nature neuroscience

Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice

Thomas Baumgartner[1,2,4], Daria Knoch[2,4], Philine Hotz[1], Christoph Eisenegger[1,3] & Ernst Fehr[1]

Humans are noted for their capacity to over-ride self-interest in favor of normatively valued goals. We examined the neural circuitry that is causally involved in normative, fairness-related decisions by generating a temporarily diminished capacity for costly normative behavior, a 'deviant' case, through non-invasive brain stimulation (repetitive transcranial magnetic stimulation) and compared normal subjects' functional magnetic resonance imaging signals with those of the deviant subjects. When fairness and economic self-interest were in conflict, normal subjects (who make costly normative decisions at a much higher frequency) displayed significantly higher activity in, and connectivity between, the right dorsolateral prefrontal cortex (DLPFC) and the posterior ventromedial prefrontal cortex (pVMPFC). In contrast, when there was no conflict between fairness and economic self-interest, both types of subjects displayed identical neural patterns and behaved identically. These findings suggest that a parsimonious prefrontal network, the activation of right DLPFC and pVMPFC, and the connectivity between them, facilitates subjects' willingness to incur the cost of normative decisions.

The capacity to curb self-interest for the sake of normatively valued goals and to obey elementary social norms is a distinguishing characteristic of the human species[1]. This capacity is well documented and takes the form of costly behaviors such as voluntary participation in collective actions, helping strangers, reciprocating favors and punishing norm violators even though the actor receives no material benefits from these actions[2]. Despite their fundamental importance for modern civilization and in the maintenance of social order, the neurobiological mechanisms underlying normative behaviors are still poorly understood. We know that individuals with injuries to the prefrontal cortex[3,4] or fronto-temporal dementia[5,6] exhibit impaired norm compliance. Although interesting, these findings cannot provide conclusive evidence for the neural mechanism behind normative decision making because normal and pathological subjects typically differ in a considerable number of characteristics; even individuals with similar lesions often have different lesion etiologies. Studies on such subjects also have limited opportunities for experimental manipulations, and there is the possibility of functional reorganization after brain lesions, rendering the interpretation of behavioral differences in terms of neural characteristics difficult. To overcome these limits, recent studies used brain imaging methods[7–15] and brain stimulation methods[16–18] to investigate the neural mechanisms behind normative decisions.

Although brain imaging methods such as functional magnetic resonance imaging (fMRI) enable the identification of the neural network that correlates with normative choices, they fail to provide information about the causal role of the network, and the connectivity between different network components, for the observed choices unless they are combined with brain stimulation tools. In contrast, non-invasive brain stimulation tools such as transcranial magnetic stimulation allow researchers to causally affect subjects' choices, but they provide no information about the neural changes that trigger the behavioral change unless they are combined with brain imaging tools. Brain stimulation tools alone are not able to detect the neural network or the change in connectivity in the components of the network that the stimulation triggers, and hence cannot identify the network that causes the behavioral change. Because of the limits of these methods when applied in isolation, we exploited the synergies that arise when they are used in conjunction. By combining these two methods, we were able to overcome the above mentioned limits of pure repetitive transcranial magnetic stimulation (rTMS) and pure fMRI studies. In particular, we were able to identify the causal effect of rTMS on the task-related activity in the stimulated brain region, which has important consequences for the interpretation of the rTMS effect, and we identified the modulation of the activity in and the connectivity between other brain regions involved in the behavioral change.

To identify the neural network that is causally involved in normative decisions, we examined the responder's behavior in the ultimatum game[19,20]. In this game, a proposer is given a sum of money that he can allocate between himself and a responder. The proposer can make one offer to the responder, who can then accept or reject the offer. In case of acceptance, the proposed allocation is implemented; in case of a rejection, neither player receives anything. The key observation in this game is that responders accept fair offers, that is, offers that are close to the equal split, whereas, in Western cultures, low offers are typically viewed as a violation of a fairness norm that deserves sanctioning[19,20]. However, punishing the proposer for a low offer is costly for the responder, who also forfeits earnings. As a result of the existence of

[1]Department of Economics, Laboratory for Social and Neural Systems Research, University of Zurich, Zurich, Switzerland. [2]Department of Psychology, Laboratory for Social and Affective Neuroscience, University of Basel, Basel, Switzerland. [3]Behavioral and Clinical Neuroscience Institute, Department of Experimental Psychology, University of Cambridge, Cambridge, UK. [4]These authors contributed equally to this work. Correspondence should be addressed to T.B. (t.baumgartner@unibas.ch), D.K. (daria.knoch@unibas.ch) or E.F. (ernst.fehr@econ.uzh.ch).

Received 11 July; accepted 8 August; published online 2 October 2011; doi:10.1038/nn.2933

a fairness norm, the responder thus faces a conflict between a decision based on economic self-interest, to accept the low offer, and a normative decision, to punish the proposer for the unfair offer by rejecting it.

Previous neuroimaging evidence[8,21] indicates that DLPFC, ventrolateral prefrontal cortex (VLPFC), VMPFC, the anterior insula, and the dorsal anterior cingulate cortex (dACC) are activated during the responder's decision phase, and subjects with higher average activation in anterior insula are more likely to reject unfair offers[8]. In addition, disrupting the function of the right DLPFC by means of low-frequency rTMS[16,17] and cathodal transcranial direct current stimulation to the right DLPFC[18] has been shown to decrease rejection rates while leaving the ability to make fairness judgments intact. However, it is not known how the interaction between different brain regions gives rise to the normative behavior of rejecting an unfair offer although different brain areas are very unlikely to act in isolation during such a complex decision but are instead more likely to work together as a network. The identification of the associated network interactions is therefore important for understanding the neurobiological mechanisms of normative social behavior.

It has been hypothesized, for example, that anterior insula activation encodes the emotional resentment associated with unfair offers[8,21]. This hypothesis is based on evidence that implicates the anterior insula in disgust[22], pain[23] and more general negative emotion processing[11,24]. A bottom-up process in which anterior insula provides information to prefrontal executive control regions might elicit the rejection of unfair offers, and disrupting this bottom-up process might cause the change in behavior observed after rTMS of right DLPFC. It has also been hypothesized that dACC activation encodes the conflict between the emotional impulse to reject an unfair offer and the economic motive to accept the money. Perhaps other executive control regions such as the DLPFC modulate and resolve this conflict, enabling a final decision by communicating with dACC, a communication that might be disturbed by rTMS of right DLPFC. Another interesting hypothesis[8] is that DLPFC encodes the economic motive of accepting the money and exerts top-down control on the anterior insula, and the behavioral change induced by rTMS of right DLPFC might therefore be caused by a disruption of this top-down control. Finally, it is not known how the neural network that is active in fairness-related decisions in social interactions relates to that in non-normative decisions involving the purchase of private goods, such as food items[25–27], risky proceeds[28] or the delay of immediate gratifications for the sake of later larger rewards[29–31]. Recent research has identified the pVMPFC as an important region that encodes the valuation of these goods[25–31]. One recent study[27], for example, found that the same region of the pVMPFC encodes the valuation of food items, nonfood items such as DVDs and risky proceeds during the decision to buy these items. Thus, perhaps activation in pVMPFC encodes the decision value of the accept or reject decision in our setting. It would be intriguing if non-normative decisions regarding the purchase of food and risky proceeds and normative decisions involving the enforcement of a fairness norm were to rely on similar valuation circuitry (as proposed by the common neural currency hypothesis[32]) in the VMPFC, and if we were able to identify the causal involvement of this brain region in fairness-related decisions.

There is thus a bewildering variety of possibilities—in fact, the hypotheses mentioned above illustrate just a few of them—and the evidence that is currently available does not provide much information on the true neural interactions. For this purpose, we combined offline rTMS with fMRI to study the neural networks involved in an experimentally induced change in normative choice. The crucial feature of offline rTMS is that the brain is not stimulated during the task, but rather during a certain period, in our case 15 min with 1 Hz, immediately before the experimental task. This low-frequency rTMS stimulation protocol is known to generate an aftereffect[33] in which the function of the stimulated brain region is disrupted. Previous evidence indicates that low-frequency (1 Hz) offline rTMS to the right DLPFC sharply reduces the responders' rejection rate of unfair offers relative to a sham stimulation, whereas rTMS to the left DLPFC does not affect the rejection rate[16], that is, no substantial differences in rejection rate between rTMS to the left DLPFC and a sham stimulation are observed. Thus, by applying a low-frequency rTMS protocol to the right DLPFC, we generated a deviant case compared with the normal behavior in this task, whereas rTMS to the left DLPFC did not generate deviant behavior. We compared the deviant to the normal case to better understand the functioning of the normal case. In other words, we know that rTMS to the right DLPFC must change the functioning of decision-relevant neural networks, whereas rTMS to the left DLPFC leaves the decision-relevant neural networks intact. This means that we can identify the neural network that is causally involved in the behavioral change by comparing the neural activations (and their interactions) after rTMS to right DLPFC (referred to as right TMS) with those after rTMS to the left DLPFC (referred to as left TMS). It should be noted that it is problematic to identify the decision-relevant neural network by comparing right TMS to sham stimulation because it is known that TMS to the PFC can be slightly irritating and inconvenient for the subjects, which may cause behavioral and neural effects. It is therefore important to keep these effects constant across treatments, which requires an active TMS condition (such as left TMS) as a control treatment. Otherwise neural activations triggered by subjects' irritation that have nothing to do with decision-making may confound the results.

We applied the offline low-frequency rTMS protocol to 32 subjects (mean age = 21.6 years, s.d. = 2.2) while they lay on the scanner bed in front of the scanner. The subjects were transferred into the scanner immediately after the stimulation, where they played the ultimatum game in the role of the responder (see Online Methods and **Supplementary Fig. 1**). Each responder played 16 anonymous games with 16 different partners and a stake size of CHF 20 (CHF 1 ≈ USD 1.20). We limited the proposer's strategy by only permitting offers of CHF 10, 8, 6 or 4 to generate enough observations on the responders' side. CHF 10 is obviously the fairest offer, as it splits the stake size equally, whereas CHF 4 is the most unfair offer. We applied low-frequency rTMS for 15 min either to the right DLPFC ($n = 15$, referred to as the right TMS group) or to the left DLPFC ($n = 17$, referred to as the left TMS group).

We found that when fairness and economic self-interest were in conflict, normal subjects (left TMS group) rejected unfair offers much more frequently than did deviant subjects (right TMS group). Normal subjects also displayed significantly higher activity in, and connectivity between, the right DLPFC and the posterior VMPFC. In contrast, when there was no conflict between fairness and economic self-interest, both types of subjects displayed identical connectivity patterns and behaved identically. Normal and deviant subjects also showed no differential activation in any other brain region during the processing of unfair offers, suggesting that a parsimonious prefrontal network, the activation of right DLPFC and pVMPFC and the connectivity between them, facilitates subjects' willingness to incur the cost of normative, fairness-related, decisions.

RESULTS

Behavioral results

We carried out a behavioral analysis of the rejection rate using a repeated-measures ANOVA of treatment (left TMS, right TMS) × offer

(4, 6, 8, 10), which revealed a main effect of offer ($F_{3,28}$ = 25.86, P < 0.001, ETA^2 = 0.74), a main effect of treatment ($F_{1,30}$ = 6.11, P = 0.019, ETA^2 = 0.17) and an interaction effect of offer × treatment ($F_{3,28}$ = 3.19, P = 0.039, ETA^2 = 0.26; **Fig. 1**). The main effect of offer indicates that the rejection rate strongly decreased with offer size. Offers of 4 were rejected, on average, in 62% of the trials, whereas the rejection rate for the offers of 6 was 25%, and the offers of 8 and 10 were rejected in only 1% of the cases. Notably, the main effect of treatment and the interaction effect of offer × treatment were qualified by substantially lower rejection rates for the unfair offers of 4 and 6 in the right compared with the left TMS group (independent t test for offers of 4 and 6, t_{30} = −2.56, P = 0.016, two-tailed). Rejection rates after right TMS were only 45% for an unfair offer of 4 and 13% for an unfair offer of 6, whereas unfair offers of 4 and 6 were rejected in 79% and 35% of the cases, respectively, after left TMS. These large behavioral differences across treatments raise the question of which changes in the neural network are responsible for the changes in behavior.

Neural effects in the stimulated brain regions

We first examined the TMS effect on the stimulated brain regions: the left and the right DLPFC. Directly comparing the left TMS group with the right TMS group in the contrast unfair offers (4/6) > fair offers (8/10), we found higher activation in the right DLPFC (Brodmann area (BA) 46, x = 45, y = 24, z = 21) in the left than in the right TMS group. In fact, the right DLPFC was strongly recruited in the unfair > fair contrast for the normal subjects (left TMS group), who rejected a large number of unfair offers, whereas the right DLPFC failed to be recruited during the processing of unfair offers in the deviant subjects (right TMS group), who showed a diminished rejection behavior (**Fig. 2**). Thus, it seems that the recruitment of the right DLPFC in normal subjects is important for their willingness to reject unfair offers with the usual high frequency, as the willingness to reject strongly declines if rTMS disrupts the function of the right DLPFC.

Moreover, activity in left DLPFC does not help to explain rejections of unfair offers or to examine the differential rejection rates across the right and left TMS groups (**Fig. 2b**), as this region was generally not significantly activated during the processing of unfair offers (P > 0.20) and there was no significant difference in activation across right and left TMS groups, even if we strongly lowered the threshold to P < 0.20 (uncorrected). Thus, although right TMS significantly reduced the activity in the stimulated brain region (the right DLPFC) during the processing of unfair offers, left TMS left brain activation in the stimulated area (the left DLPFC) unaffected (**Fig. 2**).

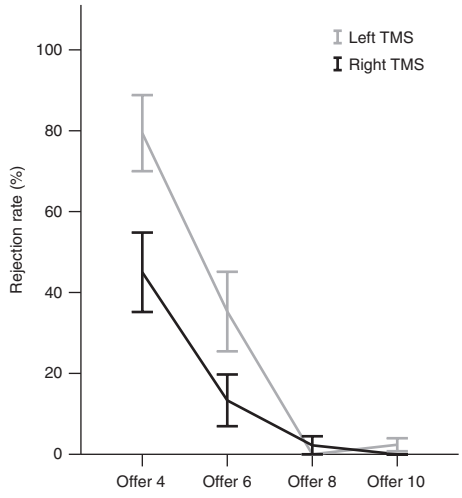

Figure 1 Rejection rates. Rejection rates (means ± s.e.m.) across treatment groups, broken down for unfair (offer 4 and 6) and fair offers (offer 8 and 10). Subjects whose right DLPFC was stimulated with rTMS exhibited a much lower average rejection rate (29.2%) for the two unfair offers of 4 and 6 than those subjects whose left DLPFC was stimulated (57.3%) (independent t test for offers of 4 and 6, t_{30} = 2.56, P = 0.016, two-tailed).

Neural network effects in remote brain regions

After our assessment of the effect of TMS on the stimulated brain regions, we examined the network effects on other brain regions. Previous neuroimaging studies[8,21] have implicated the anterior insula and the dACC in the processing of unfair offers. As these brain regions are known to be involved in the processing of negative emotions and the monitoring of conflicting cognitive or behavioral impulses[11,24,34–36], they draw our interest to whether and, if so, how left and right TMS are associated with differential neural activations in these regions. One possibility might be that the behavioral change involves the modulation of right anterior insula by the right DLPFC, as activity in the right anterior insula has been shown to correlate positively with the rejection rate for unfair offers. Alternatively, right TMS may also cause the sharp decrease in rejection rates by affecting subjects' ability to monitor conflicting motivational impulses as a result of reduced recruitment of dACC. However, our data support neither of these hypotheses, as we found similarly increased activations in both treatment groups in the bilateral inferior frontal gyrus/anterior insula regions (BA 47/13; **Supplementary Fig. 2a**) and in the dACC (BA 24/32; **Supplementary Fig. 2b** and **Supplementary Table 1**) in the contrast unfair offers > fair offers. Thus, left and right TMS caused differential activation in neither the bilateral anterior insula nor in the dACC. In addition, psycho-physiological interaction (PPI) analyses[37] (see Online Methods) revealed no differential variation in connectivity across treatment groups during unfair offers between the right DLPFC and the right anterior insula (P > 0.89), the left anterior insula (P > 0.76) or the dACC (P > 0.84). Thus, we found no evidence that the TMS-induced behavioral change is caused by a top-down modulation of anterior insula and dACC by the right DLPFC.

These findings suggest that if, as hypothesized[8,21], anterior insula and dACC encode the emotional resentment of unfair offers and the motivational conflict between self-interest and fairness, respectively, the subjects in the right and the left TMS groups are similarly engaged in these emotional and cognitive processes. The same argument holds for the observed activation in the dorsomedial prefrontal cortex (dMPFC, BA 10; **Supplementary Fig. 2b**), which has been implicated in mind-reading tasks in numerous studies[12,38,39]. Subjects in the right and left TMS group showed no differential activation in these areas, suggesting that this brain region is equally capable of processing the reading of intentions behind unfair offers in both treatment groups.

Which rTMS-induced changes in remote brain areas might then be responsible for the sharp reduction in the willingness to reject unfair offers? Beyond the stimulated right DLPFC, we found only one other area that displayed a differential activation across treatment groups during the processing of unfair offers: the pVMPFC. The left TMS group showed significantly higher activation in the pVMPFC (BA 10/11/32, P < 0.005, x = −3, y = 39, z = −9; **Fig. 3a** and **Supplementary Table 2**) compared with the right TMS group in the contrast unfair > fair offers. In fact, the left TMS group showed significantly positive activation of pVMPFC (P < 0.005), whereas the right TMS group failed to activate this region during the processing of unfair offers (**Fig. 3b**). In contrast, no brain region was more strongly activated in the right TMS group in the unfair > fair contrast (**Supplementary Table 2**).

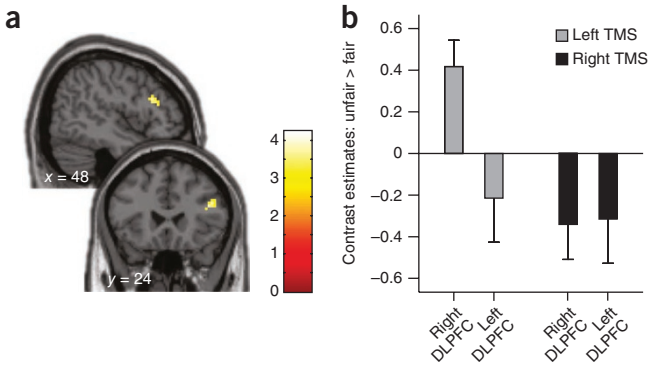

Figure 2 Differential group activation in the right DLPFC: left TMS $^{(\text{unfair} > \text{fair})}$ > right TMS $^{(\text{unfair} > \text{fair})}$. (**a**) Disrupting the right DLPFC with rTMS led to a differential group activation pattern in the disrupted right DLPFC ($x = 45$, $y = 24$, $z = 21$, $P < 0.005$, uncorrected[42], cluster extent threshold = 15 voxels; activity in the DLPFC survived small volume family-wise error (FWE) corrections at $P < 0.05$ in a 20-mm sphere defined by the peak reported in ref. 8; see Online Methods), which was qualified by increased activation in the left TMS group compared with the right TMS group during unfair offers. No such differential group activation was observed in the left DLPFC even at a strongly lowered threshold of $P < 0.20$ (uncorrected). (**b**) Bar plots represent differences (mean ± s.e.m.) in contrast estimates (unfair offers > fair offers) of homologs spherical regions of interest (ROIs) (5 mm) in the bilateral DLPFC around the peak coordinate of activation depicted in **a** (for the left side, the x coordinate was mirrored), broken down for the treatment groups (left TMS/right TMS). Bar plots indicate that a differential activation across right and left TMS group was only observed in the right DLPFC.

Because the change in neural activity induced by right TMS originates in the right DLPFC, it seems likely that communication between the right DLPFC and the pVMPFC facilitates the neural changes in the pVMPFC. Thus, right TMS might not just reduce activity in the right DLPFC and the pVMPFC, but it might also change the connectivity between the two regions during the processing of unfair offers. To examine this question, we again applied PPI analyses[37]. We found that right TMS strongly affected the connectivity in a context-dependent manner (**Fig. 4**). In particular, we observed in both treatment groups that these two brain regions displayed no change at all in connectivity (relative to baseline connectivity) during the processing of fair offers (see fair connectivity in **Fig. 4b**), which suggests that there is no special need for communication between the DLPFC and the pVMPFC after a fair offer. However, after an unfair offer, when fairness and self-interest are in conflict with each other, the left TMS group exhibited a strong increase in connectivity between these two brain regions relative to the baseline (see unfair connectivity in **Fig. 4b**), implying that, if right DLPFC activity is high (which it is after an unfair offer), pVMPFC activity is also high. In contrast, there was no significant change in connectivity between the right DLPFC and the pVMPFC in the deviant subjects ($P > 0.20$), suggesting that the usual communication of these two regions during the processing of unfair offers is disrupted. Thus, the subjects who received left TMS displayed a much higher connectivity level after unfair offers (**Fig. 4b**), whereas subjects who received right TMS displayed the same baseline connectivity level after fair and unfair offers (**Fig. 4b**). In fact, many of the voxels in pVMPFC that exhibited lower activation in the deviant subjects also lacked an increase in connectivity with the right DLPFC (**Fig. 4a**). Thus, the TMS of the right DLPFC not only reduced neural activation in this area but also removed the usual increase in connectivity between the right DLPFC and the pVMPFC after unfair offers, which was then also associated with a lower activity in the pVMPFC. These findings suggest that the increased connectivity between right DLPFC and pVMPFC during the processing of unfair offers is important for subjects' rejection behavior.

This view would receive further support if individual connectivity differences between these two regions were correlated with individual differences in rejection rates. To examine this conjecture, we computed the change in connectivity after unfair offers (relative to fair offers) for each subject in the overlap area (**Fig. 4a**). We found that subjects who displayed a larger increase in connectivity after unfair offers rejected a higher proportion of these offers (correlation = 0.421, $P = 0.016$). The importance of this connectivity finding for explaining individual differences in normative decision making was further strengthened by the following two observations. First, neither the activity in the DLPFC alone nor the activity pattern in the VMPFC alone correlated with the rejection of unfair offers (all $P > 0.10$). Thus, the strength of connectivity between the right DLPFC and VMPFC is a better predictor of rejection rates than the activity pattern in each region alone. Second, disrupting the function of the right DLPFC completely eliminated the predictive power of the DLPFC-VMPFC connectivity (correlation = −0.221, $P = 0.43$). Thus, only in those subjects with a normally functioning right DLPFC was the DLPFC-VMPFC connectivity able to predict individual differences in rejection behavior (correlation = 0.409, $P = 0.05$, one-tailed).

These activation and connectivity patterns suggest that the communication between right DLPFC and pVMPFC in normal subjects may be crucial for the upregulation of pVMPFC during the processing of unfair offers and that this upregulation is distorted in the subjects who receive right TMS. Finally, it is also noteworthy that the differential activation of pVMPFC (**Fig. 3**) cannot be the result of differential input from the anterior insula or the dACC, as these regions were not activated differentially across treatment groups (**Supplementary Fig. 2** and **Supplementary Table 1**). In addition, when we computed a PPI with either anterior insula or dACC as the seed region, we found no differential variation in connectivity across treatment groups during

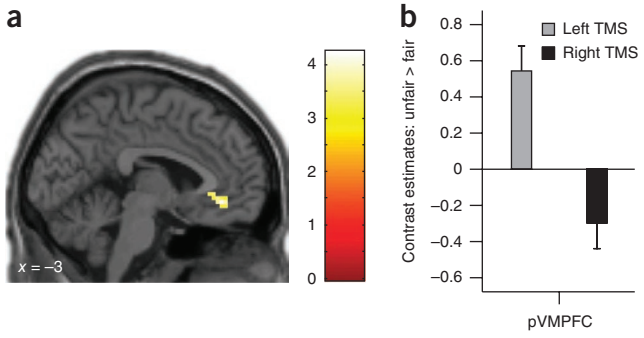

Figure 3 Differential group activation in pVMPFC: left TMS $^{(\text{unfair} > \text{fair})}$ > right TMS $^{(\text{unfair} > \text{fair})}$. (**a**) Disrupting the right DLPFC with rTMS changed neural activity not only in the disrupted brain area (depicted in **Fig. 2**), but also in another task-related remote brain region located in the pVMPFC ($x = -3$, $y = 39$, $z = -9$, thresholded at $P < 0.005$, uncorrected[42], cluster extent threshold = 15 voxels; activity in the pVMPFC survived small volume FWE corrections at $P < 0.05$ in a 20-mm sphere defined based on peaks in refs. 25–27,29,30; see Online Methods). (**b**) Bar plots represent differences (mean ± s.e.m.) in contrast estimates (unfair offers > fair offers) of a functional ROI based on the depicted activation, broken down for the two treatment groups (left TMS/right TMS). Bar plots indicate that only the left TMS group reacted to unfairness with increased activity (at $P < 0.005$) in the pVMPFC.

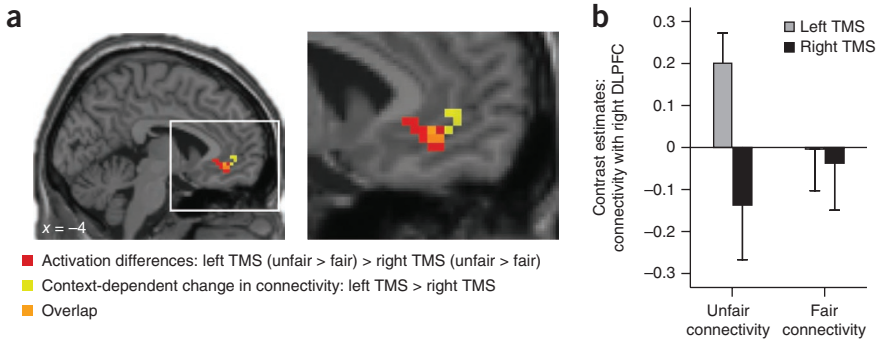

Figure 4 Treatment group differences in connectivity between right DLPFC and pVMPFC. (**a**) Overlay of the pVMPFC cluster that showed a larger change in connectivity after unfair offers (compared with fair offers) with the right DLPFC in the left compared with the right TMS group (yellow, at $P < 0.005$, cluster extent = 18 voxels[42]) and the pVMPFC cluster that showed differential activation in the contrast unfair > fair offers in the left compared with the right TMS group (red). Overlapping voxels are displayed in orange. (**b**) Bar plots based on the functional ROI (red) from **a** indicate that the differential context-dependent change in connectivity between the left and right TMS group was qualified by a differential change in connectivity during unfair offers (unfair connectivity), but not during fair offers (fair connectivity). The left TMS group therefore only showed an increased connectivity between the right DLPFC and pVMPFC at $P < 0.01$ during unfair offers, whereas the connectivity between these two brain regions did not change (relative to baseline connectivity) after fair offers. Moreover, after right TMS, the connectivity between right DLPFC and pVMPFC never deviated from the baseline (indicated by the two black bars); that is, these brain regions no longer communicated more after unfair offers. Bar plots depict mean ± s.e.m.

unfair offers between these brain regions and the pVMPFC (anterior insula, $P = 0.81$; dACC, $P = 0.52$). Thus, the anterior insula and the dACC displayed neither differential activations nor differential connectivity patterns during the processing of unfair offers across treatment groups, suggesting that they are not involved in the behavioral change induced by right TMS.

Validation of the reported behavioral and neuronal patterns

Finally, to further check the robustness and validity of the reported behavioral and neuronal findings, we conducted an additional control treatment in which subjects did not receive any rTMS before playing the ultimatum game in the scanner (no TMS condition with $n = 18$ healthy male subjects). The no TMS group enabled us to confirm that the behavior of the left TMS group was very similar and statistically indistinguishable from the behavior of the no TMS group ($P > 0.20$; **Supplementary Fig. 3**). Moreover, similar to the left TMS group, the no TMS group rejected unfair offers significantly more often than the right TMS group ($P = 0.024$). In addition, the comparison between the no TMS and the right TMS group in terms of neural activations and connectivity patterns yielded basically the same results as the comparison of the left TMS with the right TMS group (**Supplementary Analysis 1** and **Supplementary Figs. 4** and **5**). These findings confirm that the left TMS group can legitimately be treated as a normal control group in terms of both behaviors and behaviorally relevant neural activations and connectivity patterns.

DISCUSSION

We combined offline rTMS with the examination of the changes in the neural networks associated with the TMS-induced behavioral changes. By comparing the deviant cases (after right TMS) with subjects' normal behavior and normally functioning neural network after left rTMS, we were able to overcome the limits of pure fMRI and pure TMS studies and achieved a deeper understanding of the neural interactions that are causally involved in fairness-related, normative decisions.

We observed that right TMS (but not left TMS) prevented the recruitment of the right DLPFC during the processing of unfair offers. Previous pure TMS studies[16,17] did not observe this because they lacked measures of neural activation. This finding suggests that the neural activation in the right DLPFC during the processing of unfair offers is decisive in the ability to make costly normative decisions. In contrast, we find no differential neural activation across treatment groups in the left DLPFC which may explain why left TMS causes no behavioral effects.

Another important finding is that rTMS of the right DLPFC had no discernible effect on activity in areas such as the anterior insula (which has been implicated in the emotional processing of the unfairness of bargaining offers)[8,21], the dACC (which has been implicated in the monitoring of motivational and cognitive conflicts)[11,35] and the dMPFC (which has been implicated in theory of mind tasks)[38,39]. Furthermore, we found no evidence that right DLPFC is involved in a top-down regulation of these areas. Thus, activity and connectivity in these brain regions cannot explain the sharp decline in rejection rates after right TMS. Notably, emotion processing and mind reading are crucial for judging the fairness of a given offer. It is, for example, well known that many subjects attribute greedy intentions to low offers and that this attribution is a crucial input in the fairness evaluation[40], suggesting that fairness judgments are hampered if mind-reading abilities are reduced. Similarly, if subjects cannot feel the unfairness of a low offer, it is difficult to see how they can judge it to be so. The absence of a TMS effect on these areas may explain why brain stimulation of right DLPFC leaves subjects' ability to judge the (un)fairness of low offers intact[16,18]. Finally, the absence of a differential activation in anterior insula, dACC and dMPFC across treatment groups does not imply that these brain regions are generally not involved in normative choice. In fact, if these regions are crucial for the ability to attribute fairness intentions and for making fairness judgments, they are likely to be involved in the extent to which behavior and judgment are aligned. However, the absence of a differential activation across treatment groups in these regions indicates that they are not causally involved in the behavioral change across groups.

The only other region that was differentially activated during processing of unfair offers in the right compared with the left TMS group was the pVMPFC. The deviant subjects showed a lower activation than the normal subjects, suggesting that deactivation in this region reduces the ability to make costly normative decisions after right TMS. Moreover, we observe telling connectivity patterns between the right DLPFC and the very same voxels in the pVMPFC that show reduced activation after right TMS. When normal subjects faced unfair offers leading to a conflict between self-interest and fairness, they displayed a significant positive change in connectivity between the right DLPFC and the pVMPFC, which was not the case if they faced fair offers. Right TMS, however, removed this increased connectivity, suggesting that DLPFC and pVMPFC were then unable to interact with each other in the manner necessary to facilitate the normative decision.

This conjecture is particularly intriguing in light of recent evidence regarding the functional role of pVMPFC in decision making, which suggests that pVMPFC encodes the decision value of consumption goods[25–27] and normatively valued goods[7,13,14]. In fact, the peak of the rTMS-induced reduction in VMPFC activity (at MNI coordinates $x = -3$, $y = 39$, $z = -9$; **Fig. 3**) occurs in the same area of pVMPFC that has been shown to provide a common neural representation of the decision value of different consumption goods (located in voxels around MNI coordinates $x = -3$, $y = 42$, $z = -6$, ref. 27). It may be the case that pVMPFC correlates with decision values and that pVMPFC activation is causally involved in the implementation of subjects' goals. In this view, the pVMPFC may encode the decision value of a rejection (please see **Supplementary Discussion** for a precise definition), which may be upregulated in normal subjects through communication with the right DLPFC in contexts in which fairness is in conflict with self-interest. This interpretation is consistent with the fact that normal subjects showed an increase in the activation of pVMPFC when they faced unfair offers and that they rejected these offers much more frequently, whereas the deviant subjects lacked this increase in pVMPFC activation and displayed a considerably lower rejection rate.

The above interpretation of pVMPFC activation may seem puzzling in light of the fact that parts of the VMPFC have often been implicated in the experience of positive stimuli, such as attractive faces or monetary rewards. In contrast, in our setting, pVMPFC was active during the processing of unfair offers, which can hardly be described as positive stimuli. It is therefore not possible to interpret the observed pVMPFC activation as a signal of the experienced value of a stimulus; it is better interpreted as a signal of decision value, that is, the value of a rejection. In fact, recent evidence[41] suggests that value signals that encode positive experiences and value signals that predict choice may be encoded in distinct areas of the VMPFC: the passive experience of positive stimuli (attractive faces and receipt of money) activates the anterior VMPFC (aVMPFC), but these activations do not predict subsequent choices when subjects face a trade-off between money and the viewing of attractive faces. Activation in posterior VMPFC during passive consumption does, however, predict subsequent choices. In view of this evidence, it would be reassuring for our interpretation if we could also find the distinct encoding of positive stimuli in the aVMPFC. In our setting, the receipt of a fair offer is an unambiguously positive stimulus because a fair offer provides a monetary benefit and a fairness benefit without any cost. For this purpose, we computed the contrast between fair and unfair offers and found activation in only two regions: the ventral striatum and a distinct area in the aVMPFC (**Supplementary Table 1** and **Supplementary Figs. 6** and **7**). Notably, the left and the right TMS group showed no differential activations in these areas, suggesting that TMS did not change the encoding of the experience of positive stimuli.

In summation, our results suggest that the context-dependent communication between the right DLPFC and the pVMPFC is important for a neural model of normative decision making: no special communication between these regions seems to be needed in the absence of conflicting motives. However, an increased need for communication between these brain regions seems to arise in case of strongly conflicting motives, thus facilitating the choice of the costly normative option. The fact that subjects who display a larger context-dependent change in connectivity between right DLPFC and pVMPFC reject unfair offers more frequently also suggests the possibility of a deeper understanding of individual differences in normative behavior in terms of connectivity differences between these brain regions. Thus, our findings may also help to explain the implications of brain damage in these prefrontal brain regions. Finally, these results also may have implications for the therapeutic use of non-invasive TMS in (forensic) psychiatric patients displaying persistent antisocial and aggressive behaviors resulting from a hypoactivation in the VMPFC. Given that the VMPFC is not directly accessible through non-invasive brain stimulation, the fact that we could affect activation in the pVMPFC by stimulating the right DLPFC suggests that it might be possible to increase activation in the VMPFC with high-frequency rTMS or anodal transcranial direct-current stimulation of the DLPFC.

Taken together, combining non-invasive offline brain stimulation and neuroimaging provides a powerful approach for studying the neurobiological mechanisms of decision making. Imaging this form of virtual neuropsychology provides potentially important insights into the neural substrates of decision-making in the healthy brain, which may be useful for a deeper understanding and in the therapy of normative behavioral pathologies in psychiatric patients.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We thank C. Ruff, K.E. Stephan and A. Rangel for their helpful comments. This study is a part of the project on the foundations of norm compliance in the National Center of Competence in Affective Sciences. E.F. also acknowledges support from the Neurochoice Project of SystemsX, the Swiss Initiative for Systems Biology. D.K. acknowledges support from the Swiss National Science Foundation (grant no. PP00P1-123381).

AUTHOR CONTRIBUTIONS

T.B., D.K. and E.F. designed the study. T.B., D.K. and C.E. performed all of the experiments. T.B. and P.H. analyzed the data. T.B., D.K. and E.F. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Boyd, R.T. & Richerson, P. *The Origin and Evolution of Cultures* (Oxford University Press, Oxford, 2005).
2. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
3. Damasio, A.R. *Descartes' Error: Emotion, Reason and the Human Brain* (Hayrer Collins, New York, 1995).
4. Shallice, T. & Burgess, P.W. Deficits in strategy application following frontal lobe damage in man. *Brain* **114**, 727–741 (1991).
5. Neary, D. *et al.* Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* **51**, 1546–1554 (1998).
6. Miller, B.L., Darby, A., Benson, D.F., Cummings, J.L. & Miller, M.H. Aggressive, socially disruptive and antisocial behaviour associated with fronto-temporal dementia. *Br. J. Psychiatry* **170**, 150–154 (1997).
7. Moll, J. *et al.* Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc. Natl. Acad. Sci. USA* **103**, 15623–15628 (2006).
8. Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. & Cohen, J.D. The neural basis of economic decision-making in the Ultimatum Game. *Science* **300**, 1755–1758 (2003).
9. Harbaugh, W.T., Mayr, U. & Burghart, D.R. Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* **316**, 1622–1625 (2007).
10. de Quervain, D.J. *et al.* The neural basis of altruistic punishment. *Science* **305**, 1254–1258 (2004).
11. Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K. & Fehr, E. The neural circuitry of a broken promise. *Neuron* **64**, 756–770 (2009).
12. Baumgartner, T., Gotte, L., Gugler, R. & Fehr, E. The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Hum. Brain Mapp.* published online, doi:10.1002/hbm.21298 (13 May 2011).

13. Hare, T.A., Camerer, C.F., Knoepfle, D.T. & Rangel, A. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *J. Neurosci.* **30**, 583–590 (2010).
14. Tricomi, E., Rangel, A., Camerer, C.F. & O'Doherty, J.P. Neural evidence for inequality-averse social preferences. *Nature* **463**, 1089–1091 (2010).
15. Knoch, D., Gianotti, L.R., Baumgartner, T. & Fehr, E. A neural marker of costly punishment behavior. *Psychol. Sci.* **21**, 337–342 (2010).
16. Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. & Fehr, E. Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* **314**, 829–832 (2006).
17. van 't Wout, M., Kahn, R.S., Sanfey, A.G. & Aleman, A. Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport* **16**, 1849–1852 (2005).
18. Knoch, D. *et al.* Studying the neurobiology of social interaction with transcranial direct current stimulation–the example of punishing unfairness. *Cereb. Cortex* **18**, 1987–1990 (2008).
19. Güth, W., Schmittberger, R. & Schwarze, B. An experimental analyses of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367–388 (1982).
20. Henrich, J. *et al.* In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* **91**, 73–78 (2001).
21. Tabibnia, G., Satpute, A.B. & Lieberman, M.D. The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol. Sci.* **19**, 339–347 (2008).
22. Phillips, M.L. *et al.* A specific neural substrate for perceiving facial expressions of disgust. *Nature* **389**, 495–498 (1997).
23. Singer, T. *et al.* Empathy for pain involves the affective but not sensory components of pain. *Science* **303**, 1157–1162 (2004).
24. Herwig, U. *et al.* Modulation of anticipatory emotion and perception processing by cognitive control. *Neuroimage* **37**, 652–662 (2007).
25. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27**, 9984–9988 (2007).
26. Hare, T.A., O'Doherty, J., Camerer, C.F., Schultz, W. & Rangel, A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* **28**, 5623–5630 (2008).
27. Chib, V.S., Rangel, A., Shimojo, S. & O'Doherty, J.P. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* **29**, 12315–12320 (2009).
28. Boorman, E.D., Behrens, T.E., Woolrich, M.W. & Rushworth, M.F. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* **62**, 733–743 (2009).
29. Hare, T.A., Camerer, C.F. & Rangel, A. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* **324**, 646–648 (2009).
30. Kable, J.W. & Glimcher, P.W. The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10**, 1625–1633 (2007).
31. Kable, J.W. & Glimcher, P.W. The neurobiology of decision: consensus and controversy. *Neuron* **63**, 733–745 (2009).
32. Montague, P.R. & Berns, G.S. Neural economics and the biological substrates of valuation. *Neuron* **36**, 265–284 (2002).
33. Robertson, E.M., Theoret, H. & Pascual-Leone, A. Studies in cognition: the problems solved and created by transcranial magnetic stimulation. *J. Cogn. Neurosci.* **15**, 948–960 (2003).
34. Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U. & Fehr, E. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron* **58**, 639–650 (2008).
35. Carter, C.S. *et al.* Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747–749 (1998).
36. Ridderinkhof, K.R., Ullsperger, M., Crone, E.A. & Nieuwenhuis, S. The role of the medial frontal cortex in cognitive control. *Science* **306**, 443–447 (2004).
37. Friston, K.J. *et al.* Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage* **6**, 218–229 (1997).
38. Frith, U. & Frith, C.D. Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 459–473 (2003).
39. Van Overwalle, F. Social cognition and the brain: a meta-analysis. *Hum. Brain Mapp.* **30**, 829–858 (2009).
40. Blount, S. When social outcomes aren't fair: the effect of causal attributions on preferences. *Organ. Behav. Hum. Decis. Process.* **63**, 131–144 (1995).
41. Smith, D.V. *et al.* Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *J. Neurosci.* **30**, 2490–2495 (2010).
42. Lieberman, M.D. & Cunningham, W.A. Type I and type II error concerns in fMRI research: re-balancing the scale. *Soc. Cogn. Affect. Neurosci.* **4**, 423–428 (2009).

ONLINE METHODS

Participants. We studied 32 right-handed men (mean age ± s.d. = 21.6 ± 2.2 years, range 19–27 years). Subjects gave informed written consent before participating in the study, which was approved by the local ethics committee (Zurich, Switzerland). Each subject participated in only one of the two treatment conditions (rTMS of left DLPFC, rTMS of right DLPFC), and none experienced TMS or participated in the ultimatum game previously. No subject had a history of psychiatric illness or neurological disorders. There was no difference between groups with respect to age (independent *t* test, $t_{30} = 0.059$, $P = 0.954$). Subjects neither experienced serious adverse side effects nor reported scalp pain, neck pain or headaches after the experiment.

The ultimatum game. Every responder received four offers of 4, four offers of 6, three offers of 8, and five offers of 10. The sequence in which subjects received these different offers was randomized across subjects. The distribution of these offers was derived from our behavioral pilot experiments, in which the proposers generated, on average, this distribution. We asked these proposers after the pilot experiments if we could use their offers again in subsequent experiments. If they agreed and we actually used their decisions in the scanner experiment, they were paid based on player B's decision in the scanner. Thus, the responders in the scanner faced the decisions of 16 real human interaction partners and their choices actually affected the interaction partners' monetary payoffs.

Each subject received CHF 60 (CHF 1 ≈ USD 1.20) as a show-up fee in addition to the money earned in the ultimatum games. Subjects knew that 8 of the 16 bargaining trials would be randomly selected at the end of the experiment for payment. The payment was made according to the outcome of the trial; for example, both players earned nothing in a selected trial if the responder had rejected the offer in this trial. Subjects received instructions that explained the rules of the game before stimulation. Each subject was required to complete a series of test questions successfully after reading the instructions to verify comprehension.

rTMS. rTMS was administered to the dorsolateral prefrontal cortex (DLPFC) for 15 min before subjects participated in the ultimatum game (off-line procedure) using a Magstim Rapid Magnetic Stimulator (Magstim) and a commercially available nonferromagnetic figure-eight coil (70-mm diameter double-circle) with an magnetic resonance–compatible 9-m cable. Subjects lay down in the scanner room on the scanner bed, which was located in the starting position outside of the scanner. The coil was fixed and the subject's head was firmly held in place by means of a magnetic resonance–compatible coil and head holder. The TMS coil was placed over F4 and F3 using the electroencephalogram 10–20 coordination system for stimulation of the right and left DLPFC, as in previous studies[17,43–45]. The stimulation intensity was set at 54% of maximum stimulator output. The coil was held tangential to the subject's head with the handle pointing rostrally. Subjects received a single 15-min, 1-Hz rTMS train (900 pulses) over either the left DLPFC or right DLPFC. The rTMS parameters were well within currently recommended guidelines[46] and resulted in a suppression of excitability of the targeted cortical region for several minutes following completion of the rTMS train[33]. The magnetic resonance–compatible coil and head holder were removed immediately after the rTMS stimulation and the scanner bed was automatically placed into the scan position inside of the scanner. All pre-scan measurements (including, for example, the localization scan and the slice alignment procedure) were conducted before the rTMS stimulation. This procedure allowed us to start the scanner measurements and the ultimatum games 40 s after the cessation of rTMS train. In total, the 16 ultimatum games lasted approximately 6 min. The responders' decisions in the 16 games were thus well within the borders of the rTMS aftereffect[47].

fMRI image acquisition. The experiment was conducted on a 3 T Philips Intera whole-body magnetic resonance scanner (Philips Medical Systems) equipped with an 8-channel Philips sensitivity-encoded (SENSE) head coil. Structural image acquisition consisted of 180 T1-weighted transversal images (0.75-mm slice thickness). For functional imaging, a total of 150 volumes were obtained using a SENSE[48] T2*-weighted echo-planar imaging sequence with an acceleration factor of 2.0. We acquired 35 axial slices covering the whole brain with a slice thickness of 3 mm (inter-slice gap of 0.4 mm, non-interleaved acquisition, repetition time = 2,500 ms, echo time = 35 ms, flip angle = 77°, field of view = 22 mm, matrix size = 128 × 128). To optimize functional sensitivity in orbitofrontal cortex and medial temporal lobes, we used a tilted acquisition in an oblique orientation at 30° to the AC-PC line.

fMRI preprocessing. The statistical parametric mapping software package (SPM5, Wellcome Department of Cognitive Neurology) implemented in Matlab (Version R2006a) was used for the preprocessing and statistical analyses. All images were realigned to the first volume for analysis, corrected for motion artifacts and time of acquisition in a TR, normalized ($3 \times 3 \times 3$ mm^3) into standard stereotaxic space (template provided by the Montreal Neurological Institute) and smoothed using an 8-mm full-width-at-half-maximum Gaussian kernel. A band-pass filter, which was composed of a discrete cosine-basis function with a cut-off period of 128 s for the high-pass filter, was applied. To increase signal to noise ratio, we minimized global intensity changes by scaling each image to the grand mean.

General linear model. We performed random-effects analyses on the functional data for the decision phase. For that purpose, we defined a general linear model that included four regressors of interest and three regressors of non-interest. The four regressors of interest were modeled for the decision phase consisting of offer 4, offer 6, offer 8 and offer 10. Onsets for these regressors were at the time of decision screen appearance. These screens were displayed for a duration of 6 s and were modeled accordingly. In addition to the four regressors of interest, three regressors of non-interest were modeled. One of these regressors preceded the decision screen regressors and had a length of 6 s, during which the subjects in the scanner were informed that the Players A were considering the offer. Another regressor followed the decision screen regressors and informed the subjects in the scanner that a new Player A was being assigned. This regressor was modeled with a duration of 3 s. Finally, the last regressor of non-interest modeled the feedback period presented at the end of the total 16 decision rounds and had a length of 20 s (see **Supplementary Fig. 1** for a timeline of screens in a single trial of the fMRI experiment and see **Supplementary Analysis 2** for statistical analyses of the response times). All regressors were convolved with a canonical hemodynamic response function. The six scan-to-scan motion parameters produced during realignment were included as additional regressors in the SPM analysis to account for residual effects of scan-to-scan motion.

Linear contrasts of regression coefficients were computed at the individual subject level and then taken to a group-level random effects analysis of variance. The following two different contrast images were calculated for the different analyses of the decision phase at the individual level using the four regressors of interest: unfair offers (regressor offer 4 and 6) > fair offers (regressor offer 8 and 10) and fair offers (regressor offer 8 and 10) > unfair offers (regressor offer 4 and 6).

For second-level random effects analysis, the single-subject contrasts were entered into one-sample and two-sample *t* tests. We calculated two simple contrasts and two serial subtraction terms based on these *t* tests: unfair > fair offers in all subjects (irrespective of treatment groups), fair > unfair offers in all subjects (irrespective of treatment groups), left TMS $^{(\text{unfair} > \text{fair offers})}$ > right TMS $^{(\text{unfair} > \text{fair offers})}$ and right TMS $^{(\text{unfair} > \text{fair offers})}$ > left TMS $^{(\text{unfair} > \text{fair offers})}$. The two described serial subtraction terms were exclusively masked at $P < 0.005$ with the reversed second contrast of the serial subtraction term to ensure that the observed differences between the two groups were not a result of differences in the reversed second contrast.

Statistical analysis: correction for multiple comparisons. The correction for multiple comparisons was carried out using a two-step approach. First, we applied an uncorrected *P* value of 0.005 combined with a cluster-size threshold of 15 voxels to our a priori regions of interest[42], which encompassed emotion- and conflict-related areas of the anterior insula and the dorsal ACC, mentalizing-related areas of the DMPFC, control-related areas of the DLPFC and valuation-related areas of the VMPFC. We reported other brain regions that were significant at the same threshold (see **Supplementary Tables 1** and **2**). However, we are reluctant to make any interpretations based on these results because we made no a priori hypotheses. Second, we checked whether our a priori regions of interest survived small-volume FWE corrections at $P < 0.05$. The small volumes were either defined anatomically or functionally using the WFU PickAtlas toolbox[49]. The anatomically defined small volumes were based on the automated anatomical labeling atlas[50] included in the toolbox and consisted of the anterior insula and the ACC. The functionally defined small volumes were based on a 20-mm sphere

around the peak of activity reported previously. For the mentalizing-related area of the DMPFC, we used a recently published meta-analysis[39] on social cognition to define the center of the sphere in the DMPFC ($x = -3$, $y = 48$, $z = 30$). We used the average coordinates of the peaks of activations in the reviewed mentalizing tasks in ref. 39, which includes tasks involving intention/trait inferences and moral judgment tasks in economic game situations. For the control-related area in the DLPFC, we used the peak coordinate ($x = 39$, $y = 37$, $z = 26$) of a previous study[8] whose disruption by rTMS[16] has been shown to reduce subject's ability to make normative decisions. For the valuation-related area in the VMPFC, we averaged the peak coordinates ($x = 2$, $y = 41$, $z = -6$) of recent neuroimaging studies on non-normative choice[25–27,29,30], which consistently demonstrated that the activity in the posterior part of the VMPFC encodes the decision value of consumption goods, or in economic terms, the willingness to pay for consumption goods. Note that, before averaging, all coordinates were transferred into MNI-space using the WFU PickAtlas toolbox[49]. Finally, to ensure that the spherical ROIs were only composed of gray matter, these ROIs were intersected with the respective Brodmann areas (DMPFC, BA 9/10; DLPFC, BA 9/46; VMPFC, 10, 11, 24, 32). We found that all of our a priori regions of interest survived this multiple comparisons correction procedure at $P < 0.05$.

ROI analyses. We created either functional or spherical ROIs (5 mm in diameter) around the peak of activation using the MarsBaR software (**Figs. 2–4**). Functional ROIs were created by selecting all voxels that were significantly activated at $P < 0.005$ together with a cluster extent threshold of 15 voxels in the corresponding analyses.

PPI analysis. To determine whether the treatment groups (left/right TMS) differed in connectivity patterns as well as in activity, we conducted PPI analyses[37]. These analyses, usually framed in terms of effective connectivity, seek to detect context-dependent changes in connectivity (for example, enhanced connectivity during unfair compared to fair offers) between a seed region (for example, right DLPFC) and other brain regions. For this purpose, we extracted the individual mean-corrected time series of three seed regions (right DLPFC, dACC and right anterior insula) from a 5-mm spherical ROI around the peak of activation based on the corresponding contrast analyses (see **Supplementary Table 1** and **2**). Using these three time series, we conducted three independent PPI analyses as follows. The activity in remote brain regions was regressed on a voxel-wise basis against the product of these time series (either right DLPFC, dACC or right anterior insula) and the vector of the psychological variable of interest (unfair offers minus fair offers), while the physiological and the psychological variables alone served as regressors of no interest. The results of these first-level analyses were then taken to random-effects group analyses using two-sample *t* tests. The goal of these analyses was to examine whether we found brain regions showing a differential context-dependent change in connectivity between the left TMS and right TMS group. In other words, we searched for brain areas in which the left TMS and the right TMS group showed a different change in the connectivity between the seed region and other (remote) regions during the processing of unfair compared to fair offers. The rationale for this procedure is that a context-dependent and treatment group–specific connectivity pattern could provide an explanation for why the left TMS group rejects unfair offers at a much higher rate compared with the right TMS group. We focused in our analyses of connectivity patterns primarily on regions that either demonstrated an enhanced activity in the left compared with the right TMS group (right DLPFC, pVMPFC; **Supplementary Table 2**) or that were differentially activated during unfair compared with fair offers (dACC, anterior insula; **Supplementary Table 1**). As a result of this strongly reduced search volume, the significant threshold for the PPI was set at $P < 0.005$ (uncorrected), with a cluster extent threshold of 15 voxels[42]. For the sake of completeness, we also report differential connectivity effects in other regions not known to be activated during the responders' decision in the ultimatum game (**Supplementary Table 3**).

43. Knoch, D., Schneider, F., Schunk, D., Hohmann, M. & Fehr, E. Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proc. Natl. Acad. Sci. USA* **106**, 20895–20899 (2009).
44. Koch, G. *et al.* rTMS evidence of different delay and decision processes in a fronto-parietal neuronal network activated during spatial working memory. *Neuroimage* **24**, 34–39 (2005).
45. Griskova, I., Ruksenas, O., Dapsys, K., Herpertz, S. & Hoppner, J. The effects of 10-Hz repetitive transcranial magnetic stimulation on resting EEG power spectrum in healthy subjects. *Neurosci. Lett.* **419**, 162–167 (2007).
46. Rossi, S., Hallett, M., Rossini, P.M. & Pascual-Leone, A. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol.* **120**, 2008–2039 (2009).
47. Eisenegger, C., Treyer, V., Fehr, E. & Knoch, D. Time-course of 'off-line' prefrontal rTMS effects: a PET study. *Neuroimage* **42**, 379–384 (2008).
48. Pruessmann, K.P., Weiger, M., Scheidegger, M.B. & Boesiger, P. SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* **42**, 952–962 (1999).
49. Maldjian, J.A., Laurienti, P.J., Kraft, R.A. & Burdette, J.H. An automated method for neuroanatomic and cytoarchitectonic atlas–based interrogation of fMRI data sets. *Neuroimage* **19**, 1233–1239 (2003).
50. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–289 (2002).