Sciencexpress Report

EMBARGOED UNTIL 2:00 PM US ET THURSDAY, 3 OCTOBER 2013

Changing Social Norm Compliance With Noninvasive Brain Stimulation

C.C. Ruff,[1]* G. Ugazio,[1,2] E. Fehr[1]*

[1]Laboratory for Social and Neural Systems Research (SNS-Lab), Department of Economics, University of Zurich, Zurich, Switzerland. [2]Social, Cognitive and Affective Neuroscience Unit, University of Vienna, Vienna, Austria.

*Corresponding author. E-mail: christian.ruff@econ.uzh.ch (C.C.R.); ernst.fehr@econ.uzh.ch (E.F.)

All known human societies have maintained social order by enforcing compliance with social norms. The biological mechanisms underlying norm compliance are, however, hardly understood. We show that the right lateral prefrontal cortex (rLPFC) is involved in both voluntary and sanction-induced norm compliance. Both types of compliance could be changed by varying neural excitability of this brain region with transcranial direct current stimulation, but they were affected in opposite ways, suggesting that the stimulated region plays a fundamentally different role in voluntary and sanction-based compliance. Brain stimulation had a particularly strong effect for compliance based on socially constituted sanctions, while it left beliefs about what the norm prescribes and about subjectively expected sanctions unaffected. Our findings suggest that rLPFC activity is a key biological prerequisite for an evolutionarily and socially important aspect of human behavior.

Human societies depend crucially on social norms that specify the range of permissible actions for a given situation. Social norms range from the mundane (e.g., dress codes, table etiquette) to the profound (e.g., collective action, bilateral exchange, law obedience). They are considered a hallmark of human civilization because no other known species regulates social interactions to the same degrees by norms (*1*–*3*). The potential of norms to guide collective behavior can break down if norm violations are not sanctioned, because humans tend to follow prevailing norms conditional on observing others' compliance (*4*). All known human societies have therefore enforced norm compliance by threatening norm violators with punishment, both officially via legal codes and institutions, and informally in the context of private sanctions through peers (*5*, *6*). The importance of credible sanctioning threats for maintaining norm compliance is well established by ethnographic evidence (*1*, *2*), evolutionary theory (*1*, *3*), and laboratory experiments (*5*, *6*).

It has been proposed that the human brain may have developed neural processes that support norm enforcement by generating appropriate behavioral responses to social punishment threats (*7*–*10*). However, neuroscience studies on social norms have mostly focused on the neural basis of punishing others (*11*–*14*), whereas evidence for neural circuitry underlying sanction-induced compliance with norms is scarce. In mature adults, a brain network involving an area in the right lateral prefrontal cortex (rLPFC) is activated during norm-compliant behavior triggered by social punishment threats (*10*). However, it is not possible to conclude from correlative fMRI findings that norm compliance depends causally on neural activity in the rLPFC (*15*). Establishing such a causal dependence is crucial for our understanding of how social norm compliance develops in the context of brain maturation (*16*) and how it is pathologically altered and therapeutically amenable in the context of brain disorders (*9*).

We employed transcranial direct current stimulation (tDCS) (*17*) to examine whether social norm compliance depends causally on neural processing in the previously-identified rLPFC region (*10*). Participants engaged via computer terminals in anonymous social interactions that had real financial consequences. In every round, participants ("player A") received an amount of money units (MUs) and decided how much of it to transfer to a randomly assigned anonymous opponent ("player B"). In baseline rounds, this transfer was implemented, whereas in punishment rounds, player B could respond to the transfer by reducing player A's MUs [*Fig. 1, fig. S1, and supplementary* materials (SM) (*18*)]. In Western cultures, a fairness norm (*19*–*21*) prescribes to split the "cake" of MUs equally between both players. This conflicts with player A's self-interest motive to keep as many MUs as possible. In baseline rounds, player A thus typically transfers only around 10% - 25% of the MUs. In contrast, when a sanctioning threat is present, player A largely obeys the fairness norm and transfers around 40% - 50% of the MUs (*10*, *20*). The transfer difference between punishment and baseline rounds thus indexes sanction-induced norm compliance, i.e., the degree to which the sanction threat induces player A to change her transfer from the level of voluntary norm-compliance as measured in baseline rounds.

Individual differences in sanction-induced norm compliance correlate with fMRI-measured activity in the rLPFC (*10*). Based on this finding and the rLPFC's general role in the control of behavior (*22*, *23*), it has been proposed that the rLPFC may weigh fair versus selfish responses specifically when punishment threats are present (*8*, *10*). To provide causal evidence for this hypothesis, we first identified the specific rLPFC region described in (*10*) using MR-scans of 63 female participants; we then experimentally altered neural excitability in this brain area during behavioral performance in a double-blind, placebo-controlled tDCS design (SM and fig. S2). tDCS can both increase or decrease neural excitability in the stimulated region, depending on the polarity of the current flow (*17*). We thus randomly sorted participants into three stimulation groups where neural excitability in the rLPFC was enhanced with anodal tDCS, reduced with cathodal tDCS, or left unaltered by sham/placebo tDCS as control for possible non-neural effects of stimulation (see SM). Such non-neural effects did not differ between the groups (see SM) and therefore could not account for performance in the norm-compliance paradigm.

Participants were sensitive to the punishment threat and transferred more money in punishment than in baseline rounds (mean transfer difference 29.44 MUs; $P < 0.001$, GLS regression). However, in line with our hypothesis, the two active brain stimulation conditions changed sanction-induced norm compliance in opposite ways relative to the sham condition (Fig. 2A and table S2). Anodal tDCS increased the transfer difference by 33.5% (GLS regression, $p < 0.001$) whereas cathodal tDCS decreased the transfer difference by 22.7% ($P < 0.001$).

Do these effects reflect changes in altruistic behavior, with increased (decreased) monetary transfers regardless of punishment threats? This interpretation is refuted by the data on voluntary norm-compliance in baseline rounds (Fig. 2B and table S3). Voluntary transfers were actually decreased (GLS regression, $P < 0.001$) during anodal tDCS and increased ($P < 0.01$) during cathodal tDCS, relative to the sham condition. This not only confirms that tDCS affected subjects' response to the punishment threat but that these tDCS effects on sanction-induced compliance were actually stronger than the opposite effects on voluntary compliance: If tDCS had not affected sanction-induced compliance then overall transfers in punishment rounds – which are based on voluntary plus sanction-induced compliance – should also be lower after anodal and higher after cathodal stimulation. However, overall transfers in punishment rounds were in fact higher (GLS regression, $P < 0.05$) during anodal tDCS and lower ($P < 0.001$) during cathodal tDCS than in the sham condition (fig. S3).

Which task-related psychological mechanisms may have contributed to the tDCS effect? To respond appropriately, participants need to know the fairness norm and form appropriate beliefs about player B's reactions. We measured (i) the participants' perceived fairness, (ii) the anger they expected the opponent to feel, and (iii) the punishment they expected at different transfer levels (Fig. 3). All participants were clearly aware of the fairness norm and rated higher transfers as significantly fairer (ANOVA, $F(2,60) = 84.88$, $P < 0.001$), less likely to cause anger in the opponent ($F(2,60) = 218.96$, $P < 0.001$), and leading to lower punishment ($F(2,60) = 82.69$, $P < 0.001$). Importantly, the type of brain stimulation did *not* affect participants' beliefs, neither on average (all $F(2,60) < 0.94$, all $P > 0.39$) nor in their change across different transfer levels (all $F(2,60) < 0.55$, all $P > 0.74$).

Our findings do not yet show that the stimulated rLPFC region implements specifically social aspects of behavioral control. In particular, behavior in punishment rounds requires risk taking and trading off higher transfers with a lower risk of sanction. We therefore repeated the experiment in a sample of 59 new female volunteers who took the identical decisions as before, but now played against a computer pre-programmed to respond in the same way as a human opponent in punishment rounds (see SM). In this "non-social context," participants were also sensitive to punishment threats (fig. S4A) but the effects of tDCS on sanction-induced transfers were significantly weaker than during interactions with human opponents (Fig. 4A and table S3). This held for both increases in sanction-induced transfers due to anodal tDCS (GLS regression, $P = 0.009$) and decreases due to cathodal tDCS ($P = 0.001$, GLS regression). In baseline rounds of the non-social context – where no social norm prescribes sharing MUs with the computer – participants hardly transferred any MUs (fig. S4B). Such (possibly erroneous) *voluntary* transfers to the computer were therefore also less affected by tDCS than norm-related voluntary transfers to human opponents (Fig. 4B; GLS regression, $P < 0.05$ for anodal tDCS and $P < 0.001$ for cathodal tDCS).

Social punishment is thought to have played an important role for the evolution of human social behavior and cooperation (*1*–*3*). Our results show that the influence of punishment threats on human social norm compliance depends causally on neural activity in the rLPFC. This suggests a neural mechanism involving the rLPFC that aligns behavior with social norms when punishment is possible. The more pronounced involvement of this mechanism for genuinely social punishments concurs with suggestions that during human brain evolution, the steep increase in the complexity of social interactions may have shaped specific neural processes for social behavior (*8*, *24*). That tDCS affected sanction-induced and voluntary norm compliance in opposite ways suggests that these two forms of norm compliance involve distinct neural circuits; in particular, the rLPFC seems to play a fundamentally different role in voluntary and sanction-based norm compliance.

Our finding that rLPFC stimulation did not affect awareness of the fairness norm and expected sanctions suggests that the rLPFC process necessary for norm-compliant behavior is dissociated from neural mechanisms enabling humans to anticipate sanctions for norm violations and to distinguish "right" from "wrong." The rLPFC mechanism necessary for norm-compliance is probably not restricted to neural activity within this brain area, given that prefrontal cortex is involved in many aspects of behavioral control (*23*) and that brain stimulation can affect areas interconnected with the stimulation site (*25*). The anatomical connectivity (*26*) and context-dependent functions of prefrontal cortex (*27*) make it more likely that the stimulated rLPFC area integrates and coordinates activity in a network of brain regions triggered by the need for considering social punishments during action control (*8*).

Brain stimulation studies in humans have so far mostly shown unidirectional, maladaptive effects on decision making, rendering participants more impulsive (*28*), selfish (*29*), or cognitively biased (*30*). Such interventions may therefore be of limited practical use in applied settings. Our finding that changes in the neural excitability of rLPFC can enhance voluntary and sanction-induced social norm compliance may be of relevance because non-compliance with social norms constitutes a major problem in psychiatric (*31*) and neurological (*31*, *32*) disorders, during abnormal development in adolescence (*33*), and in adults in the form of criminal activity (*9*). However, the opposite influence of brain stimulation on voluntary and sanction-induced norm compliance also suggests that increasing one type of norm compliance with brain stimulation may come at the cost of decreasing the other type.

References and Notes

1. E. Sober, D. S. Wilson, *Unto Others - The Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, Cambridge, MA, 1998).
2. S. Mathew, R. Boyd, Punishment sustains large-scale cooperation in prestate warfare. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11375–11380 (2011). doi:10.1073/pnas.1105604108 Medline
3. R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531–3535 (2003). doi:10.1073/pnas.0630443100 Medline
4. U. Fischbacher, S. Gächter, E. Fehr, Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001). doi:10.1016/S0165-1765(01)00394-9
5. R. Forsythe, H. L. Horowitz, N. E. Savin, M. Sefton, Fairness in simple bargaining experiments. *Games Econ. Behav.* **6**, 347–369 (1994). doi:10.1006/game.1994.1021
6. E. Fehr, S. Gächter, Altruistic punishment in humans. *Nature* **415**, 137–140 (2002). doi:10.1038/415137a Medline
7. P. R. Montague, T. Lohrenz, To detect and correct: Norm violations and their enforcement. *Neuron* **56**, 14–18 (2007). doi:10.1016/j.neuron.2007.09.020 Medline
8. J. W. Buckholtz, R. Marois, The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nat. Neurosci.* **15**, 655–661 (2012). doi:10.1038/nn.3087 Medline
9. A. Raine, Y. Yang, Neural foundations to moral reasoning and antisocial behavior. *Soc. Cogn. Affect. Neurosci.* **1**, 203–213 (2006). doi:10.1093/scan/nsl033 Medline
10. M. Spitzer, U. Fischbacher, B. Herrnberger, G. Grön, E. Fehr, The neural signature of social norm compliance. *Neuron* **56**, 185–196 (2007). doi:10.1016/j.neuron.2007.09.011 Medline
11. A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, The neural basis of economic decision-making in the Ultimatum Game. *Science* **300**, 1755–1758 (2003). doi:10.1126/science.1082976 Medline
12. D. J. de Quervain, U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, E. Fehr, The neural basis of altruistic punishment. *Science* **305**, 1254–1258 (2004). doi:10.1126/science.1100735 Medline
13. J. W. Buckholtz, C. L. Asplund, P. E. Dux, D. H. Zald, J. C. Gore, O. D. Jones, R. Marois, The neural correlates of third-party punishment. *Neuron* **60**, 930–940 (2008). doi:10.1016/j.neuron.2008.10.016 Medline
14. M. J. Crockett, A. Apergis-Schoute, B. Herrmann, M. D. Lieberman, U. Müller, T. W. Robbins, L. Clark, Serotonin modulates striatal responses to fairness and retaliation in humans. *J. Neurosci.* **33**, 3505–3513 (2013). doi:10.1523/JNEUROSCI.2761-12.2013 Medline
15. For example, individuals with particular personality traits - such as anxious individuals or those with the tendency for Machiavellian behavior - may have generally higher LPFC activity and a generally higher propensity to respond to sanctions, but apart from their co-variation, these two variables may not directly influence one another. Alternatively, it is also possible that rather than being the cause of norm compliance, brain activation in rLPFC is merely the consequence of norm compliance. In other words, individuals who respond more strongly to the sanctioning threat may recruit rLPFC more strongly as a result of their choice.
16. N. Steinbeis, B. C. Bernhardt, T. Singer, Impulse control and underlying

functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron* **73**, 1040–1051 (2012). doi:10.1016/j.neuron.2011.12.027 Medline
17. M. A. Nitsche, W. Paulus, Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol.* **527**, 633–639 (2000). doi:10.1111/j.1469-7793.2000.t01-1-00633.x Medline
18. Materials and methods, supplemental analyses, and supplementary figures are available as supplementary materials on *Science* Online.
19. E. Fehr, U. Fischbacher, Third-party punishment and social norms. *Evol. Hum. Behav.* **25**, 63–87 (2004). doi:10.1016/S1090-5138(04)00005-4
20. C. F. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction*. (Princeton Univ. Press, Princeton, NJ, 2003).
21. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, In search of *Homo economicus*: Behavioral experiments in 15 small-scale cocieties. *Am. Econ. Rev.* **91**, 73–78 (2001). doi:10.1257/aer.91.2.73
22. A. R. Aron, T. W. Robbins, R. A. Poldrack, Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* **8**, 170–177 (2004). doi:10.1016/j.tics.2004.02.010 Medline
23. E. K. Miller, J. D. Cohen, An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001). doi:10.1146/annurev.neuro.24.1.167 Medline
24. R. I. M. Dunbar, The social brain meets neuroimaging. *Trends Cogn. Sci.* **16**, 101–102 (2012). doi:10.1016/j.tics.2011.11.013 Medline
25. J. Driver, F. Blankenburg, S. Bestmann, W. Vanduffel, C. C. Ruff, Concurrent brain-stimulation and neuroimaging for studies of cognition. *Trends Cogn. Sci.* **13**, 319–327 (2009). doi:10.1016/j.tics.2009.04.007 Medline
26. P. L. Croxson, H. Johansen-Berg, T. E. Behrens, M. D. Robson, M. A. Pinsk, C. G. Gross, W. Richter, M. C. Richter, S. Kastner, M. F. Rushworth, Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. *J. Neurosci.* **25**, 8854–8866 (2005). doi:10.1523/JNEUROSCI.1311-05.2005 Medline
27. J. Duncan, The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends Cogn. Sci.* **14**, 172–179 (2010). doi:10.1016/j.tics.2010.01.004 Medline
28. B. Figner, D. Knoch, E. J. Johnson, A. R. Krosch, S. H. Lisanby, E. Fehr, E. U. Weber, Lateral prefrontal cortex and self-control in intertemporal choice. *Nat. Neurosci.* **13**, 538–539 (2010). doi:10.1038/nn.2516 Medline
29. D. Knoch, A. Pascual-Leone, K. Meyer, V. Treyer, E. Fehr, Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314**, 829–832 (2006). doi:10.1126/science.1129156 Medline
30. G. Xue, C. H. Juan, C. F. Chang, Z. L. Lu, Q. Dong, Lateral prefrontal cortex contributes to maladaptive decisions. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4401–4406 (2012). doi:10.1073/pnas.1111927109 Medline
31. A. Bechara, M. Van Der Linden, Decision-making and impulse control after frontal lobe injuries. *Curr. Opin. Neurol.* **18**, 734–739 (2005). doi:10.1097/01.wco.0000194141.56429.3c Medline
32. B. King-Casas, C. Sharp, L. Lomax-Bream, T. Lohrenz, P. Fonagy, P. R. Montague, The rupture and repair of cooperation in borderline personality disorder. *Science* **321**, 806–810 (2008). doi:10.1126/science.1156902 Medline
33. D. P. Farrington, R. Loeber, *Child Adol Psych Cl* **9**, 733 (2000).
34. M. B. Iyer, U. Mattu, J. Grafman, M. Lomarev, S. Sato, E. M. Wassermann, Safety and cognitive effect of frontal DC brain polarization in healthy individuals. *Neurology* **64**, 872–875 (2005). doi:10.1212/01.WNL.0000152986.07469.E9 Medline
35. B. Fritsch, J. Reis, K. Martinowich, H. M. Schambra, Y. Ji, L. G. Cohen, B. Lu, Direct current stimulation promotes BDNF-dependent synaptic plasticity: Potential implications for motor learning. *Neuron* **66**, 198–204 (2010). doi:10.1016/j.neuron.2010.03.035 Medline

Acknowledgments: C.C.R. and E.F. designed research; C.C.R. and G.U. conducted research; C.C.R. and G.U. analyzed data with input from E.F.; C.C.R. and E.F. wrote the paper with comments by G.U. We thank K. Treiber and A. Schlaepfer for their assistance with data collection. This study was supported by an ERC grant to EF (Foundations of Economic Preferences), Swiss National Science Foundation (SNSF) grants to CCR and EF (CRSII3_141965 and 51NF40_144609), and the SNSF National Competence Center for Research (NCCR) in Affective Sciences.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1241399/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S4
References (*34*, *35*)

3 June 2013; accepted 24 September 2013
Published online 3 October 2013; 10.1126/science.1241399

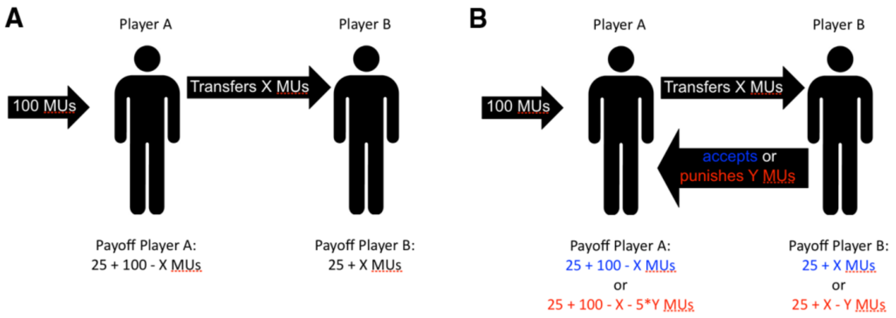

Fig. 1. **Economic game used to measure social norm compliance.** In each round, both players receive 25 money units (MUs). Player A is given an additional 100 MUs that she can share with player B by sending a transfer X (in multiples of 10 MUs). All experimental MUs are exchanged into real money at the end of the experiment. Two types of rounds are presented in random order. (**A**) Baseline round: Transfer X is implemented as proposed, measuring player A's voluntary norm compliance. (**B**) Punishment round: Player B can either accept X (blue font) or invest Y MUs from her initial endowment to punish player A (red font). Y can be any integer between 0 and 25, reducing A's payoff by 5*Y MUs. Player A is aware of this possible sanction; any increase in transfers for punishment relative to baseline rounds therefore measures sanction-induced norm compliance.

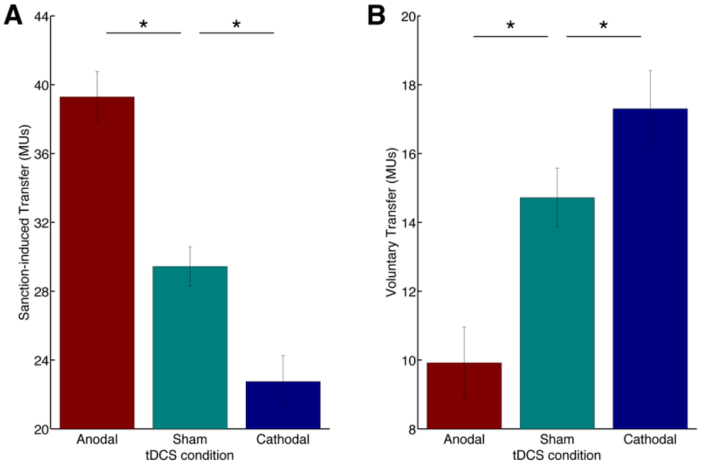

Fig. 2. rLPFC stimulation changes sanction-induced and voluntary norm compliance. (**A**) Sanction-induced norm-compliance: Average (+/− s.e.m.) transfer difference for punishment rounds minus baseline rounds. Higher values indicate that the punishment threat led to a larger adjustment of transfers toward the fairness norm of an equal split. (**B**) Voluntary norm compliance: Average (+/− s.e.m.) transfers for baseline rounds. All values determined with regression in eq. S1 (SM) ; *$P < 0.05$.

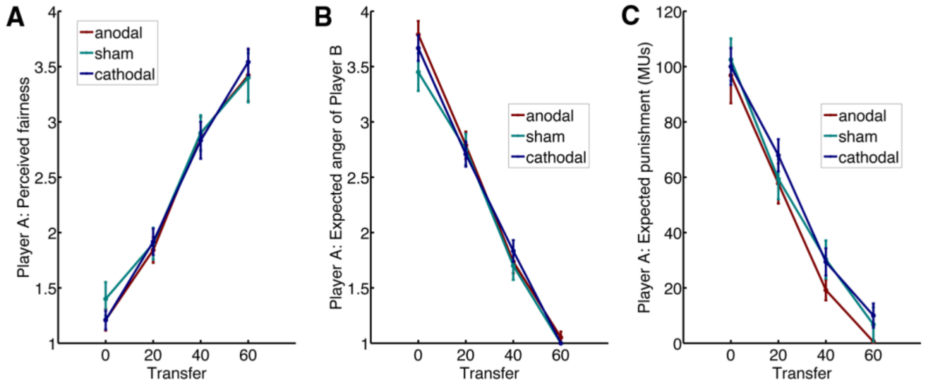

Fig. 3. rLPFC stimulation does not affect participants' beliefs about the fairness of different transfers and about player B's anticipated anger and expected punishment. (**A**) Average rating of perceived fairness for different transfer levels (scale from 1/"very unfair" to 4/"very fair"). (**B**) Average rating of anticipated anger felt by player B for different transfer levels (scale from 1/"not angry at all" to 4/"very angry"). **(C)** Average expected payoff reduction resulting from B's punishment. Error bars represent s.e.m.

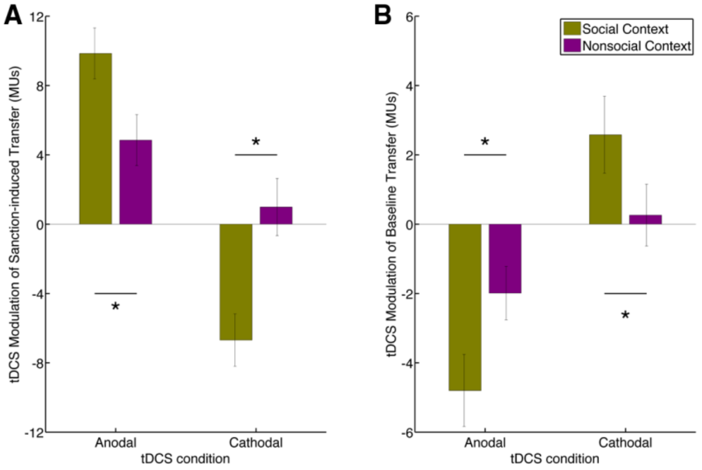

Fig. 4. rLPFC stimulation effects are stronger during social interactions. (**A**) tDCS effects on sanction-induced norm compliance during interactions with a human (Social Context) or a computer opponent (Non-social Context). Bars depict average changes in transfer difference for anodal and cathodal tDCS relative to the sham condition. (**B**) tDCS-related changes of voluntary transfers in baseline rounds. Bars represent average changes for anodal and cathodal tDCS relative to the sham condition. All values determined with regression in eq. S2 (SM); *$P < 0.05$.