deception: Multiple brain networks for lying. *Behavioural Brain Research*,
520 211(2), 164-168.

521 Maréchal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with
522 noninvasive brain stimulation. *Proceedings of the National Academy of Sciences*, 114(17),
523 4360-4364.

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

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

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

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

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

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

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

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

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

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

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

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

550

551 **Supplementary Materials (SOM) for**

552 **Perturbation of Right Dorsolateral Prefrontal Cortex (rDLPFC) Makes Power-Holders**

553 **Less Resistant to Tempting Bribes**

554

555 Yang Hu^{1,2}, Rémi Phillipe^{2,3†}, Valentin Guigon^{2,3†}, Sasa Zhao^{2,3†}, Edmund Derrington^{2,3},
556 Brice Corgnet⁴, James Bonaiuto^{2,3}, Jean-Claude Dreher^{2,3*}

557 ¹School of Psychology and Cognitive Science, East China Normal University, Shanghai,
558 China

559 ²Neuroeconomics, Reward and Decision Making Laboratory, Institut des Sciences
560 Cognitives Marc Jeannerod, CNRS, France

561 ³Université Claude Bernard Lyon 1, Lyon, France

562 ⁴EmLyon, Ecully, France

563

564 *Correspondence to: dreher@isc.cnrs.fr

565 †These authors equally contributed to this study.

566

567 **This PDF file includes:**

568 Supplementary Methods

569 Supplementary Results

570 Figures S1 to S8

571 Tables S1 to S8

572 **Supplementary Methods**

573 **tDCS Protocol**

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

590

591 **Procedure**

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

604 The main experiment included a computerized incentive task (see Task and Design for
605 details) and a follow-up paper-and-pencil rating task, which lasted about 30 min in total. The

606 rating task was aimed to measure the subjective feelings about the task and evaluations of
607 behaviors of either proposers or themselves by means of a Likert scale (0 indicated none, 100
608 indicated very much). In particular, they indicated the degree of 1) moral inappropriateness of
609 the proposers' behaviors and their decisions (had they accepted offers), 2) moral conflict
610 during the decision period, 3) the guilt they felt (had they accepted offers) in each condition.
611 They also reported the degree to which they had a power advantage over proposers and
612 whether they perceived offers from the proposers as bribes.

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

624

625 **Data Analyses**

626 ***Model-free analyses***

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

640 et al., 2016).

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

658 **Computational Modelling**

659 We adopted a basic social preference model that has been used in a modified Dictator
660 Game, i.e., a task of splitting money between oneself and a partner (Tusche & Hutcherson,
661 2018). Specifically, this model assumes that the participant, in the role of the power-holder, is
662 supposed to pit the personal profit against the proposer’s gain as well as their payoff inequity.
663 In our task, the only difference between the Bribe and Control condition is whether a moral
664 transgression of colluding with a fraudulent proposer is involved in the decision-making
665 process. Hence, bribery-related decision making would additionally bring in a moral cost that
666 might prevent the power-holder from taking the bribe. Based on our previous fMRI study using
667 a similar paradigm (Hu et al., 2021), we clearly hypothesized that there would be a moral cost
668 on the personal profit from the bribe. In addition, we explored whether such moral cost also
669 impacts the other components (i.e., the proposer’s payoff and the absolute payoff inequality)
670 involved in the trade-off during bribery-related decision-making, which remains an open
671 question. Thus, the utility function can be written as follows:

$$672 \quad SV(P_{PH}, P_P) = \beta P_{PH} + \lambda P_P + \gamma |P_P - P_{PH}|$$

673
$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \lambda_{Control}, \gamma_{Control}, & \text{if Control condition} \\ \beta_{Bribe}, \lambda_{Bribe}, \gamma_{Bribe}, & \text{if Bribe condition} \end{cases} \text{ Model 1}$$

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

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

684
$$SV(P_{PH}, P_P) = \beta P_{PH} + \lambda P_P$$

685
$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \lambda_{Control}, & \text{if Control condition} \\ \beta_{Bribe}, \lambda_{Bribe}, & \text{if Bribe condition} \end{cases} \text{ Model 2}$$

686
$$SV(P_{PH}, P_P) = \beta P_{PH} + \gamma |P_P - P_{PH}|$$

687
$$\beta, \lambda, \gamma = \begin{cases} \beta_{Control}, \gamma_{Control}, & \text{if Control condition} \\ \beta_{Bribe}, \gamma_{Bribe}, & \text{if Bribe condition} \end{cases} \text{ Model 3}$$

688 In addition, we also adopted the Fehr-Schmidt model which assumes disparate degrees
 689 of inequity aversion depending on whether one person earns more or less than the other,
 690 defined as follows:
 691

692
$$SV(P_{PH}, P_P) = P_{PH} - \alpha \max(P_P - P_{PH}, 0) - \beta \max(P_{PH} - P_P, 0)$$

693
$$\alpha, \beta = \begin{cases} \alpha_{Control}, \beta_{Control}, & \text{if Control condition} \\ \alpha_{Control}, \beta_{Bribe}, & \text{if Bribe condition} \end{cases} \text{ Model 4}$$

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

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

699
$$p(\text{accept}) = \frac{e^{\tau SV_{\text{accept}}}}{e^{\tau SV_{\text{accept}}} + e^{\tau SV_{\text{reject}}}} = \frac{1}{1 + e^{-\tau(SV_{\text{accept}} - SV_{\text{reject}})}}$$

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

704 The above model was fit using a hierarchical Bayesian approach (HBA) via the
 705 "hBayesDM" package (Ahn, Haines, & Zhang, 2017), which adopts a Markov Chain Monte
 706 Carlo (MCMC) sampling scheme to perform full Bayesian inference. We chose HBA because
 707 it has been shown to provide much more stable and accurate estimates than other estimation
 708 approaches (e. g., maximum likelihood estimation; Ahn, Krawitz, Kim, Busemeyer, & Brown,
 709 2011). Convergence of the MCMC chains was assessed through the Gelman-Rubin R-hat
 710 Statistics (Gelman & Rubin, 1992). Here, R-hat values of all estimated parameters of each
 711 tDCS group for all models were smaller than 1.02, indicating adequate convergence of the
 712 MCMC chains.

713 For model comparisons, we adopted the leave-one-out information criterion (LOOIC) as
 714 the index for model evidence. Compared with other point estimate information criteria (e.g.,
 715 Akaike information criterion, AIC), LOOIC score can be more reliable by providing the estimate
 716 of out-of-sample predictive accuracy in a fully Bayesian way (Vehtari, Gelman, & Gabry, 2017).
 717 Conventionally, the lower LOOIC score indicates better out-of-sample prediction accuracy of
 718 the candidate model. A difference score of 10 on the information criterion scale is considered
 719 decisive (Burnham & Anderson, 2004). We selected the model with the lowest LOOIC for all
 720 tDCS groups as the winning model for subsequent analysis of key parameters. We also
 721 performed the posterior predictive check (PPC) both at the individual and group level following
 722 the procedure suggested by Zhang *et al* (2020) and used by our previous studies (Hu *et al.*,
 723 2021; Qu, Hu, Tang, Derrington, & Dreher, 2020) to examine whether the prediction of the
 724 model could capture the features of real behaviors of participants.

725 For each individual, we obtained the posterior mean of individual-level key parameters of
 726 the winning model for each condition (i.e., β , λ , γ of Model 1). To examine how bribery-elicited
 727 moral cost affect each parameter and how tDCS treatment modulated such effects, we
 728 implemented mixed-effect linear regression on each parameter separately, by including *tDCS*
 729 *group*, *task condition*, and their interactions as the fixed-effect predictors. We also allowed
 730 intercepts to vary across participants as the random effects. For further analyses and
 731 illustration purpose, the individual-level differential parameters between the Bribe and Control
 732 condition were also calculated to characterize the bribery-specific effect (i.e., $\Delta\beta = \beta_{\text{Bribe}} -$
 733 β_{Control} , $\Delta\lambda = \lambda_{\text{Bribe}} - \lambda_{\text{Control}}$, $\Delta\gamma = \gamma_{\text{Bribe}} - \gamma_{\text{Control}}$; same below; see **Figure S8**). To further establish

734 the link between the tDCS treatment, the bribery-elicited moral cost on these parameters, and
735 the choice behaviors, we implemented post-hoc mediation analyses using the *bootM* package
736 with tDCS group as the predictor, the differential parameters as the mediator, and the
737 differential acceptance rate (i.e., $\Delta\text{Accept} = \text{Accept}_{\text{Bribe}} - \text{Accept}_{\text{Control}}$) as the dependent
738 variable. Statistical inference was confirmed by using a bootstrapping procedure to test the
739 mediation effect (i.e., 5000 bootstraps).

740 **Supplementary Results**

741 **No tDCS effect was observed in other behavioral measures**

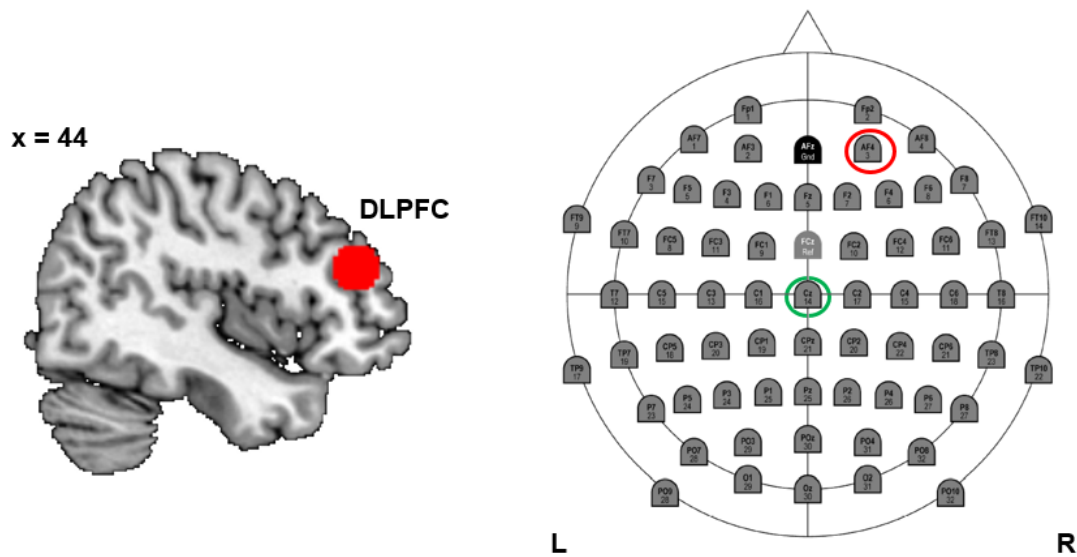
742 We investigated whether a similar effect of tDCS over rDLPFC existed in other behavioral
743 measures. Analyses on log-transformed DT revealed that participants responded slightly
744 slower in the Bribe condition (vs. Control; a main effect of task condition: $F_{(1,325)} = 5.97$, $p <$
745 0.001 , $partial-\eta^2 < 0.001$) and more quickly when the offer proportion increased (a main effect
746 of offer proportion: $F_{(1,17012)} = 67.03$, $p < 0.001$, $partial-\eta^2 = 0.004$). In addition, we observed a
747 two-way interaction between *task condition* and *offer proportion* ($F_{(1,16937)} = 16.59$, $p < 0.001$,
748 $partial-\eta^2 = 0.001$; see **Figure S2**). *Post-hoc* analyses indicated that participants responded
749 faster when the offer proportion increased in both conditions ($z_s < -3.15$, $p_s < 0.002$) but the
750 slope was less steep in the Bribe condition (vs. Control; $z = 4.07$, $p < 0.001$; see **Table S5** for
751 details of the regression output).

752 In addition, we also examined whether tDCS over rDLPFC affected subjective ratings, in
753 order to rule out alternative accounts that might explain the effect of tDCS on bribe-taking
754 behaviors. First, compared with the Control condition, participants in the Bribe condition felt a
755 higher level of moral conflict during the decision period ($F_{(1,116)} = 103.50$, $p < 0.001$, $partial-\eta^2$
756 $= 0.157$). They thought that the proposer's offering act ($F_{(1,116)} = 21.65$, $p < 0.001$, $partial-\eta^2 =$
757 0.472) and their hypothetical acceptance were more morally inappropriate ($F_{(1,115)} = 157.73$, p
758 < 0.001 , $partial-\eta^2 = 0.578$). They also felt more guilty for their hypothetical acceptances of
759 offers provided by the proposer ($F_{(1,115)} = 101.64$, $p < 0.001$, $partial-\eta^2 = 0.469$). However, none
760 of these measures were modulated by tDCS ($F_s < 1.01$, $p_s > 0.36$, $partial-\eta^2_s < 0.02$) nor its
761 interaction with task conditions ($F_s < 1.34$, $p_s > 0.26$, $partial-\eta^2_s < 0.03$). Second, participants
762 from the three tDCS groups reported similar levels of the sense of power over the proposer
763 ($F_{(2,116)} = 0.52$, $p = 0.597$, $partial-\eta^2 = 0.009$) and the sense of being bribed ($F_{(2,116)} = 1.04$, $p =$
764 0.357 , $partial-\eta^2 = 0.018$).

765 Regarding task-irrelevant measures, no difference between the three tDCS groups was
766 found in emotional state, as measured by the Multidimensional Mood Questionnaire (MDMQ)
767 (Steyer, 2014), reported before the main task (the awake-tired [AT] subscale: $F_{(2,115)} = 0.85$, p
768 $= 0.429$, $partial-\eta^2 = 0.015$; the calm-nervous [CN] subscale: $F_{(2,114)} = 0.22$, $p = 0.804$, $partial-$
769 $\eta^2 = 0.004$; the good-bad [GB] subscale: $F_{(2,115)} = 0.44$, $p = 0.645$, $partial-\eta^2 = 0.008$) or after
770 (AT: $F_{(2,116)} = 0.39$, $p = 0.677$, $partial-\eta^2 = 0.007$; CN: $F_{(2,116)} = 1.18$, $p = 0.312$, $partial-\eta^2 = 0.020$;
771 GB: $F_{(2,116)} = 0.95$, $p = 0.389$, $partial-\eta^2 = 0.016$). Cognitive reflection ability, as measured by
772 the Cognitive Reflection Test (Frederick, 2005), was unaffected by the tDCS manipulation ($\chi^2_{(4)}$
773 $= 5.28$, $p = 0.260$; see **Table S6** and **S7** for a descriptive summary of these measures).

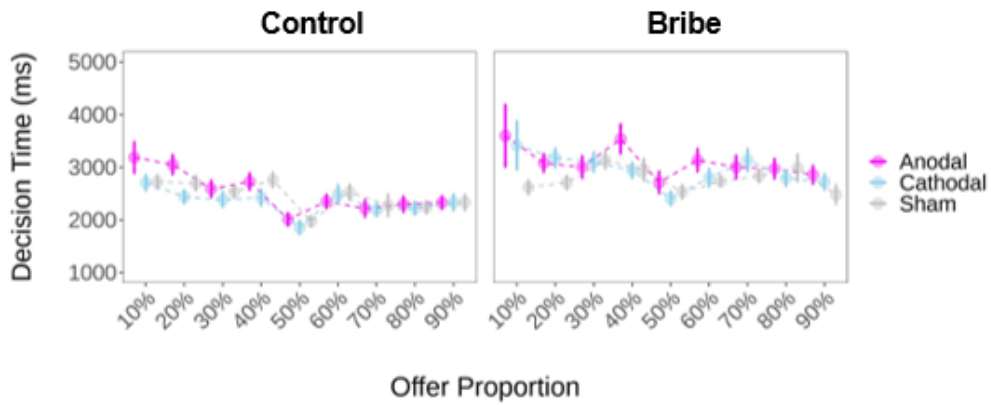
774 **Inverse temperature did not influence the tDCS effect on choice behavior and key** 775 **parameters in the winning model**

776 As the inverse temperature parameter (τ) varied between tDCS groups ($F_{(2, 116)} = 4.67, p$
777 $= 0.019, partial-\eta^2 = 0.08$; see **Table S4** for the descriptive summary), we performed control
778 analyses on the choice behavior and key parameters (i.e., β and γ) by including τ as a between-
779 group covariate to rule out the confounding effect of τ . Results showed that the main findings
780 related with the tDCS effect on behaviors (tDCS Group \times Condition \times Offer Proportion three-
781 way interaction: $\chi^2_{(2)} = 7.93, p = 0.019$) and key parameters (tDCS Group \times Condition two-way
782 interaction: $\beta: F_{(2, 116)} = 11.71, p < 0.001, partial-\eta^2 = 0.12$; $\gamma: F_{(2, 116)} = 16.14, p < 0.001, partial-$
783 $\eta^2 = 0.14$) still held after we took the effect of τ into account (see **Table R4** for complete
784 regression outputs). These findings indicated that the inverse temperature might not well
785 explained the tDCS effect on behaviors and its underlying computations.

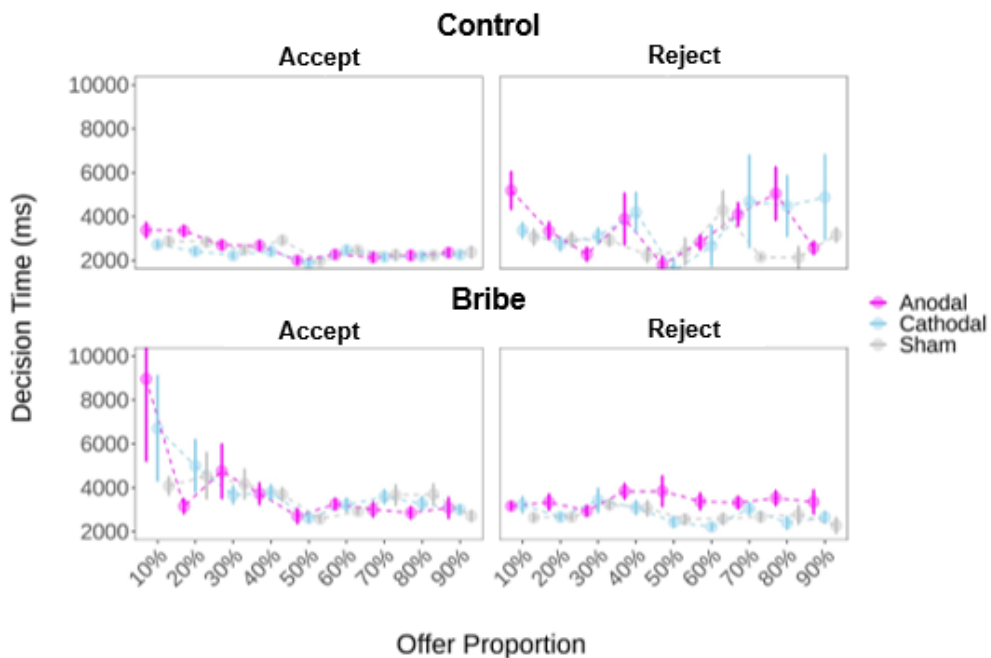


787
 788 **Figure S1. Display of the tDCS electrode localization.** Based on previous literature highly
 789 relevant to the current study (Knoch *et al.*, 2006; Strang *et al.*, 2014), we chose the position
 790 centering around the MNI coordinate of 39/37/22 as our target site (the left panel; a sphere
 791 of a 10mm radius was used for visualization). This location approximately corresponds to the
 792 electrode position of AF4 in the 10-20 system of 64-channel EEG cap (the right panel;
 793 marked with a red circle). The vertex was chosen as the reference electrode based on the
 794 study by Marechal *et al* (2017), which corresponds to the electrode position of Cz (the right
 795 panel; marked with a green circle).

A



B



796

797

798

799

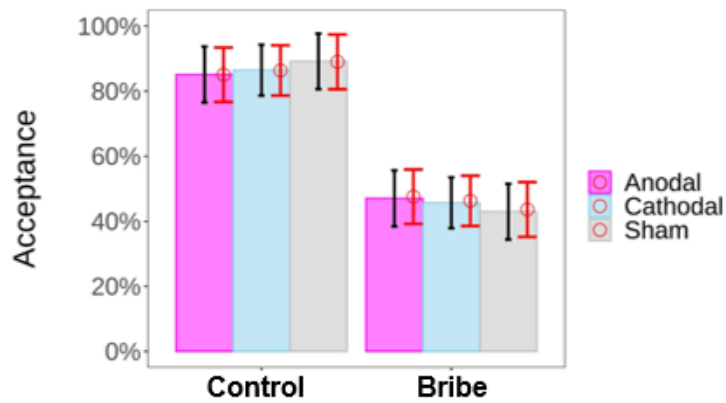
800

801

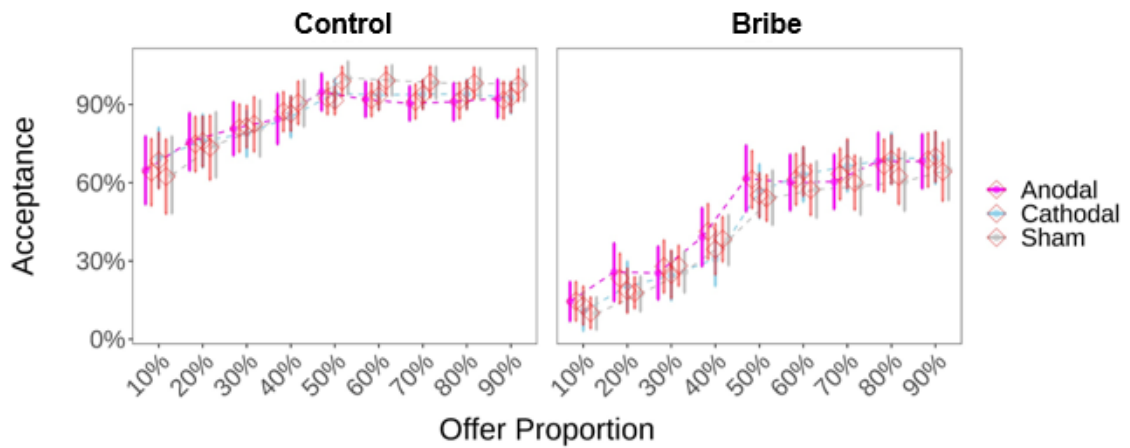
802

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.

A



B



803

804

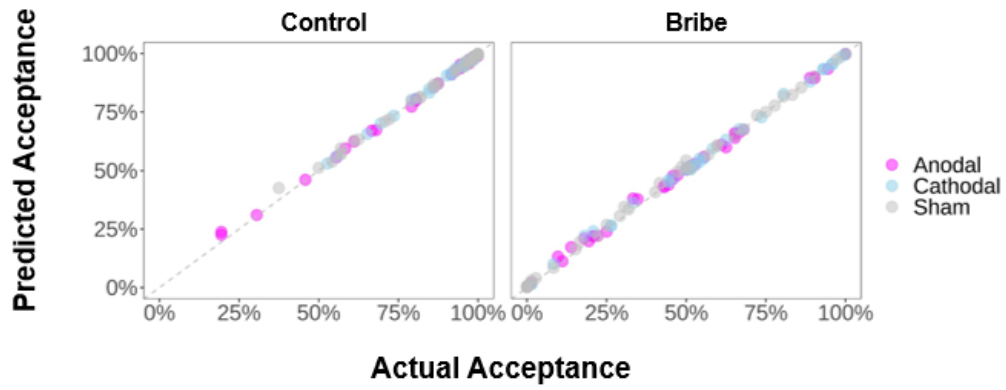
805

806

807

808

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.



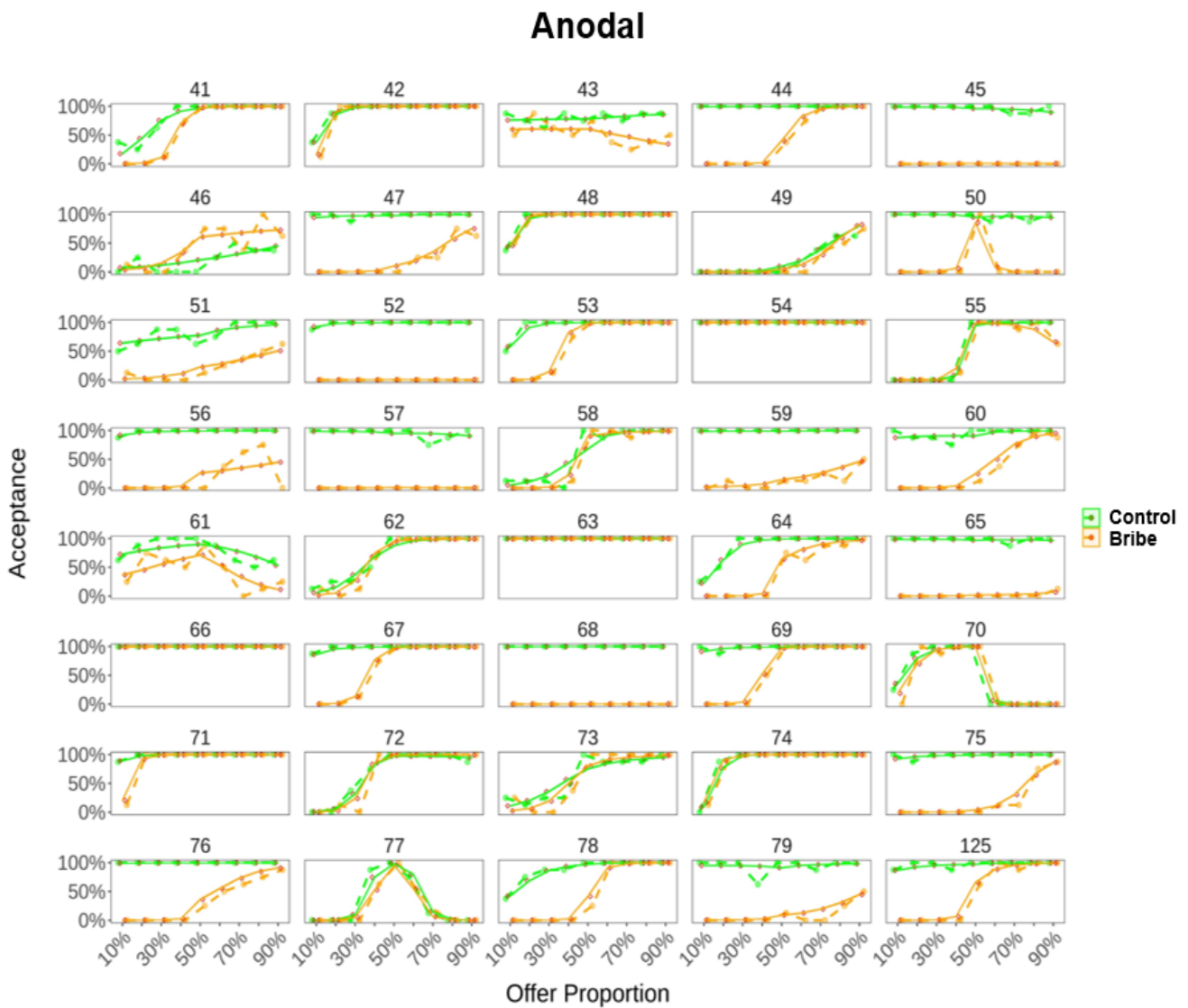
809

810

811

812

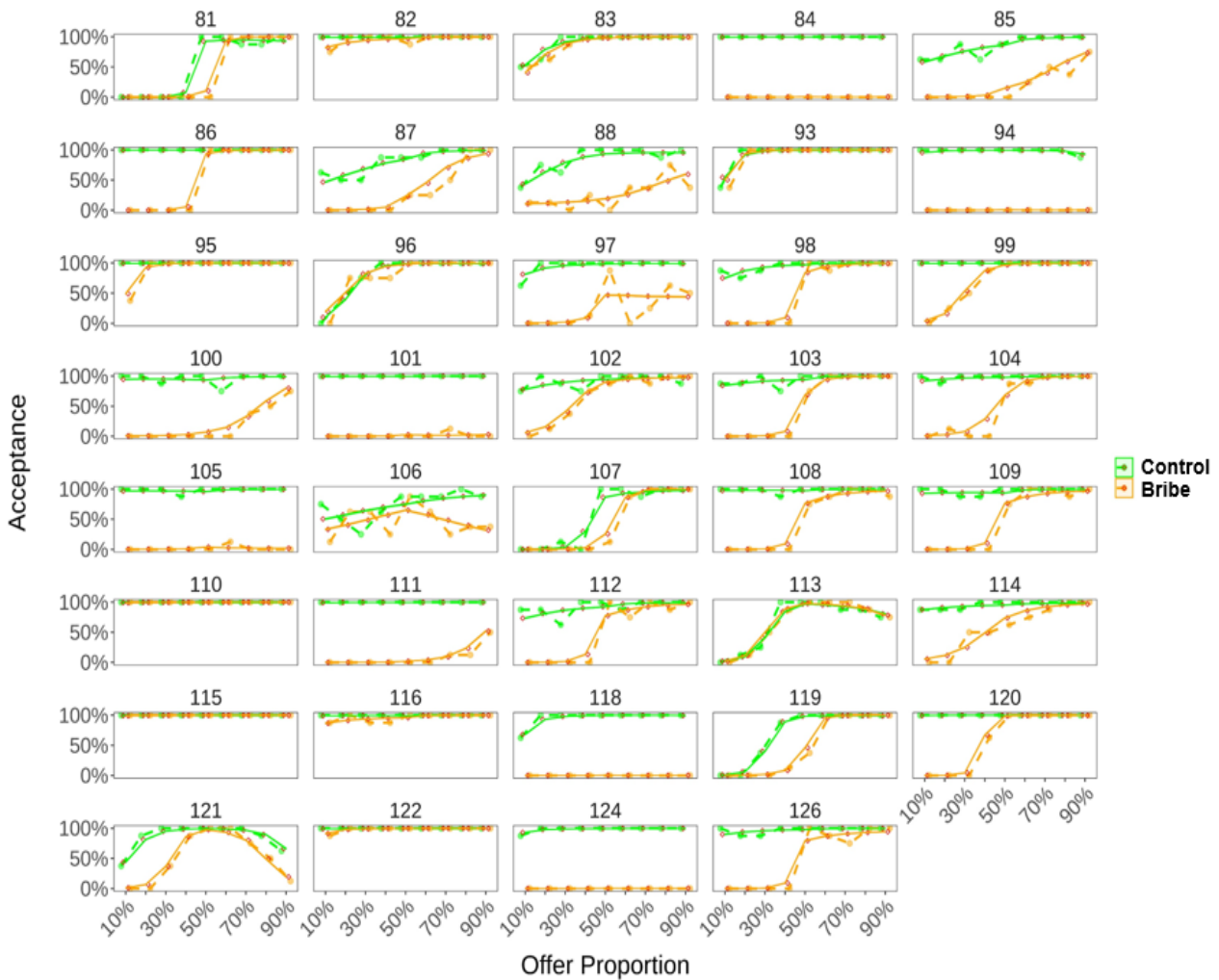
Figure S4. Posterior predictive check at the individual level. Relationship between predicted acceptance rates and actual acceptance rates across individuals. Filled dots represent individual data. Error bars represent 95% CI.



813
 814
 815
 816
 817
 818

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.

Cathodal



819

820 **Figure S6. Posterior predictive check at the individual level for the Cathodal group.**

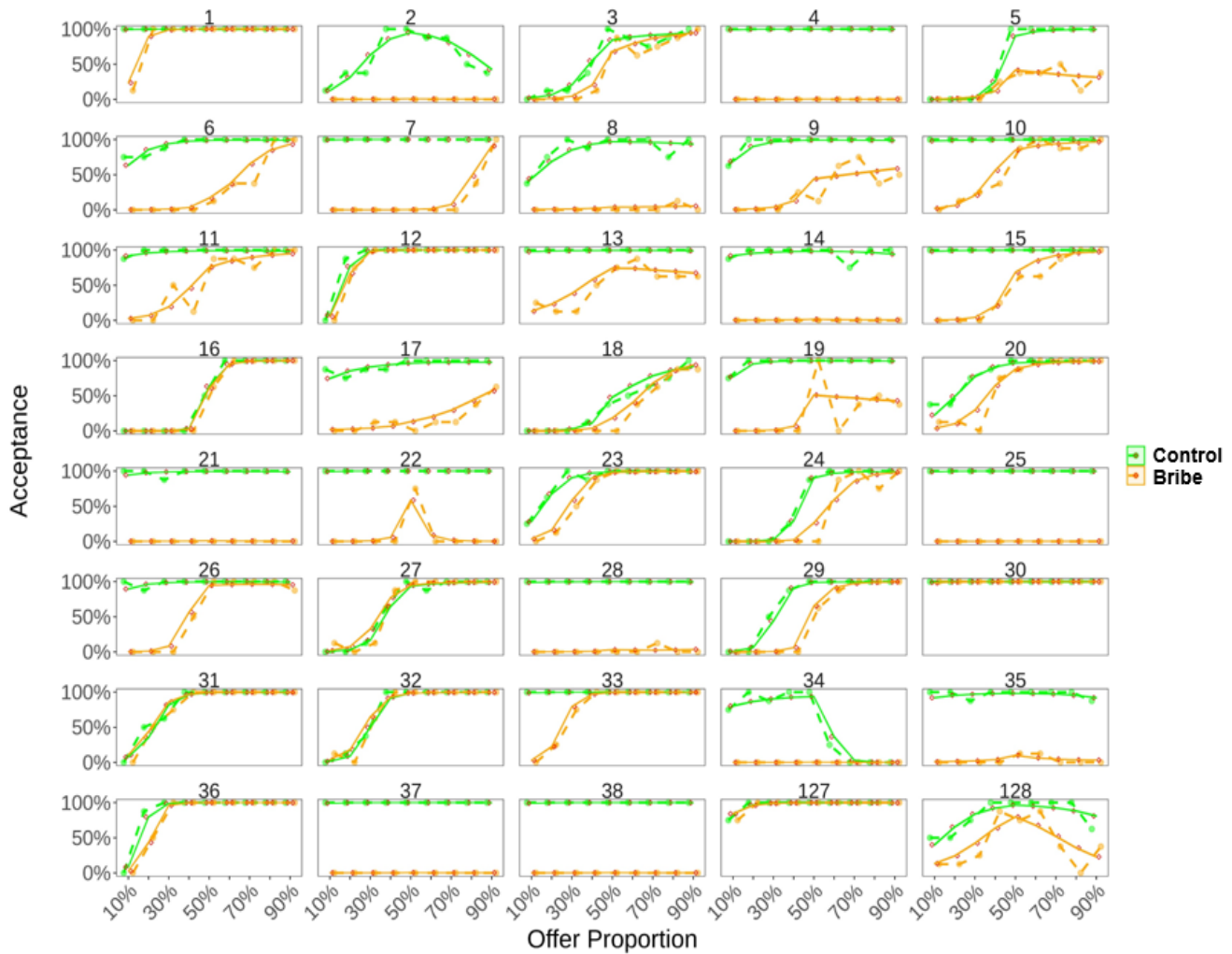
821 Mean predicted (red circles; connected by solid lines) and actual acceptance rates (filled

822 dots; connected by dashed lines) plotted as a function of task condition and offer proportion

823 across individuals in the Cathodal group. Numbers refer to subject ID. Solid lines that are

824 actually shaded areas represent 95% CI based on 4000 posterior samples.

Sham



825

826

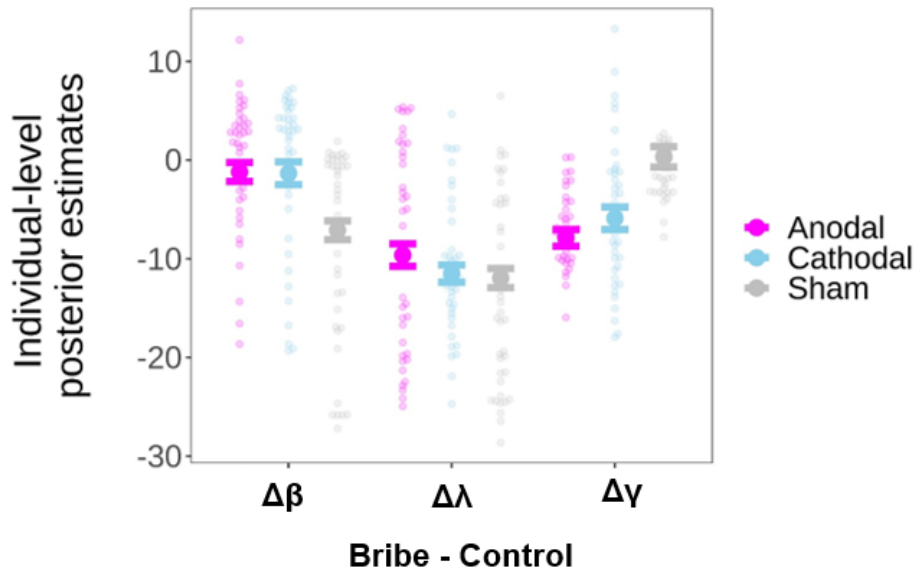
827 **Figure S7. Posterior predictive check at the individual level for the Sham group.** Mean

828 predicted (red circles; connected by solid lines) and actual acceptance rates (filled dots;

829 connected by dashed lines) plotted as a function of task condition and offer proportion

830 across individuals in the Sham group. Numbers refer to subject ID. Solid lines that are

actually shaded areas represent 95% CI based on 4000 posterior samples.



831

832 **Figure S8. The tDCS effect on differential parameters of the winning model.** This is
 833 another way to illustrate the interaction effect on key parameters. Differential parameters are
 834 calculated as follows: $\Delta\beta = \beta_{\text{Bribe}} - \beta_{\text{Control}}$, $\Delta\lambda = \lambda_{\text{Bribe}} - \lambda_{\text{Control}}$, $\Delta\gamma = \gamma_{\text{Bribe}} - \gamma_{\text{Control}}$. Each large
 835 filled dot represents the group-level mean; each smaller filled dot represents the data of a
 836 single participant. Error bars represent the SEM; Significance: * $p < 0.05$, ** $p < 0.01$, *** $p <$
 837 0.001 .

838 **Supplementary Tables**839 **Table S1 Results of mixed-effect logistic regressions predicting acceptance**

	All <i>b</i> (SE)	Control <i>b</i> (SE)	Bribe <i>b</i> (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

840 Note: ^a This variable was standardized before the analyses. Reference levels in dummy variables were
841 set as follows: tDCS Group = Sham, Condition = Control. Table also shows goodness-of-fit statistics:
842 AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: * $p < 0.05$, ** $p <$
843 0.01 , *** $p < 0.001$.

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

845

	All	Control ^b	Bribe ^b
	<i>b</i> (SE)	<i>b</i> (SE)	<i>b</i> (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 (Anodal) × 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

846 Note: ^a This variable was standardized before the analyses.

847 ^b We did not incorporate interactions between tDCS Group and offer proportion, as none of these effects
 848 was significant in the regression using all trials. Reference levels in dummy variables were set as follows:
 849 tDCS Group = Sham, Condition = Control, Decision = acceptance. Table also shows goodness-of-fit
 850 statistics: AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. Significance: ^{*}*p* <
 851 0.05, ^{**}*p* < 0.01, ^{***}*p* < 0.001.

852 **Table S3 Descriptive statistics of task-relevant subjective rating**

853

		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

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

856 **Table S4 Descriptive statistics of other measures**

857

		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

858

859 Note: ^aData of the pre-task MDMQ measures from one participant in the Cathodal group was missing

860 ^bData of pre-task MDMQ measures (only in CN subscale) from one participant in the Sham group was
861 missing.

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

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

866

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

867

868 **Table S6 Results of linear regressions predicting parameters in the winning model**
 869

	β	λ	γ
	<i>b</i> (SE)	<i>b</i> (SE)	<i>b</i> (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

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

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

	Acceptance <i>b</i> (SE)	β <i>b</i> (SE)	γ <i>b</i> (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

875
 876 Note: ^aThis variable was standardized before the analyses. We did not implement the same analysis
 877 for $\Delta\lambda$ because no tDCS effect or related interaction on λ was observed in the regression analysis.
 878 Reference levels in dummy variables were set as follows: tDCS Group = Sham, Condition = Control.
 879 Table also shows goodness-of-fit statistics: AIC = Akaike Information Criterion, BIC = Bayesian
 880 Information Criterion. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

881 **Table S8 Results of regressions used for the mediation analyses**

882

	Path c (Total Effect)	Path a	Path a*b and c' (Direct and Indirect Effect)
	$\Delta\text{Accept\%}$ b (SE)	$\Delta\beta$ b (SE)	$\Delta\text{Accept\%}$ 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)
$\Delta\beta$			-0.03*** (0.003)
R^2	0.01	0.17	0.60

	$\Delta\text{Accept\%}$ b (SE)	$\Delta\gamma$ b (SE)	$\Delta\text{Accept\%}$ 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)
$\Delta\gamma$			-0.04*** (0.01)
R^2	0.01	0.25	0.33

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

887 **References**

- 888 Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neuro-computational mechanisms of
889 reinforcement learning and decision-making with the hBayesDM package. *Computational*
890 *Psychiatry*, 1, 24-57.
- 891 Ahn, W.-Y., Krawitz, A., Kim, W., Busemeyer, J. R., & Brown, J. W. (2011). A model-based fMRI
892 analysis with hierarchical Bayesian parameter estimation. *Journal of neuroscience,*
893 *psychology, and economics*, 4(2), 95.
- 894 Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory
895 hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.
- 896 Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes.
897 R package version 0.999999-0. 2012. URL: <http://CRAN.R-project.org/package=lme4>.
- 898 Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in
899 model selection. *Sociological methods & research*, 33(2), 261-304.
- 900 Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., . . . Graves, S. (2016).
901 Package 'car'.
- 902 Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*,
903 19(4), 25-42.
- 904 Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences.
905 *Statistical science*, 457-472.
- 906 Hu, Y., Hu, C., Derrington, E., Corgnet, B., Qu, C., & Dreher, J. C. (2021). Neural basis of corruption
907 in power-holders. *Elife*, 10. doi:10.7554/eLife.63922
- 908 Huang, Y., Datta, A., Bikson, M., & Parra, L. C. (2019). Realistic vOlumetric-Approach to Simulate
909 Transcranial Electric Stimulation -- ROAST -- a fully automated open-source pipeline.
910 *Journal of Neural Engineering*, 16(5).
- 911 Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal
912 fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829-832.
- 913 Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research*
914 *Methods*, 49(4), 1494-1502.
- 915 Maréchal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with
916 noninvasive brain stimulation. *Proceedings of the National Academy of Sciences*, 114(17),
917 4360-4364.
- 918 Qu, C., Hu, Y., Tang, Z., Derrington, E., & Dreher, J. C. (2020). Neurocomputational mechanisms
919 underlying immoral decisions benefiting self or others. *Social Cognitive and Affective*
920 *Neuroscience*, nsaa029.
- 921 R Core Team. (2014). R: A language and environment for statistical computing.
- 922 Steyer, R. (2014). MDMQ Questionnaire (English Version of Mdbf) [Online] Jena: Friedrich-
923 Schiller-Universität Jena, Institut für Psychologie, Lehrstuhl für Methodenlehre und
924 Evaluationsforschung. Available online at: <https://www.metheval.uni-jena.de/mdbf.php>
925 (Accessed April 4, 2016).
- 926 Strang, S., Gross, J., Schuhmann, T., Riedl, A., Weber, B., & Sack, A. (2014). Be nice if you have
927 to-The neurobiological roots of strategic fairness. *Social Cognitive and Affective*
928 *Neuroscience*, nsu114.
- 929 Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by
930 changing attribute representations in domain-general and domain-specific brain circuits.

931 *Elife*, 7, e31185.

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

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

938