



HAL
open science

Perturbation of Right Dorsolateral Prefrontal Cortex Makes Power Holders Less Resistant to Tempting Bribes

Yang Hu, Rémi Philippe, Valentin Guigon, Sasa Zhao, Edmund Derrington,
Brice Corgnet, James J Bonaiuto, Jean-Claude Dreher

► **To cite this version:**

Yang Hu, Rémi Philippe, Valentin Guigon, Sasa Zhao, Edmund Derrington, et al.. Perturbation of Right Dorsolateral Prefrontal Cortex Makes Power Holders Less Resistant to Tempting Bribes. *Psychological Science*, 2022, 33, pp.412 - 423. 10.1177/09567976211042379 . hal-03842209

HAL Id: hal-03842209

<https://cnrs.hal.science/hal-03842209>

Submitted on 16 Nov 2022

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.

Perturbation of Right Dorsolateral Prefrontal Cortex Makes Power Holders Less Resistant to Tempting Bribes

Yang Hu^{1,2}, Rémi Philippe^{2,3}, Valentin Guigon^{2,3}, Sasa Zhao^{2,3}, Edmund Derrington^{2,3}, Brice Corgnet^{4,5}, James J. Bonaiuto^{2,3}, and Jean-Claude Dreher^{2,3}

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

² Neuroeconomics, Reward and Decision Making Laboratory, Institut des Sciences Cognitives Marc Jeannerod, Centre Nationale de la Recherche Scientifique (CNRS), Lyon, France

³ UFR Biosciences, Université Claude Bernard Lyon

⁴ EmLyon Business School

⁵ Groupe d'Analyse et de Théorie Economique, Lyon Saint-Etienne (GATE L-SE), France

Abstract

Bribery is a common form of corruption that takes place when a briber suborns a power holder to achieve an advantageous outcome at the cost of moral transgression. Although bribery has been extensively investigated in the behavioral sciences, its underlying neurobiological basis remains poorly understood. Here, we employed transcranial direct-current stimulation (tDCS) in combination with a novel paradigm (N = 119 adults) 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 as the relative value of the offer increased. 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 generating profits for oneself via taking bribes and reshapes the concern 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.

Introduction

As one of the most common forms of corruption, bribery is pervasive in governments, enterprises, and other organizations all over the world (Dreher et al., 2007). In real life, bribes usually occur in interpersonal contexts in which there is an asymmetry in power between the parties involved, such as when a power holder can exert an influence in the briber's interest (Köbis et al., 2016). Hence, bribes often result in mutual benefits via collaboration between the two parties involved but transgress moral principles and legal rules. Although bribery-related issues have been widely investigated in the social sciences (Abbink, 2006; Mauro, 1995; Serra & Wantchekon, 2012), the neurobiological roots of bribery and the underlying computations involved in deciding whether to accept a bribe 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 et al., 2008) and the account of social

40 preference (Fehr & Krajbich, 2014). In a simplified situation, a power holder makes a choice on the basis
41 of a relative subjective value between accepting and rejecting the bribe, calculated by pitting personal
42 profits against the other-regarding interests. Moreover, accepting a bribe often involves the transgression
43 of a moral principle and results in moral costs, which affects the subjective-value computation (Crockett
44 et al., 2014). A recent study identified the moral cost to the power holder of colluding with a fraud
45 committed by the briber, which depreciates the decision weights on per-sonal gains from the bribe and
46 thus decreases the accep-tance rates (Hu et al., 2021). Notably, the moral cost of taking the bribe is
47 critically distinguished from the psy-chological cost of dishonesty (Fischbacher & Föllmi- Heusi, 2013;
48 Gneezy et al., 2018; Mazar et al., 2008). In these studies, the moral cost occurs if an individual cheats for
49 personal profit, whereas in the bribery sce-nario, the moral cost for a power holder is elicited by collusion
50 with a briber to obtain morally tainted benefits via taking a bribe. It is well established that the right
51 dorsolateral pre-frontal cortex (rDLPFC) is critically involved in modulat-ing human social and moral
52 behaviors. Specifically, previous studies using an ultimatum game have consis-tently showed that
53 decreasing the neural excitability of rDLPFC—either by low-frequency repetitive transcranial magnetic
54 stimulation or by cathodal transcranial direct- current stimulation (tDCS)—makes the respondents more
55 likely to accept disadvantageous offers (Knoch et al., 2006, 2008; Speitel et al., 2019). In the moral domain,
56 inhibiting rDLPFC and related anterior prefron-tal areas with cathodal tDCS improves deceptive behav-iors
57 by reducing the reaction time to tell lies and increasing skillful lies (Karim et al., 2010). Using a dif-ferent
58 task, a brain-lesion study illustrated that patients with DLPFC lesions selectively increased self-serving
59 cheating behaviors (Zhu et al., 2014). Concerning the anodal tDCS effect over rDLPFC on social and moral
60 behaviors, the current evidence is less clear. There is no evidence supporting the hypothesis that a
61 responder’s intolerance of inequity is increased in the ultimatum game after they receive anodal tDCS
62 (Speitel et al., 2019). Regarding moral behaviors, par-ticipants who receive anodal tDCS are more likely to
63 behave honestly (Maréchal et al., 2017). Yet there is also evidence that anodal tDCS over DLPFC speeds up
64 dishonest decisions, suggesting an opposite effect (Mameli et al., 2010). Moreover, a recent functional
65 MRI (fMRI) study indicates that the DLPFC guides anticor-rupt behaviors contextually and selectively
66 modulates bribery-specific computations across individuals (Hu et al., 2021). Together, these results
67 suggest that the rDLPFC should play a pivotal role in bribery-related decision- making, but it remains
68 unclear how disrupting the rDLPFC specifically impacts corrupt acts and the com-putations underlying such
69 decision-making. Here, to examine whether rDLPFC exerts a causal influence in determining whether a
70 power holder would accept a bribe or not, we manipulated the neural excitability of rDLPFC via tDCS and
71 measured corrupt behaviors of power holders using a novel paradigm. Specifically, 120 healthy
72 participants were randomly assigned to three tDCS groups to causally modulate (anodal or cathodal tDCS)
73 or maintain (sham tDCS) the neural excitability of rDLPFC (see Fig. 1; see also Fig. S1 in the Supplemental
74 Material available online). Par-ticipants played the role of a power holder who decided whether another
75 (fictitious) person in a separate game would earn a given amount of money in a fraudulent manner (the
76 bribe condition) or in a morally proper manner (the control condition). Thus, the fictitious per-son,
77 denoted as a proposer, made an offer to influence the power holder’s decision. The task for the partici-
78 pants was to decide whether to accept or reject the offer made by the proposer. If the offer was accepted,
79 both the proposer and the participant would profit from the offer, whereas neither would earn any money
80 if the participant rejected the offer (see Fig. 2). Because mak-ing a decision in the bribe condition
81 additionally creates the ethical concern of colluding with a briber (which is not the case in the control
82 condition), this design allowed us to uncover the specific role of the rDLPFC in bribery-related decision-
83 making.

84 On the basis of our recent study on corruption and of previous literature that revealed a role of moral cost
85 on ethical decision-making, we hypothesized that participants would be generally less willing to accept
86 the offers in the bribe condition than in the control condition. More importantly, according to the tDCS
87 literature mentioned above, we expected that participants who received cathodal tDCS over the rDLPFC
88 would be more likely to accept offers in the bribe condition than would participants who received sham
89 stimulation in the control condition, especially when larger offers were proposed. In contrast, we did not
90 form a specific hypothesis about how anodal tDCS affects corrupt behaviors because of its mixed effect on
91 social and moral behaviors. Moreover, we tested several computational models and identified the one
92 that best characterized actual behaviors for all tDCS groups, which allowed us to delineate how rDLPFC
93 specifically contributes to the computations underlying corrupt acts.

94 **Method**

95 **Participants**

96 One hundred twenty French-speaking students from University of Lyon I and local residents (54 women;
97 age: $M = 22.4$ years, $SD = 4.4$) were recruited via online advertisements. The sample size was adopted on
98 the basis of previous tDCS studies on similar topics (Maréchal et al., 2017; Ruff et al., 2013), which are
99 standard in the field. All participants were psychiatrically and neurologically healthy and were not taking
100 any medications, as confirmed by a standardized clinical screening. The tDCS study was approved by the
101 local ethics committees. All experimental protocols and procedures were conducted in accordance with
102 institutional review board guidelines for experimental testing and complied with the latest revision of the
103 Declaration of Helsinki.

104 **Task and design**

105 Participants were randomly assigned to three tDCS treatment conditions with 40 persons in each: (a)
106 anodal stimulation (18 women; age: $M = 22.6$ years, $SD = 5.5$), (b) cathodal stimulation over the rDLPFC
107 (17 women; age: $M = 21.9$ years, $SD = 2.6$), or (c) sham stimulation (19 women; age: $M = 22.6$ years, $SD =$
108 4.8). Participants were blind to condition (see the Supplemental Material for the tDCS protocol). The main
109 experiment included a computerized incentive task and a follow-up paper-and-pencil rating task, which
110 lasted around 30 min in total (see the Supplemental Material for procedure details). In the computerized
111 task, participants were assigned the role of the power holder who decides to accept or reject financial
112 offers (see Fig. 2a). In a cover story, they were informed that they would be presented with a series of
113 choices from an independent group, whose data were collected previously by the experimenter.
114 Specifically, participants were led to believe that this independent group of online attendants (denoted as
115 proposers hereafter) played a game of chance. This independent group did not actually exist, and the
116 choices made by this group were predetermined by the task software. Each proposer was presented with
117 two options that would earn them different payoffs. The larger payoff ranged from €60 to €130 (see details
118 below), and the smaller payoff was fixed at €5. One of the two payoffs was randomly indicated by the
119 computer as the one to be received. According to the rules of the game, the proposer should report the
120 payoff indicated by the computer, which determined the final payoff (i.e., the control condition).
121 However, the response of the proposer was never checked by the experimenters. This allowed the
122 proposer to lie by reporting the alternative payoff that had not been indicated by the computer when this
123 would earn the proposer more profit (i.e., the bribe condition). In other words, the only difference
124 between the two conditions was that in the bribe condition, the proposer cheated for a larger payoff by
125 reporting the nonchosen larger payoff, whereas in the control condition, the proposer honestly reported

126 the chosen larger payoff. Importantly, participants were told that each proposer had been informed that
127 whether or not they obtained the payoff of the reported option crucially depended on the decisions of a
128 power holder (i.e., the participants themselves). To obtain the profits in the reported option, the proposer
129 could share a portion of the money from their potential gain (i.e., the reported larger payoff) to influence
130 the power holder's decision. The task for the power holder was to decide whether to accept or reject the
131 offer on the basis of the information above. If the power holder accepted the offer, both the power holder
132 and the proposer would benefit from the payoff. If the power holder rejected the offer, neither of them
133 earned anything. Participants were informed that they would be paid at the end of the experiment based
134 on one of their decisions in a randomly selected trial. Several aspects of this task merit additional notes.
135 First, participants were informed that each decision was independent, and we matched each decision with
136 different proposers to avoid possible learning effects or strategic responses. Second, each participant was
137 actually paid €30 at the end, as required by the ethics approval board. Finally, we designed the task so
138 the proposers always reported the option with a larger pay-off, so their personal profits after sharing with
139 the power holder were always more than the €5 option. This ensured that selfish motivation was the only
140 source that drove the proposer to cheat for a higher payoff and ruled out other motivations perceived by
141 participants that might influence their subsequent behaviors. We implemented a 3×2 mixed design by
142 manipulating the tDCS treatment (a between-subject factor) and the task condition (a within-subject
143 factor). Crucially, we operationally defined corrupt behaviors as the acceptance of offers made by the
144 proposer only when the proposer lied (the bribe condition). Compared with accepting offers in the control
145 condition, accepting offers in the bribe condition incurred the moral cost of colluding with the proposer's
146 dishonesty. We also manipulated the offer proportion, which was defined as the proportion of the amount
147 the proposer decided to share with the power holder from the payoff the proposer would have earned in
148 the reported option, which ranged from 10% to 90% (in steps of 10%; nine levels). This allowed us to
149 investigate whether and how the degree of temptation of a bribe modulated corrupt behaviors. To further
150 increase the variance of offers, we set potential gains that could be earned by the proposer (i.e., the larger
151 payoff, which ranged from €60 to €130 in steps of 10; eight levels). This yielded 72 trials, each involving a
152 unique offer, which appeared once in each condition. Each trial began with a screen displaying two payoff
153 options in the game of chance: the computer's choice (indicated by a computer icon) and the proposer's
154 offer. Participants were asked to decide whether to accept or reject the offer by pressing relevant buttons
155 with either the left or right index finger at their own pace. A yellow bar appeared below the corresponding
156 option for 0.5 s once the decision was made. Each trial ended with an intertrial interval of random duration
157 ($M = 1$ s; see Fig. 2b). The order of these trials was randomized across participants to reduce the
158 confounding effect of the condition order. In addition, the positions of payoffs were randomized within
159 participants, and those of the choice options were counterbalanced across participants. All stimuli were
160 presented using Presentation software (Version 14; Neurobehavioral Systems, 2009). After completing the
161 experiment, participants were asked to perform a follow-up rating task in which they reported their
162 subjective feelings about the task. Then they filled out a series of task-irrelevant control measures (see
163 the Supplemental Material for details). They were debriefed, paid, and thanked at the end of the
164 experiment.

165 166 **Data analyses**

167
168 One participant in the cathodal group was excluded because technical issues prevented complete data
169 recording, thus leaving a total of 119 participants whose data were further analyzed (overall: 54 women;

170 age: M = 22.4 years, SD = 4.5; anodal group: 18 women; age: M = 22.6 years, SD = 5.5; cathodal group: 17
171 women; age: M = 22.0 years, SD = 2.5; sham group: 19 females; age: M = 22.6 years, SD = 4.8). Overall,
172 participants did not report any uncomfortable feelings after the experiment and were not able to correctly
173 identify the treatment to which they were assigned, $\chi^2(1, N = 119) = 1.89, p = .169$. Because no difference
174 in age, $F(2, 116) = 0.26, p = .775$, or gender, $\chi^2(2, N = 119) = 0.13, p = .939$, was observed between tDCS
175 groups, we did not include these variables as covariates for later analyses. Behavioral analyses were
176 conducted using R (Versions 3.5.3 and 3.6.3; R Core Team, 2019, 2020). Model-based analyses were
177 performed using the hierarchical Bayesian approach via the hBayesDM package (Version 1.1.1; Ahn et al.,
178 2017). For method details, see the Supplemental Material.

179

180 **tDCS procedure**

181

182 The tDCS was administered using a multichannel stimulator (neuroConn, Munich, Germany) and pairs of
183 standard electrodes covered with conductive paste. On the basis of previous literature closely relevant to
184 the current study (Knoch et al., 2006; Strang et al., 2014), we designated our target site as the position
185 centering around the following Talairach coordinates: $x = 39, y = 37, z = 22$. This location approximately
186 corresponds to the electrode position of AF4 in the 10-10 electroencephalography (EEG) system (see Fig.
187 1, right; marked with a black circle). The vertex, which corresponded to the electrode position of Cz, was
188 chosen as the reference electrode on the basis of the study by Maréchal et al. (2017). To illustrate the
189 strength of the stimulation, we performed current-flow simulations with the realistic volumetric-approach
190 to simulate transcranial electric stimulation (ROAST) tool (Version 3.0; Huang et al., 2019;
191 <https://github.com/andypotatohy/roast>). For additional methodological details, see the Supplemental
192 Material.

193

194 **Results**

195

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

197

198 We first tested our main hypothesis regarding choice behavior. Using mixed-effect logistic regression, we
199 observed that participants were less likely to accept an offer in the bribe condition than in the control
200 condition—a main effect of task condition: $\chi^2(1, N = 17,136) = 126.94, p < .001$ —and more likely to do so
201 when the offer proportion increased—a main effect of offer proportion: $\chi^2(1, N = 17,136) = 96.34, p <$
202 $.001$. We also detected a significant two-way interaction between task condition and offer proportion,
203 $\chi^2(1, N = 17,136) = 33.05, p < .001$. Post hoc analyses indicated that participants in the bribe condition
204 were more likely to accept offers when the offer proportion increased than participants in the control
205 condition were ($z = 5.41, p < .001$). More importantly, we found a significant three-way interaction
206 between tDCS group, task condition, and offer proportion with respect to whether the offer was accepted,
207 $\chi^2(2, N = 17,136) = 8.04, p = .018$ (see Fig. 3). To follow up the three-way interaction, we performed post
208 hoc analyses on choice for each tDCS group. These analyses incorporated task condition, offer proportion,
209 and their interaction as fixed-effect predictors. We found that participants in the bribe condition who
210 received either type of tDCS stimulation were more likely to accept offers when the offer proportion
211 increased than participants in the control condition were (anodal: $z = 4.67, p < .001$; cathodal: $z = 4.34, p$
212 $< .001$), which was not the case in the sham group ($z = 0.67, p = .501$; see Table S1 in the Supplemental
213 Material for details). Notably, we did not observe any main effect of tDCS or related interaction on a series
214 of other behavioral measures, including decision time, task-related subjective ratings, and task-irrelevant
215 measures (see Fig. S2 and Tables S2–S4 in the Supplemental Material for details).

216

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

219
220 Bayesian model comparison showed that Model 1 (shown below) yielded the lowest leave-one-out infor-
221 mation criterion (LOOIC) scores and outperformed other competitive models (Models 2–4; see the
222 Supplemental Material for details):
223

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

$$\beta, \lambda, \gamma = \begin{cases} \beta_{\text{control}}, \lambda_{\text{control}}, \gamma_{\text{control}}, & \text{if control condition} \\ \beta_{\text{bribe}}, \lambda_{\text{bribe}}, \gamma_{\text{bribe}}, & \text{if bribe condition} \end{cases}$$

224
225
226
227
228 In this model, SV denotes the subjective value of the choice. PP and PPH represent the offer's payoff for
229 the proposer and power holder respectively, given different choices (i.e., to accept or reject the offer). β
230 and λ measure the decision weights on personal profits and proposer's gain from the offer, respectively; γ
231 measures the sensitivity to the absolute-payoff inequality between the power holder and the proposer.
232 The posterior pre-dictive check revealed that the proportion of accep-tance predicted by this model could
233 capture the proportion of observed acceptance across individuals (both conditions for all groups: $r_s > .99$,
234 $p_s < .001$; see Figs. S3–S7 in the Supplemental Material for the poste-rior predictive check at various levels),
235 which further justified the validity of our model. To examine how bribery-elicited moral cost affected each
236 parameter and how tDCS treatment modulated such effects, we implemented mixed-effects linear
237 regression on each parameter separately, including tDCS group, task condition, and their interactions as
238 the fixed-effect predictors. We also allowed intercepts to vary across participants as the random effects.
239 As a result, we first found a main effect of task condition for all three parameters, namely that participants
240 devalued the personal gains, β : $F(1, 116) = 18.04$, $p < .001$, $\eta p^2 = .092$; the proposer's gains, λ : $F(1, 116) =$
241 172.64 , $p < .001$, $\eta p^2 = .481$; and the absolute-payoff differences, γ : $F(1, 116) = 96.33$, $p < .001$, $\eta p^2 = .320$,
242 in the bribe con-dition relative to the control condition. Furthermore, we observed a main effect of tDCS
243 treatment on γ , $F(2, 116) = 20.42$, $p < .001$, $\eta p^2 = .166$. Post hoc analyses showed that participants in the
244 anodal group decreased their concern for the absolute-payoff differences rela-tive to participants in the
245 sham group, $t(116) = 3.05$, $p = .003$ (false-discovery-rate [FDR] corrected), Cohen's $d = 0.55$, 95%
246 confidence interval (CI) = [0.19, 0.92], which was even further reduced in the cathodal group (relative to
247 the anodal group), $t(116) = 3.35$, $p = .002$ (FDR corrected), Cohen's $d = 0.61$, 95% CI = [0.24, 0.98] (see the
248 Supplemental Material for details). More intriguingly, we found an interaction effect between tDCS group
249 and task condition on decision weights on personal gains, β : $F(2, 116) = 11.71$, $p < .001$, $\eta p^2 = .116$, and
250 absolute-payoff differences, γ : $F(2, 116) = 16.14$, $p < .001$, $\eta p^2 = .320$, but not on proposers' gains, λ : $F(2,$
251 $116) = 2.35$, $p = .100$, $\eta p^2 = .025$. Post hoc analyses for β showed that compared with participants who
252 received sham tDCS, participants who received cathodal tDCS had decreased weights on personal gains in
253 the control condition, $t(213) = -2.21$, $p = .042$ (FDR corrected), Cohen's $d = 0.59$, 95% CI = [-1.13, -0.06],
254 but they had increased weights in the bribe condition, $t(213) = 2.55$, $p = .035$ (FDR corrected), Cohen's $d =$
255 0.68 , 95% CI = [0.15, 1.22]. Anodal tDCS induced a similar effect of β in the control condition, $t(213) =$
256 -3.55 , $p = .001$ (FDR corrected), Cohen's $d = 0.95$, 95% CI = [-1.48, -0.41], but the enhancement effect was
257 not statistically significant in the bribe con-dition, $t(213) = 1.58$, $p = .172$ (FDR corrected), Cohen's $d = 0.42$,
258 95% CI = [-0.11, 0.95]. Regarding γ , post hoc analyses showed that compared with participants in the sham
259 group, participants in both the anodal group, $t(228) = 5.91$, $p < .001$ (FDR corrected), Cohen's $d = 1.42$, 95%
260 CI = [0.93, 1.91], and the cathodal group, $t(228) = 7.46$, $p < .001$ (FDR corrected), Cohen's $d = 1.80$, 95% CI
261 = [1.31, 2.29], were less aversive to absolute-payoff differences (i.e., the general inequality) in the control

262 condition. However, in the bribe condition, participants in the cathodal group were less averse to the
263 absolute-payoff inequality compared with both the sham group, $t(228) = 2.15$, $p = .049$ (FDR corrected),
264 Cohen's $d = 0.52$, 95% CI = [0.04, 1.00], and the anodal group, $t(228) = 3.45$, $p = .002$ (FDR corrected),
265 Cohen's $d = 0.83$, 95% CI = [0.35, 1.32]; see Figure 4 for the summary for key parameters; see Fig. S8 in the
266 Supplemental Material for the visualization of the tDCS effect on differential parameters; also see Tables
267 S5–S7 in the Supplemental Material for details of statistical analyses).

268 269 **Applying tDCS over rDLPFC modulates bribery-elicited moral cost on choice behaviors by mediating key** 270 **parameters of the computation**

271
272 To further establish the link between the tDCS treatment, the bribery-elicited moral cost on these param-
273 eters, and choice behaviors, we implemented post hoc mediation analyses with tDCS group as the
274 predictor, the differential parameters as the mediator (i.e., $\Delta\beta = \beta_{\text{bribe}} - \beta_{\text{control}}$, $\Delta\gamma = \gamma_{\text{bribe}} - \gamma_{\text{control}}$),
275 and the differential acceptance rate as the dependent variable (i.e., $\Delta_{\text{accept}} = \text{accept}_{\text{bribe}} -$
276 $\text{accept}_{\text{control}}$). A bootstrapping procedure was applied to the mediation effect (i.e., 5,000 boot-
277 strapped samples). We found that although the tDCS treatment did not directly modify the bribery-specific effect
278 on choice behaviors (i.e., total effect, path c: $p_s > .3$ for both tDCS effects), the differential parameters
279 mediated the impact of tDCS treatment on the bribery-specific effect on the behaviors (i.e., direct effect
280 [path c']: $p_s < .001$ in both tDCS effects for $\Delta\beta$ and in the anodal tDCS for $\Delta\gamma$, $p = .007$ in the cathodal tDCS
281 for $\Delta\gamma$; indirect effect [path ab] for $\Delta\beta$ —anodal: $b = -0.27$, 95% CI = [-0.40, -0.15]; cathodal: $b = -0.26$,
282 95% CI = [-0.39, -0.12]; indirect effect [path ab] for $\Delta\gamma$ —anodal: $b = 0.21$, 95% CI = [0.13, 0.30]; cathodal:
283 $b = 0.18$, 95% CI = [0.07, 0.28]; see Figure 5; also see Table S8 in the Supplemental Material for detailed
284 regression outputs).

285 286 **Discussion**

287
288 In the present study, we combined tDCS with a novel task that captured the essence of real-life bribery to
289 examine whether rDLPFC causally influences the corrupt behaviors of a power holder. As predicted,
290 participants were less likely to accept a bribe compared with a standard offer (i.e., the offer in the control
291 condition), even when the bribe became more tempting. These results are consistent with those of other
292 studies on moral decision-making (Crockett et al., 2014; Mazar et al., 2008; Qu et al., 2020) and confirm
293 the role of moral cost for power holders when they decide whether to take a bribe. Model-based analyses
294 further revealed how the computations made during bribery-related decision-making are influenced.
295 Specifically, participants depreciated personal gains (β) earned by taking the bribes, which replicates the
296 findings of our recent fMRI study on corruption (Hu et al., 2021). In addition, we also observed stronger
297 negative weights for both the proposer's gains (λ) and absolute differences between their payoffs (γ) in
298 the bribe condition than in the control condition. This aligns with previous findings showing contextual
299 modulation of subjective valuation to a partner (Bhanji & Delgado, 2014; Delgado et al., 2005) or to a
300 fairness concern (Gao et al., 2018; Hu et al., 2018). Together, the results of the present study reveal that
301 such bribery-elicited moral cost reshapes not only the valuation of self-profits but also other-regarding
302 interests and thus helps to prevent the power holder from being corrupted. More interestingly, the
303 disruption of rDLPFC (i.e., in both the anodal and cathodal groups) made participants, as power holders,
304 more likely to accept bribes (vs. standard offers) as the size of the prospective pay-off increased, but this
305 finding did not hold for the sham group. Importantly, this tDCS effect over rDLPFC did not influence other
306 measures (e.g., decision time, subjective ratings), suggesting that general cognitive or affective processes
307 are less likely to constitute the underlying mechanism. Taking a model-based approach, we further showed
308 that disrupting rDLPFC also alters the computations that contribute to bribery decisions.

309

310 Specifically, cathodal tDCS over rDLPFC mitigated the effect of the moral cost on personal gains due to
311 bribe taking ($\Delta\beta$). This finding is consistent with a previous brain-lesion study in which patients with lesions
312 of DLPFC selectively reduced the moral cost to personal profits (Zhu et al., 2014). Moreover, altering the
313 rDLPFC excitability via cathodal tDCS enhanced the effect of the bribery-elicited moral cost on fairness
314 concerns ($\Delta\gamma$). As noted previously, studies using a standard ultimatum game consistently showed that
315 inhibiting the rDLPFC by low-frequency repetitive transcranial magnetic stimulation (Knoch et al., 2006)
316 or cathodal tDCS (Knoch et al., 2008; Speitel et al., 2019) increases the tolerance of unfairness. Although
317 we replicated these findings by showing a less negative γ for the cathodal group than the sham group in
318 the control condition, we found that participants in the cathodal group become more aversive to the
319 inequity between them-selves and the proposer. Collectively, these results in the cathodal group indicate
320 a dual role of rDLPFC during bribery-related decision-making: It not only over-rides selfish motivation
321 when it conflicts with moral principles (Carlson & Crockett, 2018) but also integrates the moral cost in
322 modulating fairness concerns. This account is further supported by the mediation analyses, which
323 established the link between rDLPFC, computations underlying bribery-related decision-making, and final
324 behaviors. It is worth noting that the excitation of rDLPFC via anodal tDCS had a similar effect as cathodal
325 tDCS in modulating bribe-taking behaviors and the computations underlying bribery-related decision-
326 making. There is no a priori reason to believe that anodal and cathodal tDCS should induce opposite
327 behavioral effects in the moral domain. Indeed, previous evidence is mixed concerning the anodal effect
328 on moral behaviors, which varies in different paradigms. Although Maréchal et al. (2017) showed that
329 anodal tDCS over rDLPFC increased honesty in a die-rolling task, another tDCS study with an instrumental-
330 deception paradigm indicated the opposite effect (Mameli et al., 2010). In agreement with this, an fMRI
331 study has also shown that DLPFC is recruited more in dishonest individuals when they have a chance to
332 cheat (Greene & Paxton, 2009). Moreover, the classical polarity effect of tDCS (i.e., anodal excitation and
333 cathodal inhibition) has been shown to be much less common in the cognitive domain than in the motor
334 domain (Jacobson et al., 2012). A systematic review has revealed highly variable effects of tDCS over the
335 DLPFC on cognitive functions such as working memory (Tremblay et al., 2014). Such inconsistent effects
336 also exist in the social domain. For example, although inhibiting rDLPFC with cathodal tDCS consistently
337 enhances the tolerance to unfairness (Knoch et al., 2008; Speitel et al., 2019), no evidence suggests that
338 anodal tDCS increases fairness concerns (Speitel et al., 2019). Lastly, there are large individual variations
339 in tDCS effects on modulating behaviors (López-Alonso et al., 2014; Wiethoff et al., 2014) and in the
340 relationship between DLPFC engagement and moral behaviors (Hu et al., 2021; Yin & Weber, 2019).
341 Together, our findings confirm that the classical polarity effect of tDCS, originally observed in the primary
342 motor cortex, should not be expected to be directly applied to other brain areas and to social and moral
343 behaviors such as corruption. Some limitations of the present study should be noted. First, bribery-elicited
344 moral cost merits further consideration. In our task, taking bribes was presumed to carry the only moral
345 cost, that of colluding in fraud. In the control condition, no fraud was taking place, and therefore the offer
346 was not considered to be a bribe. However, it is likely that an extra moral cost might be involved simply
347 because of the action of accepting bribes. Because of the present design, it is impossible to isolate this
348 putative moral cost because it always covaries with the other moral cost. Second, because our sample
349 consisted of healthy adults mainly of college age, researchers should be cautious about generalizing these
350 findings to individuals who actually hold power in companies or governmental agencies, who are usually
351 older. Future studies are needed to address these issues. Overall, the present study provides empirical evi-
352 dence that perturbing rDLPFC via tDCS causally influences a power holder's decisions of whether to accept
353 a bribe and modifies the computations underlying bribery-related decision-making. These findings shed
354 light on the neurobiological substrates of corrupt acts and open a new window to investigate corruption
355 using a multidisciplinary research approach.

356

357

358 Transparency

359
360 Action Editor: Daniela Schiller Editor: Patricia J. Bauer Author Contributions R. Philippe, V. Guigon, and S.
361 Zhao contributed equally to this study. Y. Hu and J.-C. Dreher conceived of and designed the study. R.
362 Philippe, V. Guigon, and S. Zhao collected the data. Y. Hu analyzed the data. Y. Hu wrote the first draft of
363 the manuscript, and J.-C. Dreher, J. J. Bonaiuto, E. Derrington, and B. Corgnet made critical edits. All
364 authors approved the final manuscript for submission. Declaration of Conflicting Interests The author(s)
365 declared that there were no conflicts of interest with respect to the authorship or the publication of this
366 article. Funding This research was funded by the IDEXLYON project of the Université de Lyon (Project
367 INDEPTH) under the Pro-gramme Investissements d’Avenir (ANR-16-IDEX-0005), by a grant from the LABEX
368 CORTEX project of the Université de Lyon (Grant No. ANR-11-LABX-0042) under the pro-gram
369 Investissements d’Avenir (Grant No. ANR-11- IDEX-007) by Agence Nationale de la Recherche (ANR) to J.-
370 C. Dreher, and by a grant from the China Postdoctoral Science Foundation (2019M660007) to Y. Hu. Open
371 Practices All data and analysis code have been made publicly avail-able via OSF and can be accessed at
372 <https://osf.io/ve837/>. The design and analysis plan for the experiment were not preregistered. This article
373 has received the badges for Open Data and Open Materials. More information about the Open Practices
374 badges can be found at [http://www .psychologicalscience.org/publications/badges](http://www.psychologicalscience.org/publications/badges).
375

References

- Abbink, K. (2006). Laboratory experiments on corruption. In S. Rose-Ackerman (Ed.), *International handbook on the economics of corruption* (pp. 418–437).
- Edward Elgar. Ahn, W.-Y., Haines, N., & Zhang, L. (2017). Revealing neurocomputational 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. <https://doi.org/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, USA, 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), 1611–1618. <https://doi.org/10.1038/nn1575>
- Dreher, A., Kotsogiannis, C., & McCorrison, 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* (2nd ed., 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 in equity aversion. *Proceedings of the National Academy of Sciences, USA*, 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, USA*, 106(30), 12506–12511.
- Hu, Y., He, L., Zhang, L., Wölk, T., Dreher, J.-C., & Weber, B. (2018). Spreading inequality: Neural computations underlying paying-it-forward reciprocity. *Social Cognitive and Affective Neuroscience*, 13(6), 578–589. <https://doi.org/10.1093/scan/nsy040>
- Hu, Y., Hu, C., Derrington, E., Corghet, B., Qu, C., & Dreher, J.-C. (2021). Neural basis of corruption in power-holders. *eLife*, 10, Article e63922. <https://doi.org/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), Article 056006. <https://doi.org/10.1088/1741-2552/ab208d>
- 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), 71–85.
- 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., Nordio, F., Consonni, D., Sartori, G., & 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. <https://doi.org/10.1016/j.bbr.2010.03.024>

- 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, USA*, 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.
- Neurobehavioral Systems. (2009). Presentation (Version 14) [Computer software]. www.neurobs.com
- 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*, 15(2), 135–149.
- Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556. <https://doi.org/10.1038/nrn2357>
- R Core Team. (2019). R: A language and environment for statistical computing (Version 3.5.3) [Computer software]. <https://www.r-project.org/>
- R Core Team. (2020). R: A language and environment for statistical computing (Version 3.6.3) [Computer software]. <https://www.rproject.org/>
- 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. (Eds.) (2012). *Research in experimental economics: Vol. 15. New advances in experimental research on corruption*. Emerald Group.
- Speitel, C., Traut-Mattausch, E., & Jonas, E. (2019). Functions of the right DLPFC and right TPJ in proposers and responders in the ultimatum game. *Social Cognitive and Affective Neuroscience*, 14(3), 263–270. <https://doi.org/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. <https://doi.org/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. (2019). I lie, why don't you: Neural mechanisms of individual differences in self-serving lying. *Human Brain Mapping*, 40(4), 1101–1113. <https://doi.org/10.1002/hbm.24432>
- Zhu, L., Jenkins, A. C., Set, E., Scabini, D., Knight, R. T., Chiu, P. H., King-Casas, B., & Hsu, M. (2014). Damage to dorso lateral prefrontal cortex affects tradeoffs between honesty and self-interest. *Nature Neuroscience*, 17(10), 1319–1321. <https://doi.org/10.1038/nn.3798>

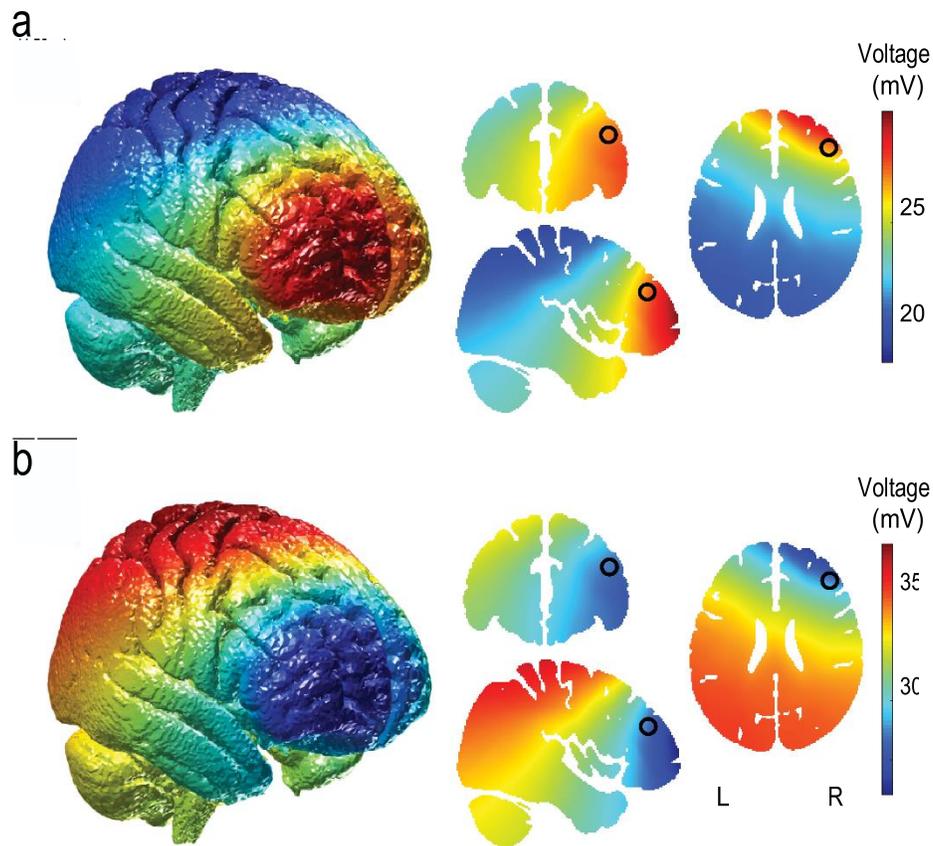


Fig. 1.

Electric field simulation for (a) anodal and (b) cathodal transcranial direct-current stimulation (tDCS). The position centering around the Talairach coordinate of $x = 39$, $y = 37$, $z = 22$ (marked with a black circle in the images on the right) was chosen as the target site. This location approximately corresponds to the electrode position of AF4 in the 10-10 electroencephalography (EEG) system. The vertex was chosen as the reference electrode and corresponds to the electrode position of Cz. The voltage indicates strength of tDCS across the whole brain. L = left; R = right.

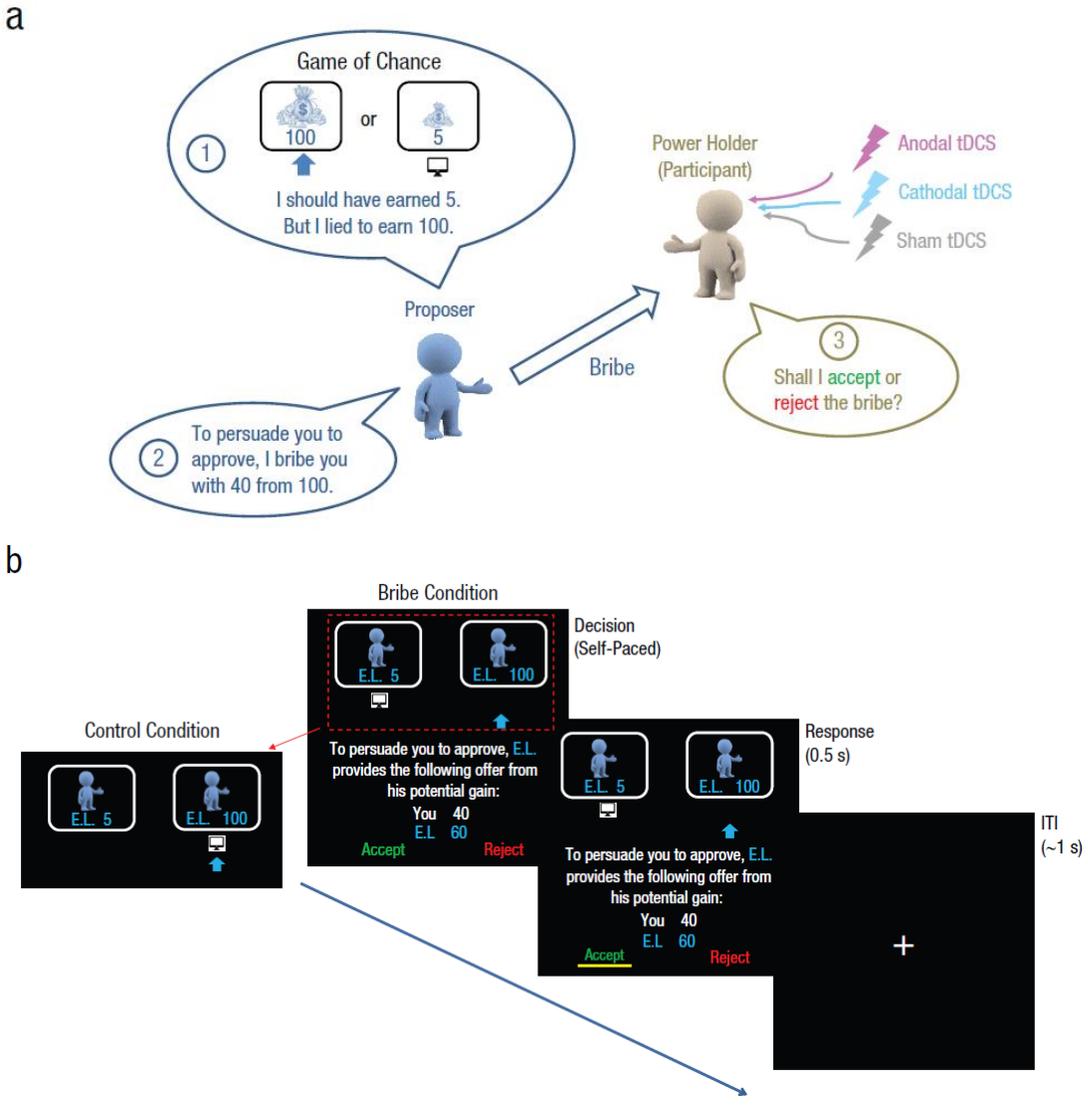


Fig. 2.

Illustration of the transcranial direct-current stimulation (tDCS) manipulation and behavioral paradigm (a) and an example trial sequence (b). All participants were assigned randomly to three tDCS groups (i.e., anodal, cathodal, or sham). The task involved two roles: a proposer (i.e., a fictitious participant in a previous online study in which a game of chance was played) and a power holder (i.e., the real participant in the current study). In the control condition, the proposer truthfully reported the larger payoff selected by the computer. In the bribe condition, shown here in (a), 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. In the example trial from the bribe condition (b), a proposer (“E.L.”) lied by reporting the nonselected 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 their potential gain (i.e., €40 out of €100). 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 below the corresponding option for 0.5 s to highlight the choice, which was followed by an intertrial interval (ITI) with a fixation cross (M = 1 s, range = 0.6–1.4 s). 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).

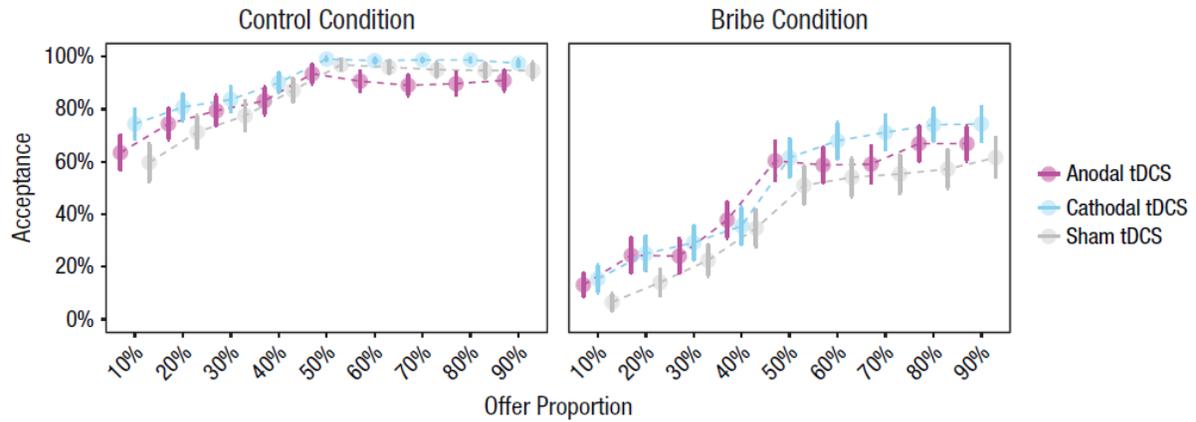


Fig. 3.

Mean acceptance rate of the standard offer (control condition) and bribes (bribe condition) as a function of transcranial direct-current stimulation (tDCS) group (anodal, cathodal, or sham) and offer proportion (10% to 90% in steps of 10%). Error bars represent standard errors of the mean.

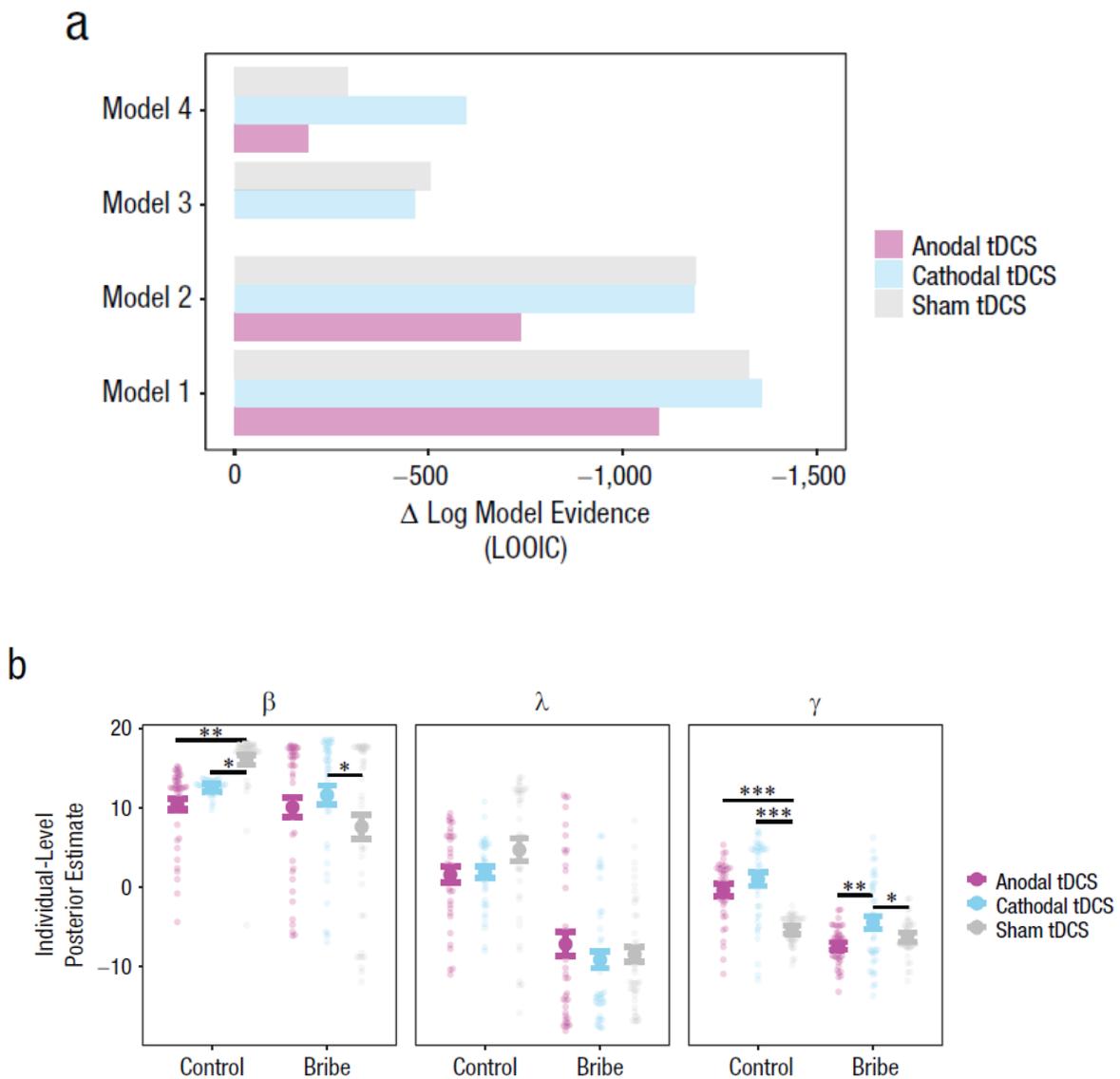


Fig. 4.

Model-based results. Bayesian evidence for each of the four models across the three transcranial direct-current stimulation (tDCS) groups (a) was calculated as the difference between the model's own leave-one-out information criterion (LOOIC) score and that of the model with the worst accuracy of out-of-sample prediction (in this case, Model 2 of the anodal group). The posterior mean of individual-level key parameters of the winning model (Model 1) is shown in (b) as a function of condition and tDCS group. The parameters 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 dot represents the group-level mean; each smaller dot represents the data of a single participant. Error bars represent standard errors of the mean. Asterisks indicate between-group differences (* $p < .05$, ** $p < .01$, *** $p < .001$; all p s false-discovery-rate corrected).

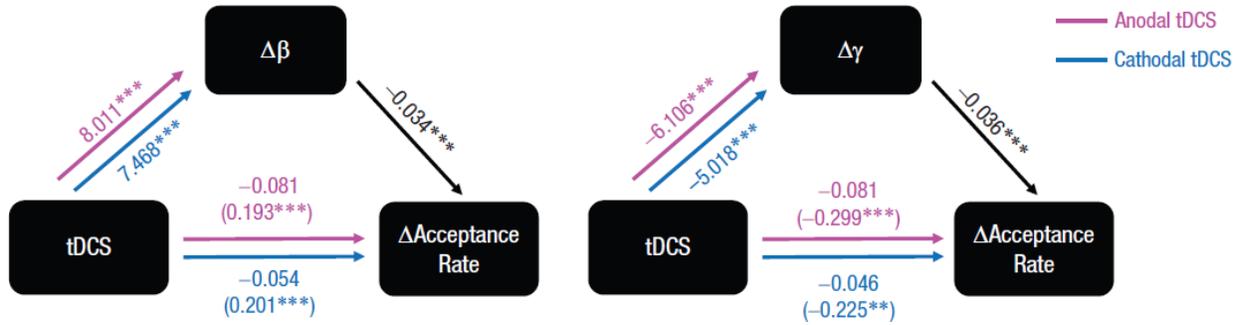


Fig. 5.

Results of the mediation analysis showing the influence of receiving transcranial direct-current stimulation (tDCS) on the differential acceptance rate of the offer (bribe vs. control), as mediated by the differential parameters $\Delta\beta$ (left) and $\Delta\gamma$ (right). Unstandardized coefficients are shown; differently colored coefficients on paths a and c show results for each type of tDCS separately. On the path from tDCS to differential acceptance rate, values outside parentheses reflect total effects, and values inside parentheses reflect direct effects after controlling for the mediator. Five thousand bootstrap samples ($N = 5,000$) were used to test the significance of the indirect effect. Asterisks indicate significant paths (** $p \leq .01$, *** $p \leq .001$).