doi:10.1093/scan/nsu114 SCAN (2015) 10, 790–796

Be nice if you have to — the neurobiological roots of strategic fairness

Sabrina Strang,[1,*] Jörg Gross,[2,3,*] Teresa Schuhmann,[2] Arno Riedl,[3,4,5] Bernd Weber,[1,6] and Alexander T. Sack[2]

[1]Center for Economics and Neuroscience, University of Bonn, 53127 Bonn, Germany, [2]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, The Netherlands, [3]Department of Economics (AE1), School of Business and Economics, Maastricht University, 6200 MD Maastricht, The Netherlands, [4]Center for Economic Studies (CESifo), 81679 München, Germany, [5]Institute for the Study of Labor (IZA), 53113 Bonn, Germany and [6]Department of Epileptology, University Hospital Bonn, 53105 Bonn, Germany

Social norms, such as treating others fairly regardless of kin relations, are essential for the functioning of human societies. Their existence may explain why humans, among all species, show unique patterns of prosocial behaviour. The maintenance of social norms often depends on external enforcement, as in the absence of credible sanctioning mechanisms prosocial behaviour deteriorates quickly. This sanction-dependent prosocial behaviour suggests that humans strategically adapt their behaviour and act selfishly if possible but control selfish impulses if necessary. Recent studies point at the role of the dorsolateral prefrontal cortex (DLPFC) in controlling selfish impulses. We test whether the DLPFC is indeed involved in the control of selfish impulses as well as the strategic acquisition of this control mechanism. Using repetitive transcranial magnetic stimulation, we provide evidence for the causal role of the right DLPFC in strategic fairness. Because the DLPFC is phylogenetically one of the latest developed neocortical regions, this could explain why complex norm systems exist in humans but not in other social animals.

Keywords: decision-making; prosocial behaviour; strategic fairness; social norms

INTRODUCTION

Humans among all animals are unique in their ability to establish highly complex social norm systems (Tomasello and Rakoczy, 2003; Fehr and Rockenbach, 2004; Gintis, 2003; Ostrom, 2000; Sethi and Somanathan, 1996; Fehr and Fischbacher, 2003). Fairness and cooperation norms often demand to restrict immediate self-interest in favour of benefits of the group, the society or another individual in need. The widespread prevalence of such norms in human societies is puzzling from an evolutionary perspective (Fehr and Fischbacher, 2004; Melis and Semmann, 2010), as they are informal, often vaguely defined and, as such, should be easy to circumvent. Especially in large groups with anonymous interactions, free-riding should dominate (Bowles and Gintis, 2003). Indeed, experiments have shown that without credible punishment threats, fair and cooperative behaviour, like sharing with others or contributing to a group project, can deteriorate quickly (Fehr and Gächter, 2002; Egas and Riedl, 2008; Gächter *et al.*, 2008; Ule *et al.*, 2009). On the other hand, there is convincing evidence that fair and cooperative behaviour can emerge and be maintained when there is the threat that freeriding will be sanctioned (Fehr and Gächter, 2000, 2002; Fehr and Fischbacher, 2004; Spitzer *et al.*, 2007; Egas and Riedl, 2008; Gächter *et al.*, 2008; Ule *et al.*, 2009). This indicates that humans are sensitive to punishment threats, enabling them to act selfishly when they can and to act strategically fairly when they have to.

The neurobiological basis of this ability to adapt behaviour strategically and thereby controlling immediate selfish impulses has been explored in recent studies. These suggest that activity in the right prefrontal cortex is associated with the control of selfish impulses (Wout *et al.*, 2005; Knoch *et al.*, 2006; Knoch and Fehr, 2007; Knoch *et al.*, 2009) and the ability to adapt behaviour strategically (Spitzer *et al.*, 2007; Ruff *et al.*, 2013). Knoch *et al.* (2006) found that the disruption of the right dorsolateral prefrontal cortex (DLPFC), using transcranial magnetic stimulation (TMS), led participants to accept an offer that yielded a higher financial payoff for themselves much more frequently than to reject it in favour of a financially less attractive but fair outcome. A functional magnetic resonance imaging (fMRI) study by Spitzer *et al.* (2007) showed that acting fairly because of strategic reasons is correlated with increased activity in the right DLPFC. Most recently, using transcranial direct current stimulation (tDCS), Ruff *et al.* (2013) demonstrated that suppressing neural excitability (with cathodal tDCS) of the right lateral prefrontal cortex (LPFC) led to a lower degree of strategic fairness, whereas enhancing neural excitability (with anodal tDCS) of the right LPFC increased strategically fair behaviour. Interestingly, cathodal tDCS over right LPFC also decreased immediate selfish responses, while enhancing the right LPFC using anodal tDCS led to a higher degree of immediate selfishness.

This result pattern is intriguing and yet puzzling at the same time. The fact that suppressing neural activity of the right LPFC with cathodal tDCS decreases immediate selfishness is in conflict with an earlier result of Knoch *et al.* (2006) who found an increase of immediate selfishness when disrupting the dorsal part of the right LPFC using repetitive TMS (rTMS). However, these two studies differ in various methodological aspects, which might account for this apparent contradiction. Ruff *et al.* (2013) applied a different intervention method and used a broader target region. The strength of the study by Ruff *et al.* (2013) clearly lies in the differential effects revealed on strategic fairness as well as immediate selfishness contrasting anodal (enhancing) *vs* cathodal (suppressing) tDCS over the same brain region, i.e. the right LPFC. This focus on the right LPFC naturally neglected the possible contribution of and comparison with left LPFC, as tested, e.g.in Knoch *et al.* (2006). Moreover, all mentioned studies used between-subject experimental designs, which allow inferences on the population level but do not allow investigating individual differences in these effects across stimulation conditions. Hence, although these studies strongly suggest that the right LPFC is functionally relevant for decisions that

Received 24 December 2013; Revised 18 August 2014; Accepted 1 September 2014
Advance Access publication 3 September 2014

*These authors contributed equally to this work.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. [263472]; granted to A.T.S.). The authors would like to thank Franziska Dambacher and Jeannetta Boschma for their assistance.

Correspondence should be addressed to Sabrina Strang, Center for Economics and Neuroscience, University of Bonn, Nachtigallenweg 86, 53127 Bonn, Germany. E-mail: strang@uni-bonn.de

involve trade-offs between immediate selfish goals on the one hand and fair and cooperative behaviour on other hand, it remains necessary to further investigate to what extent the right, and not the left, LPFC is involved in controlling immediate selfish impulses and strategically fair behaviour. In the current study we used a within-subject design applying rTMS (or sham) either to the right or left DLPFC of male participants in order to test on the individual level whether the right and/or left DLPFC are causally linked to, first, the control of immediate selfishness and, second, the strategic acquisition of this control mechanism when the threat of norm enforcement demands it. Moreover, to explore whether a shift in beliefs or perception could explain the results, we elicited participants' beliefs about norm enforcement and perception of the fairness of behaviours while under the different TMS and sham conditions.

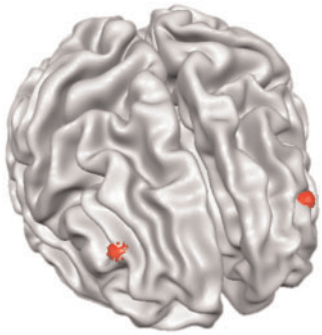

Fig. 1 Target area for the magnetic brain stimulation. Each target was selected based on the individual anatomical image obtained in a separate MRI measurement. The red dots represent the two target points in Talairach space: $x = \pm 39$, $y = 37$, $z = 22$.

MATERIALS AND METHODS

Subjects

We studied 17 male participants (mean age 23.5 years, ranging between 20 and 41 years) with normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. None of the participants had taken part in a TMS experiment before. They received medical approval for participation and gave written informed consent after being instructed about the procedure. The study was approved by the local Medical Ethical Commission.

TMS-procedure

Participants were tested in four sessions separated by at least 1 week. In the first session a T1-weighted magnetic resonance image was acquired. The other three sessions were TMS sessions. Each participant took part in each of the three TMS conditions (left DLPFC, right DLPFC and sham). The condition order was counterbalanced across participants.

A surface reconstruction on the MRI images was made to recover the spatial surface of the cortical sheet based on the white-grey matter boundary using Brain Voyager QX 2.4 (BrainInnovation, Maastricht, The Netherlands). We then identified the right and left DLPFC based on the coordinates established by Sanfey *et al.* (2003) and Knoch *et al.* (2006); $x = \pm 39$, $y = 37$, $z = 22$, radius = 6 (Figure 1). The coordinates, given in Talairach space, were transformed to each participant's individual brain space.

Biphasic TMS pulses were applied using the MagVenture R30 stimulator (MagVenture A/S, Farum, Denmark) and a figure-of-eight coil (MC-B70, inner radius 10 mm, outer radius 50 mm). The maximum output of this coil and stimulator combination is ~1.9 T and 150 A/μS. At the beginning of the first TMS session, individual resting motor thresholds (RMTs) were determined. The mean RMT was 34.6% (s.d. = 3.7), ranging from 28 to 40% of maximal stimulator output (MSO). Stimulation intensity was applied at 110% RMT.

For sham stimulation, a figure-of-eight placebo coil (MC-BP70) was used. This coil produced the same acoustic stimulation as the active coil while not inducing a magnetic field. The coil was manually held tangentially to the skull over the right/left DLPFC using the online visualization function of Brain Voyager TMS Neuronavigation. Participants received 15 min, 1 Hz rTMS (900 pulses) offline stimulation over the left or right DLPFC. Sham stimulation was applied either over the left DLPFC or right DLPFC, balanced over participants. Participants were told beforehand that intensity of the TMS stimulation could vary across sessions. In the debriefing, we asked participants how the different TMS conditions might have affected their behaviour. None of the participants indicated any directed hypotheses.

Task

Resembling the task used in Spitzer *et al.* (2007), two different games were used, a standard Dictator Game (DG) and a Dictator Game with punishment option (DGp). In both games, two players, a dictator and a recipient, interact with each other. Each player receives an initial endowment of 25 monetary units (MUs, 1 MU = 0.16 euro cents). Additionally, the dictator receives 100 MUs and can distribute these freely between himself and the recipient. In DG, the recipient is passive and the game ends after the dictator has made a decision. In DGp, the recipient can punish the dictator after being informed about the distribution. To punish the dictator, the recipient has to spend his own MUs. For every MU the recipient uses for punishment, the dictator's payoff is reduced by 5 MUs. Thus, in case the dictator does not transfer any MUs and the recipient applies maximum punishment by spending his 25 MUs, both participants end up with 0 MUs.

Experimental procedure

Dictators and recipients were invited separately. First, 60 recipients were invited to the BEElab (Behavioural & Experimental Economics laboratory, Maastricht). They received written instructions about the rules of both games (DG and DGp). In the instructions, dictators and recipients were neutrally labelled as player A and B. To maximize the number of observations per recipient, we implemented the so-called strategy method (Selten, 1967). Recipients were asked how many of their MUs they would spend for punishment for every possible transfer the dictator could make. That is, we gathered a punishment response for each possible dictator transfer. Importantly, this method does not imply that punishment choices of recipients are hypothetical. The chosen punishment is real, as it has real monetary consequences for the recipient as well as the dictator. Specifically, depending on how much the dictator later actually chose to transfer to the recipient, the corresponding punishment decision was selected and final payoffs were calculated and paid accordingly (Brandts and Charness, 2011). Additionally a photo was taken of every recipient.

The 17 participants invited to the TMS sessions were allocated the role of dictators. They received written instructions about the rules of both games (DG and DGp) and were asked to answer a set of comprehension questions before the TMS stimulation. At the beginning of each session, after TMS stimulation, they saw a group picture of the 20 recipients who they would be paired with in the following 20 rounds. This was done to emphasize that in each round they interact with a different real person and that each decision bears real consequences. In half of the rounds, participants were paired with recipients who could punish (DGp condition), and in the other half, they were paired with

recipients who could not punish (DG condition). These conditions were randomized over rounds and participants were informed about the condition by a symbol on the computer screen. Hence, in the DG condition, participants knew that they would not face any consequences for acting selfishly, whereas in the DGp condition, punishment by the recipient could decrease their payoffs substantially. In each round, participants were asked how much, if any, of the 100 MUs they wanted to transfer to the other participant. To avoid learning effects, dictators received no feedback about the punishment decisions of recipients until the end of the experiment. We opted for not giving feedback because providing feedback would have had the problematic downside that potential TMS effects on learning could not have been disentangled from the effects we are interested in. Experienced punishment often has a larger impact on behaviour than imagined punishment. In repeated public goods game experiments with punishment, for example, first round cooperation (imagined punishment) is often smaller than cooperation in later rounds (experienced punishment; Fehr and Gächter, 2000, 2002; Egas and Riedl, 2008). Thus, by giving no feedback we are likely observing a lower bound of the effect of (potential) punishment on dictator transfers. This means that the incentive for acting strategically fairly is on the lower side and, therefore, inhibiting effects of our TMS intervention are likely to be on the conservative side too.

To test whether fairness perception or punishment beliefs were systematically affected by our TMS manipulation, participants saw five hypothetical transfers (from 0 to 50 MUs in steps of 10) from a hypothetical dictator and were asked to make fairness judgments for each transfer. Furthermore they were asked about their punishment expectation and own punishment expenses, were they in the role of the recipient.

Decisions in both games and the elicitation of fairness perceptions, punishment beliefs and hypothetical own punishment expenses were completed within 5–6 min after the rTMS stimulation. For each session, 1 of the 20 rounds was randomly selected and paid out in cash. Participants knew about this procedure upfront and were informed about the selected rounds and the associated earnings after the last session.

Analysis

Dictator transfers are censored below by zero. We therefore fitted random-intercept Tobit regressions (Gelman and Hill, 2007) to the data using R and JAGS (see Lunn *et al.*, 2009). In each regression model, variables, coding the session number as well as the sequence of the conditions were added to control for potential learning and order effects (see Supplementary Information). To test whether participants behave more selfishly when TMS is applied over the right DLPFC compared with sham and TMS over the left DLPFC, transfer decisions in the DG without punishment were regressed on dummy predictors coding the three TMS condition (sham condition as baseline).

To test the causal involvement of the right and/or left DLPFC in the ability to act strategically fairly, we first classified participants into 'adapters' and 'non-adapters'; those who gave more in the DGp during sham were classified as 'adapters' and those who gave less or equal were classified as 'non-adapters'. For each participant the transfer difference across DG and DGp as a measure for strategic adaption was calculated and regressed on dummy predictors coding the three TMS condition (sham condition as baseline) and a dummy variable indicating whether a dictator was a non-adapter to test for changes in strategic fairness across TMS conditions for adapters and non-adapters separately.

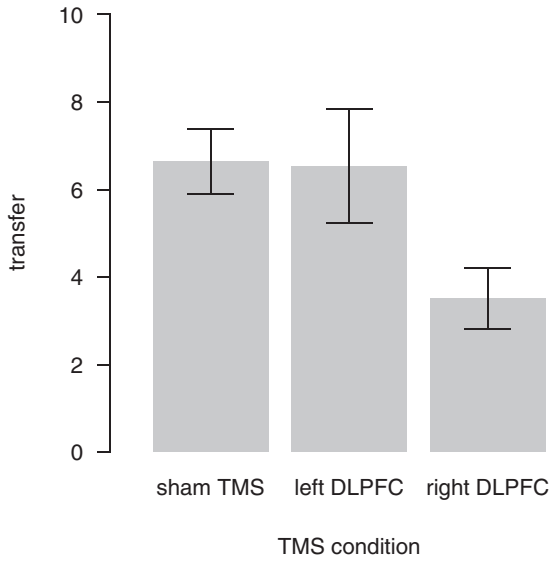

Fig. 2 Average transfers in the DG condition. Mean transfers to recipients without the ability to punish (DG condition) for each TMS session. Error bars show the within-subject standard errors of the mean (Cousineau, 2005; Morey, 2008).

RESULTS

On average, dictators transferred relatively little, although significantly more than zero, to recipients in the DG during sham (average transfer: 6.6, one sample *t*-test, $t(16) = 2.7$, $P < 0.05$, two-sided). Transfer rates are smaller than observed in some other DG experiments but similar to experiments with large social distance between dictators and recipients (Hoffman *et al.*, 1996; Camerer, 2003). The low transfers in sham give little room for observing the hypothesized effect of increased selfishness when inhibiting the right or left DLPFC. Nevertheless, we find that, on average, participants gave significantly less (almost 50%; 3 MUs on average) to recipients when TMS was applied to the right DLPFC compared with TMS over the left DLPFC and sham, respectively (see Figure 2; random-intercept Tobit regression, rDLPFC dummy, 95% confidence interval (CI): −14.4 to −3.9, $P < 0.05$). There was no significant difference in transfers between sham and TMS over the left DLPFC (random-intercept Tobit regression, lDLPFC dummy, 95% CI: −6.9 to 3.1).

To test our second hypothesis, whether the right DLPFC is also causally involved in the ability to act strategically fairly, we analysed the change in strategic adaption over the TMS conditions. Not all participants showed strategic adaption during sham. Four participants gave constantly nothing, regardless of DG and TMS condition. These participants also reported that they did not expect any punishment from the recipients for unfair transfers. One participant offered very low amounts to the recipients (between 0 and 15 MUs) and did not change offers across DG and DGp and one participant actually gave more to recipients without punishment power (DG) than to recipients with punishment power (DGp). However, a majority of the dictators (11 of 17) adapted strategically during sham and were classified as 'adapters'. On average, with sham TMS, adapters transferred 6 MUs in DG but a 5-fold of it (32 MUs) in DGp.

Figure 3 shows how this strategic adaption was affected by the disruption of the right and left DLPFC by plotting the mean transfers of DGp and DG for the three TMS conditions. When facing a recipient with punishment ability, participants who adapted strategically during sham did so significantly less when TMS was used to disrupt the right DLPFC compared with sham (random-intercept regression, rDLPFC

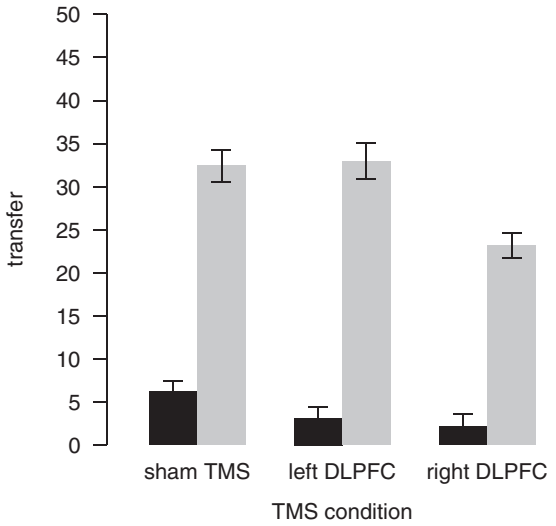

Fig. 3 Strategic adaption across TMS conditions. Mean transfers of adapters in DG (black) and DGp (grey) across the TMS conditions. Error bars show the within-subject standard errors of the mean (Cousineau, 2005; Morey, 2008).

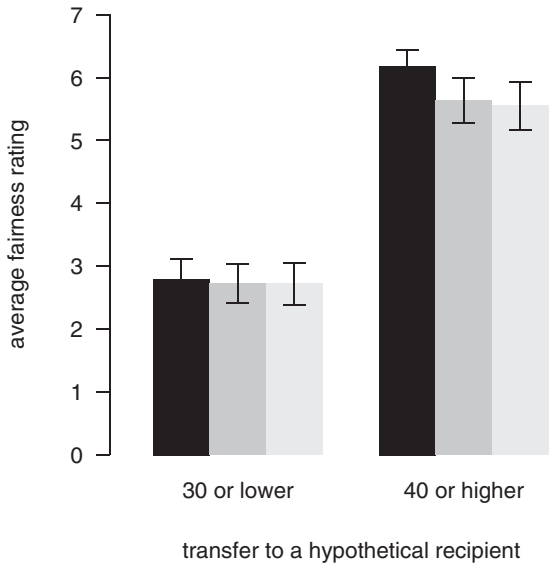

Fig. 4 Fairness judgments. Fairness judgments of adapters (from 1 = 'very unfair' to 7 = 'very fair') for low (≤30) and high (≥40) hypothetical transfers after each TMS session (black: sham; dark grey: TMS over the left DLPFC; light grey: TMS over the right DLPFC). Error bars show the within-subject standard errors of the mean (Cousineau, 2005; Morey, 2008).

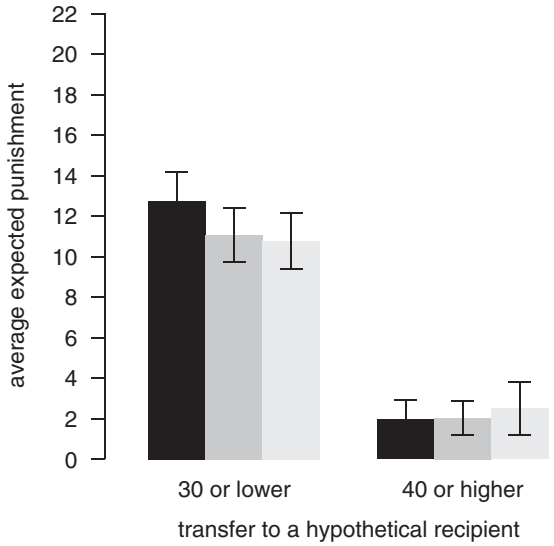

Fig. 5 Expected punishment. Expected punishment of adapters (from 0 to 25 punishment points) for low (≤30) and high (≥40) hypothetical transfers after each TMS session (black: sham; dark grey: TMS over the left DLPFC; light grey: TMS over the right DLPFC). Error bars show the within-subject standard errors of the mean (Cousineau, 2005; Morey, 2008).

dummy, 95% CI: −8.0 to −4.6, $P<0.05$). We observed the highest strategic adaption when the left DLPFC was disrupted. However, the difference to sham stimulation was not significant (random-intercept regression, lDLPFC dummy, 95% CI: −0.6 to 4.3).

Most recipients were willing to use their MUs to punish unfair behaviour by the dictators. Consequently, by giving less to recipients with punishment ability during TMS over the right DLPFC, dictators earned 4.1 MUs less on average in each interaction compared with sham (random-intercept regression, rDLPFC dummy, 95% CI: −4.7 to −3.3, $P<0.05$).

In principle, it could be possible that TMS stimulation alters the judgement of how (un)fair an offer is and/or the belief about how likely it is that unfair offers will be punished. To test for this we explored whether fairness judgements or beliefs about punishment were different across the three TMS conditions. As expected, we find that adapting dictators judge offers to be fairer the more is offered to the recipient (random-intercept regression, offer predictor, 95% CI: 0.10 to 0.15, $P<0.05$). These fairness judgements did not significantly change across the TMS conditions (see Figure 4; random-intercept regression, offer × left TMS, 95% CI: −0.05 to 0.02; offer × right TMS, 95% CI: −0.04 to 0.03). Regarding punishment, during sham, adapting dictators expected that unfair offers would be punished more severely (see Figure 5; random-intercept regression, offer predictor, CI: −0.56 to −0.37, $P<0.05$). Importantly, this expectation did not change significantly across the TMS conditions (random-intercept regression, offer × right, CI: −0.05 to 0.22; offer × left, CI: −0.04 to 0.22).

Hence, neither fairness judgements nor expected punishment of adapters were significantly affected by the TMS conditions. Interestingly, in comparison with adapters, non-adapters reported different fairness judgements as well as punishment beliefs. They showed a significantly smaller increase of rated fairness for increasing offers (random-intercept regression, non-adapter × offer, CI: −0.18 to −0.10, $P<0.05$), which did not change significantly across TMS conditions (random-intercept regression, non-adapter × offer × left, CI: −0.01 to 0.10; non-adapter × offer × right, CI: −0.03 to 0.08). Compared with adapters, they also believed that dictators are punished significantly less severely in general (random-intercept regression, non-adapter dummy, CI: −43.87 to −15.29, $P<0.05$). Thus, fairness judgements of non-adapters were less sensitive to changes in transfers and they generally expected less punishment.

When dictators were asked to imagine to be in the role of the receiver confronted with unfair transfers by a hypothetical dictator, adapters reported that they would be less willing to spend their MUs for punishment while the right DLPFC was disrupted by TMS, as compared with sham and TMS over the left DLPFC (Figure 6, random-intercept Tobit regression, rDLPFC dummy, 95% CI: −7.8 to −1.3, $P<0.05$). Non-adapters indicated that they would punish significantly less compared with adapters in general (random-intercept regression, non-adapter dummy, CI: −60.1 to −4.9, $P<0.05$).

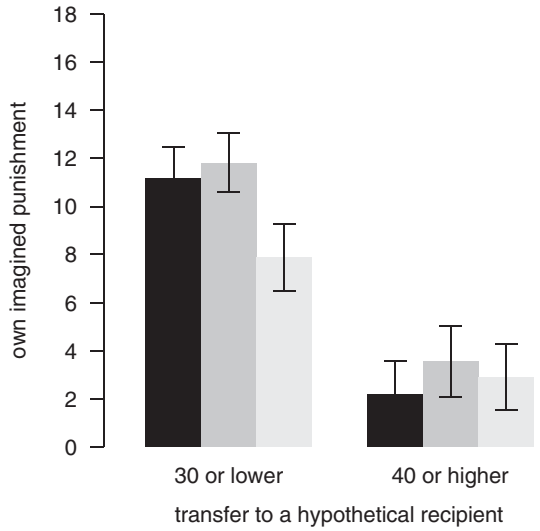

Fig. 6 Own imagined punishment. Own imagined punishment of adapters (from 0 to 25 punishment points) for low (≤30) and high (≥40) hypothetical transfers after each TMS session (black: sham; dark grey: TMS over the lDLPFC; light grey: TMS over the rDLPFC). Error bars show the within-subject standard errors of the mean (Cousineau, 2005; Morey, 2008).

DISCUSSION

Our results reveal that experimental perturbation of the right, but not the left, DLPFC systematically altered, first, the degree of prosocial behaviour and, second, the ability to act strategically fairly. The first result is consistent with previous findings by Knoch *et al.* (2006) who studied responses to fair or selfish behaviour. We extent their finding and show that even when subjects can actively decide to behave selfishly or fairly they are on average more selfish when the rDLPFC is disrupted.

The average transfers during sham in the DG were relatively low compared with other studies. Behaviour in the DG is known to be sensitive to framing (List, 2007; Bardsley, 2008). Especially social distance has been shown to influence transfer rates (Hoffman *et al.*, 1996; Leider *et al.*, 2009). Leider *et al.* (2009) showed that dictators give significantly more to receivers that are socially close. Hoffman *et al.* (1996) used a double-blind procedure that maximizes social distance and find transfer rates of ∼6–8%, similar to the rates we observed in the DG with sham TMS. In our experiment, receivers and dictators were invited separately to the experiment, and receivers were, thus, not present when dictators made their transfer decisions. This certainly increased social distance between dictators and receivers and might explain the relatively low transfers observed in our study. Further, it has been shown that when participants have to exert effort to earn their endowment, transfer rates decrease (Cherry *et al.*, 2002; Oxoby and Spraggon, 2008). In our study, dictators had to come to the lab four times (three times more often than the receivers) and had to undergo the TMS procedure three times. Thus, compared with the receivers, the effort they invested was much higher, and dictators might have thought that they deserved to keep more.

The relatively low offers of dictators in the DG with sham TMS left little room for a decrease of offers due to disruption of the right or left DLPFC. The fact that we nevertheless find a significant decrease in offers after TMS over the right DLPFC in comparison with sham TMS and TMS over the left DLPFC is reassuring for the interpretation that the rDLPFC controls selfish impulses not only when responding to offers (Knoch *et al.*, 2006) but also when actively making offers.

During sham and TMS over the left DLPFC, most dictators adapted their behaviour strategically by transferring more of their money to recipients who were able to punish them. Similar to DG transfers, our DGp transfers were lower than in comparable studies (Spitzer *et al.*, 2007; Ruff *et al.*, 2013), and similar reasons as discussed above might play a role. Moreover, in the DGp, participants did not receive immediate feedback about punishment and, hence, could not experience but only anticipate punishment. It is conceivable that participants thought that punishment threats are not credible and without feedback they could not update their beliefs. Some participants indeed reported that they did not expect any punishment from the recipients for unfair transfers. These participants mostly responded rationally to their beliefs and did not adapt their behaviour to the punishment threat. The majority of participants, however, reported that they expected to be punished for unfair offers and also that punishment would increase with the unfairness of the offer. These participants also adapted their behaviour accordingly in the sham TMS condition. However, when disrupting the right DLPFC, the same dictators not only shared their money less generously with recipients but also showed significantly less strategic adaption in their unfair behaviour in case a punishment threat was present. Thus, participants who consistently adapted their behaviour strategically during sham and TMS over the left DLPFC were less capable of doing so when TMS was applied to the right DLPFC. Our study therefore provides causal evidence for the functional role of the right DLPFC not only in overriding immediate selfish impulses but also in acting strategically fairly, an ability paramount for obeying fairness norms.

Interestingly, and in contrast to Knoch *et al.* (2006), where the disruption of the right DLPFC led to an increase in earnings, this lack of adaption was maladaptive because recipients were willing to spend their money to punish unfair behaviour, which decreased the payoff of the dictators substantially. Hence, by disrupting the right DPLFC participants not only failed to comply with widely shared fairness norms but thereby also failed to maximize their own payoff.

Our results also indicate that the failure to adapt strategically is neither explained by altered fairness perception nor by changes in beliefs about recipients' punishment behaviour as suggested by Sanfey *et al.* (2014). In line with previous findings (Knoch *et al.*, 2006; Ruff *et al.*, 2013), fairness judgments were not significantly affected by TMS, indicating that disrupting the right DLPFC impaired the control of selfish impulses without altering fairness perception. This suggests that fairness perception and decisions on complying with a fairness norm are to some degree independently represented in the brain, enabling us to know what is right, but do otherwise. There is also no evidence that beliefs about recipients' punishment behaviour were affected by our TMS intervention. Across all three TMS conditions, participants either believed that there would be no or little punishment (non-adapters) or that there would be punishment and that it would increase with the unfairness of offers (adapters). These results indicate that perturbation of the right DLPFC can alter strategic behaviour, but not the underlying motive or belief system that led to strategic fairness in the first place. Importantly, participants under TMS of the right DLPFC reported that they themselves would use less money for punishing unfair transfers. That is, although perceiving small transfers as unfair, participants in the right DLPFC TMS condition indicated not to be willing to spend money to punish unfair behaviour of others. Consistent with the results of Knoch *et al.* (2006), this suggests that even in the role of the recipient, participants would act more selfishly by withholding costly punishment. This implies that less social norm violations would be punished and more selfish behaviour would be tolerated, pointing to a possible involvement of the right DLPFC not only in norm compliance but also in the enforcement of norms that demand the restriction of selfish behaviour. Further research needs to be conducted to investigate the role of the DLPFC in norm enforcement directly.

An inherent characteristic of a within-subject design is that participants have to engage in a task repeatedly. This may lead to memory effects or habit formation, which may influence behaviour in later sessions. We controlled for this by counterbalancing the order of TMS conditions and also controlled for it in our statistical analysis. Evidence that order effects may only be of limited importance also comes from experiments showing that when participants restart an experimental task, behaviour is similar to the one in the previous task. This so-called restart effect was first observed by Andreoni (1988) and has been replicated numerous times (Andreoni and Croson, 2008). Hence, memory effects or habit formation are unlikely to confound our results.

Ruff *et al.* (2013) employed the same paradigm (DG and DGp) and investigated differences in strategic adaption across groups of female participants, while decreasing and increasing the neural excitability of the right LPFC using cathodal and anodal tDCS. Using a different intervention method and male participants, our findings are mostly in agreement with their results. They show that strategic adaption is significantly lower when decreasing excitability of the right LPFC, while fairness judgements are unaffected. Transfers in the DG condition were generally higher in their sample but adaption rates were comparable in size. Ruff *et al.* (2013) had an additional non-social control condition, showing that strategic adaption is only altered in a social context. In contrast to Ruff *et al.* (2013) who report an increase in transfers in the DG when disrupting the right LPFC with cathodal tDCS, we find that transfers to recipients who cannot punish (DG condition) significantly decrease when disrupting the right DLPFC. This finding is also in line with previous findings of increased selfishness (Wout *et al.*, 2005; Knoch *et al.*, 2006; Spitzer *et al.*, 2007) and the subsequent interpretation of a causal role of the right DLPFC in controlling selfish impulses. One possible explanation for the observed differences to Ruff *et al.* (2013) may be that the spontaneous reaction that is controlled by a secondary process is in fact not always selfish but sometimes prosocial to some extent (Rand *et al.*, 2013; Schulz *et al.*, 2014), and that this is different between male (current study) and female participants (Ruff *et al.*, 2013). Future research specifically designed to explore these questions is needed to identify the exact role of the right DLPFC in voluntary giving (DG). Another possible explanation for the observed differences lies in the different techniques used to stimulate/suppress neural activity (TMS *vs* tDCS). In addition to the possible differences in neurophysiological effects induced by both techniques, they also differ in spatial resolution, thereby potentially affecting different (sub) regions within the LPFC (Priori *et al.*, 2009).

While our study demonstrated a crucial involvement of the right DLPFC in strategic fairness, the specific interplay of the right DLPFC and other brain areas is not resolvable with our study design. A recent study, combining fMRI with TMS, suggests that the DLPFC and the ventromedial prefrontal cortex (VMPFC) are part of a frontal cortex network, responsible for orchestrating normative choice (Baumgartner *et al.*, 2011). From this perspective, and in line with other research linking the VMPFC to the computation of abstract value signals guiding simple choice (see e.g. Chib *et al.*, 2009; Levy and Glimcher, 2011; McNamee *et al.*, 2013; Gross *et al.*, 2014), one possible interpretation is that the VMPFC is computing the expected value, integrating selfish goals with expected punishment for violating cooperation norms, whereas the DLPFC is linked to the execution of actions based on this valuation.

The complex norm system we observe in human societies, which is not present in other social animals, might be related to the fact that phylogenetically the DLPFC is one of the latest neocortical regions (Fuster, 2001). In a similar vein, the DLPFC is also ontogenetically one of the latest developing brain regions (Shaw *et al.*, 2008; Gogtay *et al.*, 2004), which in the context of our current findings could help to explain why young children up to the age of 3–8 years are not able to fully implement social rules like sharing resources with others (Fehr *et al.*, 2008). Taken together, our study provides strong evidence for a direct neurobiological basis of social norm compliance, a cornerstone for the functioning of human society.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, *37*, 291–304.

Andreoni, J., Croson, R. (2008). Partners versus strangers: random rematching in public goods experiments. In: Plott, C., Smith, V., editors. *Handbook of Experimental Economics Results*, Vol. 1, Amsterdam: North-Holland, pp. 776–83.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, *14*, 1468–76.

Bowles, S., Gintis, H. (2003). Origins of human cooperation. In: Hammerstein, P., editor. *Genetic and Cultural Evolution of Cooperation*. Cambridge: MIT Press, pp. 429–44.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, *11*(2), 122–33.

Brandts, J., Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, *14*, 375–98.

Camerer, C. (2003). Behavioral game theory: experiments in strategic interaction. *Princeton, New Jersey*. Princeton University Press.

Cherry, T.L., Frykblom, P., Shogren, J.F., Cherry, B.T.L. (2002). Hardnose the dictator. *The American Economic Review*, *92*, 1218–21.

Chib, V.S., Rangel, A., Shimojo, S., O'Doherty, J.P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*, 12315–20.

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's methods. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–5.

Egas, M., Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Biological Sciences*, *275*, 871–8.

Fehr, E., Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, *90*, 980–994.

Fehr, E., Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–40.

Fehr, E., Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*, 785–91.

Fehr, E., Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*, 185–90.

Fehr, E., Rockenbach, B. (2004). Human altruism: economic, neural, and evolutionary perspectives. *Current Opinions in Neurobiology*, *14*, 784–90.

Fehr, E., Bernhard, H., Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*, 1079–84.

Fuster, J.M. (2001). The prefrontal cortex - an update. *Neuron*, *30*, 319–33.

Gächter, S., Renner, E., Sefton, M. (2008). The long-run benefits of punishment. *Science*, *322*, 1510.

Gelman, A., Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, pp. 237–78.

Gintis, H. (2003). The Hitchhiker's Guide to altruism: gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology*, *220*, 407–18.

Gogtay, N.N., Giedd, J.N.J., Lusk, L.L., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences USA*, *101*, 8174–9.

Gross, J., Woelbert, E., Zimmermann, J., Okamoto-Barth, S., Riedl, A., Goebel, R. (2014). Value signals in the prefrontal cortex predict individual preferences across reward categories. *The Journal of Neuroscience*, *34*(22), 7580–6.

Hoffman, E., McCabe, K., Smith, V.L., Hoffman, B.E., Mccabe, K., Smith, V.L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, *86*, 653–60.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*, 829–32.

Knoch, D., Fehr, E. (2007). Resisting the power of temptations - The right prefrontal cortex and self-control. *Annals of the New York Academy of Sciences*, *1104*, 123–34.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of Sciences USA*, *106*, 20895–9.

Leider, S., Möbius, M.M., Rosenblat, T., Do, Q.A. (2009). Directed altruism and enforced reciprocity in social networks. *Quarterly Journal of Economics*, *124*, 1815–51.

Levy, D.J., Glimcher, P.W. (2011). Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *Journal of Neuroscience*, *31*, 14693–707.

List, J. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, *115*, 482–93.

Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. (2009). The BUGS project: evolution, critique and future directions. *Statistics in Medicine*, *28*, 3049–67.

McNamee, D., Rangel, A., O'Doherty, J.P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, *16*, 479–85.

Melis, A.P., Semmann, D. (2010). How is human cooperation different? *Philosophical Transactions of the Royal Society, Biological Sciences*, *365*, 2663–74.

Morey, R.D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–4.

Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, *14*, 137–58.

Oxoby, R.J.M., Spraggon, J. (2008). Mine and yours: property rights in dictator games. *Journal of Economic Behavior and Organization*, *65*, 703–13.

Priori, A., Hallett, M., Rothwell, J.C. (2009). Repetitive transcranial magnetic stimulation or transcranial direct current stimulation? *Brain Stimulation*, *2*, 241–5.

Rand, D.G., Greene, J.D., Nowak, M.A. (2013). Spontaneous giving and calculated greed. *Nature*, *489*, 427–30.

Ruff, C.C., Ugazio, G., Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, *342*, 482–4.

Sanfey, A., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*, 1755–8.

Sanfey, A.G., Stallen, M., Chang, L.J. (2014). Norms and expectations in social decision-making. *Trends in Cognitive Sciences*, *18*, 172–4.

Schulz, J.F., Fischbacher, U., Thöni, C., Utikal, V. (2014). Affect and fairness: dictator games under cognitive load. *Journal of Economic Psychology*, *41*, 77–87.

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperiments. *Beiträge zur experimentellen Wirtschaftsforschung*. Tübingen: Mohr, pp. 136–68.

Sethi, R., Somanathan, E. (1996). The evolution of social norms in common property resource use. *The American Economic Review*, *86*, 766–88.

Shaw, P., Kabani, N.J., Lerch, J.P., et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of Neuroscience*, *28*, 3586–94.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, *56*, 185–96.

Tomasello, M., Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind and Language*, *18*, 121–47.

Ule, A., Schram, A., Riedl, A., Cason, T. (2009). Indirect punishment and generosity toward strangers. *Science*, *326*, 1701–4.

Wout, M., Kahn, R., Sanfey, A., Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport*, *16*, 1849–52.