

Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction

Thomas Baumgartner,^{1,*} Bastian Schiller,^{1,*} Jörg Rieskamp,² Lorena R. R. Gianotti,¹ and Daria Knoch¹

¹Social and Affective Neuroscience, Department of Psychology, Birmanngasse 8 and ²Center for Economic Psychology, Department of Psychology, University of Basel, Missionsstrasse 62a, CH-4055 Basel, Switzerland

Individuals react to violation of social norms by outgroup members differently than to transgressions of those same norms by ingroup members: namely outgroup perpetrators are punished much more harshly than ingroup perpetrators. This parochial punishment pattern has been observed and extensively studied in social psychology and behavioral economics. Despite progress in recent years, however, little is known about the neural underpinnings of this intergroup bias. Here, we demonstrate by means of transcranial magnetic stimulation (TMS) that the transient disruption of the right, but not the left temporo-parietal junction (TPJ), reduces parochial punishment in a third-party punishment paradigm with real social groups. Moreover, we show that this observed TMS effect on parochial punishment is mediated by a classical punishment motive, i.e. retaliation. Finally, our data suggests that a change in perspective-taking might be the underlying mechanism that explains the impact of right TPJ disruption on retaliation motivation and parochial punishment. These findings provide the first causal evidence that the right TPJ plays a pivotal role in the implementation of parochial behaviors.

Keywords: parochialism; outgroup hostility; ingroup love; normative behavior; third-party punishment; transcranial magnetic stimulation; temporo-parietal junction; intergroup bias

INTRODUCTION

Parochialism, i.e. the differential treatment of ingroup and outgroup members, constitutes a pervasive psychological phenomenon that has a strong impact on human society. It is qualified by a preference for altruistic behavior toward the members of one's ethnic, racial or any other social group, combined with a tendency for indifference, mistrust or even hostility toward outgroup members (Brewer, 1999; Hewstone *et al.*, 2002). For example, people are more likely to help ingroup than outgroup members when they are in an emergency situation (Levine *et al.*, 2005) or when they can relieve these in- and outgroup members from pain (Hein *et al.*, 2010); they cooperate more with ingroup than with outgroup members (Tajfel *et al.*, 1971; Brewer, 1979; Goette *et al.*, 2006; Fehr *et al.*, 2008; Ben-Ner *et al.*, 2009; Chen and Li, 2009; Koopmans and Rebers, 2009), and they punish norm violations less strongly when ingroup members rather than outgroup members commit them (Bernhard *et al.*, 2006; Goette *et al.*, 2006). Hence, the differential behaviors toward members of different social groups underlie acts such as discrimination, terrorism or open intergroup conflict (Bowles, 2009; Bruneau and Saxe, 2010; De Dreu *et al.*, 2010; Haushofer *et al.*, 2010).

Recent evidence from functional magnetic resonance imaging (fMRI) studies suggests that areas involved in social cognition (Gallagher and Frith, 2003; Lieberman, 2007; Van Overwalle, 2009), including the dorsomedial prefrontal cortex (DMPFC) and bilateral temporo-parietal junction (TPJ), might play a role in differentiating between ingroup and outgroup members in perception (Harris and Fiske, 2006), judgments (Freeman *et al.*, 2010; Mathur *et al.*, 2010; Falk *et al.*, 2012) and behavior (Baumgartner *et al.*, 2012).

However, functional imaging studies, although indispensable, do not permit causal inferences about the effect of brain processes on human perception, judgment and behavior, because the observed neural activations could simply be an epiphenomenon or a consequence, and not

necessarily the cause of the perception, judgment or behavior. In contrast, brain stimulation techniques, such as transcranial magnetic stimulation (TMS), which interfere non-invasively with the activity of defined areas in the human cortex, allow researchers to draw causal conclusions about the behavioral impact of the stimulated brain region. No previous study, however, has used brain stimulation techniques to provide causal evidence about the impact of brain areas on parochialism. Using inhibitory low-frequency repetitive TMS (rTMS), we disrupted two key areas involved in social cognition (i.e. right and left TPJ) in order to investigate the causal role these areas play in the differential behavior toward outgroup and ingroup members.

For that purpose, we applied a third-party punishment paradigm with real social groups (soccer fans, see 'Materials and Methods' section for details) and real monetary stakes and consequences for all interaction partners. Third-party punishment is argued to be one of the decisive factors regulating the enforcement of social norms in human society (Bendor and Swistak, 2001; Fehr and Fischbacher, 2004) because third-parties are willing to punish norm violators despite the fact that the norm transgression does not directly affect them. Crucially for this study, third-party punishment has been shown to be strongly shaped by parochialism (Bernhard *et al.*, 2006; Goette *et al.*, 2006; Baumgartner *et al.*, 2012), and lacks any strategic component which could otherwise obscure the parochial behavioral tendencies.

The applied third-party punishment paradigm consists of two decision stages. The first stage took place in the form of an Internet experiment a few weeks before the second stage. Soccer fans were paired with either a member of their own team (ingroup) or someone from a rival team (outgroup) and they played a simultaneous Prisoner's Dilemma Game (PDG) in the roles of either Player A or B. Each player was endowed with 20 points and each had to decide simultaneously whether to keep the points or pass all of them to the other player. Passed points were doubled. Thus, keeping the points equals defection (denoted as D throughout the article) and passing the points equals cooperation (denoted as C throughout the article). For example, if A retained the 20 points while B transferred the 20 points (behavioral pattern DC), then A earned a total of 60 points (40 points from the transfer plus the original 20 points) and B earned nothing.

Received 15 June 2012; Revised 21 February 2013

Advance Access publication 11 March 2013

This project was supported by a grant to DK by the Swiss National Science Foundation (PP00P1_123381). J.R. acknowledges the Swiss National Science Foundation (100014_130352).

Correspondence should be addressed to Thomas Baumgartner, Social and Affective Neuroscience, Department of Psychology, University of Basel, Birmanngasse 8, CH-4055 Basel, Switzerland. E-mail: t.baumgartner@unibas.ch

*The first two authors contributed equally to this work.

In the second stage, which took place in the TMS laboratory, subjects in the role of a third-party (Player C) received the opportunity to punish Player A's or Player B's behavior by assigning punishment points. For that purpose, Player C received an endowment of 10 points at the beginning of each trial, which C could use to finance the assignment of punishment points. Assigning one punishment point costs Player C one point and the sanctioned player three points. Importantly, Player C only could punish the behavior of one player (either A or B) during each of the punishment trials played. In order to simplify the nomenclature, we recoded all Player C's decisions in such a way that A always refers to the player that C can punish, while B always refers to the player that C cannot punish. Player C could retain any points not used for punishment as income (Figure 1).

As we wanted to examine the causal impact of the right and left TPJ on third-parties' parochialism, third-parties were confronted with different combinations of Player A's and B's group affiliations: (i) Both Player A and Player B were ingroup members (referred to in the following as IN/IN), (ii) Player A was an ingroup member and Player B was an outgroup member (referred to as IN/OUT) or (iii) Player A was an outgroup member and Player B an ingroup member (referred to as OUT/IN) (please see Figure 1, and also the supplementary Figure 1 for an example of a decision screen).

With this design, we were able to explore the impact of rTMS on parochial punishment patterns in an intergroup constellation and in an ingroup constellation (Figure 1). The most pronounced and explored parochial punishment pattern in the literature is the stronger punishment of an outgroup member who defects against an ingroup member compared with the less harsh punishment of an ingroup member who defects against an outgroup member (difference value $OUT/IN - IN/OUT$). This punishment pattern represents the comparison of Player C's level of punishment against outgroup perpetrators with that against ingroup perpetrators. We therefore label this punishment pattern as parochial punishment in an 'intergroup constellation' (Figure 1). A few studies (e.g. Bernhard *et al.*, 2006) further reported stronger punishment of an ingroup member who defects against an ingroup member compared with an ingroup member who

defects against an outgroup member (difference value $IN/IN - IN/OUT$). Both perpetrators in this punishment pattern stem from the same social group, i.e. ingroup. We therefore label this 'ingroup constellation'. Thus, by creating these difference values, our study allows us to examine whether parochial punishment behavior is causally modulated by rTMS of the right or left TPJ with regard to the intergroup constellation, the ingroup constellation or both. We expected a potential rTMS effect to be most pronounced in the most conflicting trials, i.e. in DC trials (Player A, who can be punished, defects and Player B cooperates), because studies consistently indicate that this is the situation where the strongest parochial punishment occurs (Goette *et al.*, 2006; Baumgartner *et al.*, 2012).

In order to answer these research questions, we applied low-frequency rTMS for 20 min to either the right TPJ or the left TPJ or sham TMS to healthy subjects in the role of Player C. Low-frequency rTMS for the duration of several minutes leads to a suppression of activity in the stimulated brain regions that outlasts the duration of the TMS train for about half of the duration of the stimulation (Eisenegger *et al.*, 2008; Baumgartner *et al.*, 2011). Each subject participated in only one of the three treatment conditions, and none had experienced TMS previously. We used a stereotaxic infrared system (Brainsight) to localize the right or left TPJ on each subject's individual anatomical brain scan based on the peak coordinates reported in our functional neuroimaging study on parochialism (Baumgartner *et al.*, 2012).

Taken together, our study allows us to clarify the following research questions: *First*, are the right TPJ, left TPJ or both areas causally involved in parochialism? *Second*, if we find such causal evidence, what is the direction of the effect (i.e. a more pronounced parochial punishment pattern or less differential behavior toward ingroup and outgroup members)? *Third*, does rTMS affect all parochial punishment patterns (i.e. parochialism in the intergroup constellation and ingroup constellation) similarly? *Fourth*, we measured several emotions and motivations (anger, justice, improvement of future behavior and retaliation) known to affect punishment decisions (Keller *et al.*, 2010). Thus, if we observe an impact of rTMS on one or both of the

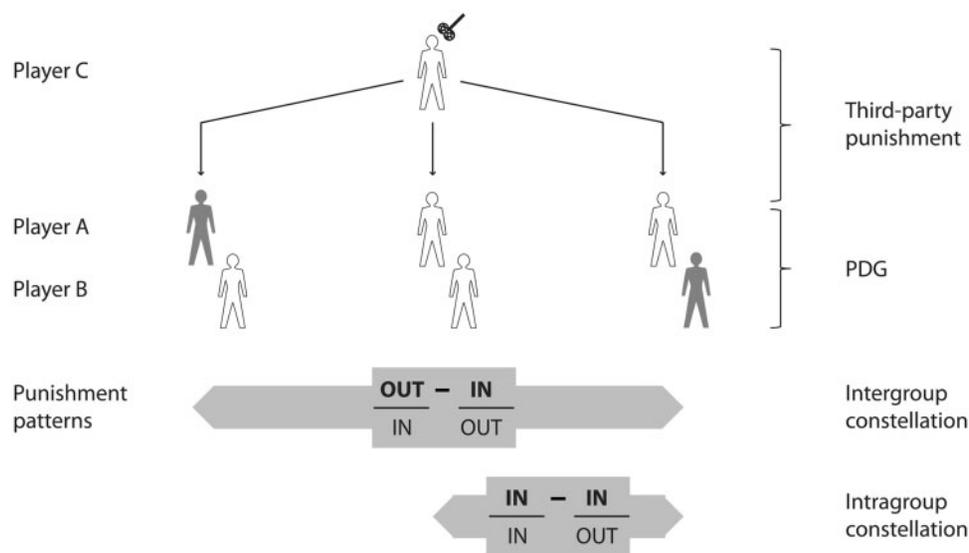


Fig. 1 Schematic representation of the study design. Subjects in the role of Player C received rTMS. After stimulation they were confronted with decisions of Players A and B in a PDG and had the opportunity to assign punishment points to Player A. Players A and B were either ingroup members (IN; depicted in white color) or outgroup members (OUT; depicted in gray color). In order to investigate the impact of rTMS on the two most consistently reported and explored parochial punishment patterns, we calculated the following difference values: OUT/IN vs IN/OUT and IN/IN vs IN/OUT , separately for CC, CD, DC and DD trials.

parochial punishment patterns, are we able to identify a punishment-related emotion or motivation that mediates the effects of disrupting the right or left TPJ on parochialism?

MATERIALS AND METHODS

Subjects

We studied 36 healthy men (mean age \pm s.d. = 24.3 ± 4.2 years). Subjects gave informed written consent prior to participating in the study, which was approved by the local ethics committee. Subjects were randomly assigned to one of the three treatment conditions (rTMS of left TPJ = 12 subjects, rTMS of right TPJ = 13 subjects and sham stimulation = 11 subjects), and none had experienced TMS or a third-party punishment game previously. No subject had a history of psychiatric illness or neurological disorders. Subjects neither experienced serious adverse side effects nor reported scalp pain, neck pain or headaches after the experiment.

Transcranial magnetic stimulation

We applied low-frequency (1 Hz) rTMS to the TPJ for 20 min (1200 pulses) before subjects participated in the third-party punishment paradigm (off-line paradigm), using a Magstim Rapid Magnetic Stimulator (Magstim, Winchester, MA, USA) and a commercially available figure-of-eight coil (70-mm diameter double-circle, air-cooled). To investigate a possible hemispheric laterality in the role of the TPJ on parochialism, we applied rTMS to the right TPJ or to the left TPJ. The creation of stimulation groups receiving rTMS to the same homologous area, but on different hemispheres, is important to control for the potential side effects of rTMS (Abler *et al.*, 2005), including discomfort, irritation and mood changes (see Supplementary Table 2 and Supplementary Figure 5 for statistical evidence that this control procedure was in fact successful). Furthermore, we had a control condition in which we applied sham stimulation for 20 min to the right or left TPJ.

In order to localize the right or left TPJ, we acquired individual anatomical brain images (T1-weighted) of each subject. These individual brain images were used to plan, guide and monitor the stimulation in real time using a stereotaxic infrared system (Brainsight), ensuring that every pulse was delivered to the predetermined cortical location (Sack *et al.*, 2009). The coordinates for stimulation were taken from our functional neuroimaging study (Baumgartner *et al.*, 2012), in which we found correlational evidence that the areas in the left and right TPJ might play a role in parochialism (MNI-coordinates for left TPJ: $x = -45$, $y = -60$, $z = 21$, MNI-coordinates for right TPJ: $x = 57$, $y = -60$, $z = 30$). We transformed these MNI-coordinates to each subject's native brain space using the Brainsight software, in order to make them suitable for neuronavigation with the stereotaxic infrared system.

Stimulation intensity was set at 110% of the individual resting motor threshold (MT), as determined following current guidelines (Rossi *et al.*, 2009). We did not observe any differences with respect to stimulation intensity for the left vs right TMS groups (independent t -test: $t_{(23)} = -0.385$, $P = 0.70$). For the left TPJ, stimulation intensity was on average at 61.6% of maximal stimulator output, and for the right TPJ, the stimulation intensity was on average at 62.3% of maximal stimulator output. The coil was held tangentially to the subject's head with the handle pointing caudally. The rTMS parameters are well within currently recommended guidelines (Rossi *et al.*, 2009) and result in a suppression of excitability of the targeted cortical region for several minutes following completion of the rTMS train (Eisenegger *et al.*, 2008; Baumgartner *et al.*, 2011). The whole paradigm lasted ~ 9 min (including the questions regarding emotions, motives and perspective-taking) and was thus well within the borders of the rTMS after-effect.

Social groups and ingroup identification scale

To explore the impact of rTMS on parochialism, we decided to use naturally occurring social groups. For that purpose, we recruited strong fans of two soccer teams. These two soccer teams are among the best teams in Switzerland (they have won most of the recent championships) and there is a strong rivalry between the two fan groups. Recent work showed that sports team loyalty is one of the most powerful sources for group discrimination and parochialism (Ben-Ner *et al.*, 2009). In order to measure each subject's strength of identification with their favored soccer team, we applied the Sport Spectator Identification Scale (5-point Likert scale; Wann and Branscombe, 1993). The analysis of this identification scale revealed that all recruited subjects in fact strongly identified with their favored soccer team (mean \pm s.e.m. = 3.64 ± 0.10) and that there were no significant differences between the three treatment groups (left TMS, right TMS and sham, for details please see 'Results' section).

The third-party punishment paradigm

We applied a third-party punishment paradigm with real social groups and the involvement of real monetary stakes. The paradigm consisted of two decision stages, one conducted on the Internet using an online questionnaire tool (Unipark) and one conducted in the TMS laboratory. The subjects knew in all stages of the paradigm that there were no repeated interactions and that all interactions were conducted in complete anonymity in order to exclude reputation effects. In the first stage, subjects in the role of Player A or B played a simultaneous PDG with ingroup and outgroup members, where they could decide either to cooperate or defect. Thus, four behavioral patterns were possible: Players A and B cooperate (CC), Players A and B defect (DD), Player A cooperates and Player B defects (CD), and Player A defects and Player B cooperates (DC). In the second stage (in the TMS laboratory), subjects in the role of Player C were confronted with 30 of these decisions and had to decide whether to punish Player A's behavior by assigning punishment points. For that purpose, they received an endowment of 10 points at the beginning of each punishment trial which they could either keep or use to punish Player A. One point assigned for punishment reduced the punished player's income by three points. Points not used for punishment were exchanged into real money and paid to Player C at the end of the experiment (10 points = 1 Swiss Franc, \sim \$1). Player C was always informed about the group affiliations of Players A and B, that is whether the players were from his own or another social group. The PDG decisions of Players A and B were preselected such that each Player C was confronted with the same 30 decision situations, which were presented in random orders (for details, please see Supplementary Table 3). The group affiliation and the behavioral decisions of Players A and B were presented both verbally and visually. Please see Supplementary Figure 1 for an example of a decision screen third-parties saw during the TMS session. The software package z-Tree (Fischbacher, 2007) was used for presenting these screens and for collecting behavioral and timing data.

Finally, please note that all Players C took also part in the PDG. The PDG session took place 6–8 weeks before the TMS session was conducted. At the point in time when Players C played the PDG, they did not know anything about the later TMS session. The advantage of this procedure (i.e. Players C experiencing both situations) is 2-fold: first, the experience of the online PDG substantially increases the credibility of the third-party session conducted in the TMS laboratory where Players C were confronted with real decisions of other players made in the online PDG. Second, it enabled us to investigate pre-existing differences in parochial tendencies across treatment groups.

Trait questionnaires

In order to corroborate our TMS findings (see 'Results' section for details), we measured the following well-established trait questionnaires. The moral foundations questionnaire (Graham *et al.*, 2011) is a 20-item questionnaire (short version), which was developed on the basis of a theoretical model of five universally available (but variably developed) sets of moral intuitions: Harm/Care, Fairness/Reciprocity, Ingroup/Loyalty, Authority/Respect and Purity/Sanctity. It measures on six-point Likert scales how strongly a subject weights these different moral intuitions when deciding whether something is right or wrong. The social dominance orientation questionnaire (Jost and Thompson, 2000) is a two-dimensional 16-item approach to the measurement of social dominance orientation using 7-point Likert scales. One dimension measures general opposition to equality and the other dimension measures support for group-based dominance. We also applied six items of the personal norm of reciprocity scale (Perugini *et al.*, 2003) which measures (on two 7-point Likert scales) a subject's tendency to reward another person's positive behavior (positive reciprocity, three items) and to punish negative behavior (negative reciprocity, three items). Finally, we applied the injustice sensitivity scale from the observer perspective (Schmitt *et al.*, 2005). This 10-item questionnaire measures on a 7-point Likert scale how strongly subjects in the role of an observer (i.e. a third-party) psychologically and emotionally react to observed injustice. All trait questionnaires were measured 30 min after the end of the TMS administration, i.e. when the rTMS effect was expected to have dissipated (Eisenegger *et al.*, 2008).

Punishment-related emotions and motivations

In order to examine whether disrupting key areas of social cognition impacts punishment-related emotions and motivations, we obtained ratings for four emotions and motivations known to play a role in punishment-related decision-making (anger, retaliation, improvement of future behavior and justice; Keller *et al.*, 2010). At the end of the punishment paradigm, subjects were confronted with some of the previously experienced decision situations and were asked to indicate on 7-point Likert scales how strongly these emotions and motivations had affected their decision to assign punishment points. Due to time limits of the TMS after-effect, we only obtained these emotions/motives for DC trials (Player A, who can be punished, defects, whereas Player B cooperates). We decided to focus on these trials, because here the strongest parochial punishment patterns can be found (Bernhard *et al.*, 2006; Baumgartner *et al.*, 2012).

Item on perspective-taking

After the ratings of the punishment-related emotions and motivations, subjects were asked to answer the following statement regarding trials with unilateral defection (DC trials) committed by an ingroup or an outgroup perpetrator in the role of Player A: 'Putting myself in the position of Player A helped me to make my punishment decision'. Subjects had to indicate on 6-point Likert scales whether they agreed with the statement or not (ranging from 'I agree not at all' to 'I agree completely').

Statistical analyses

All statistical analyses were run with SPSS 20 for PC (SPSS Inc., Chicago, IL, USA). Please see 'Results' section for details about the statistical tests conducted, including independent, dependent and one sample *t*-tests, chi-square tests as well as repeated measures ANOVAs. Results were considered significant at the level of $P < 0.05$ (two-tailed). In case of significant multivariate effects, *post hoc* paired *t*-tests were computed using the Bonferroni correction according to Holm (1979). As effect size measure η^2 is reported. We controlled for different

covariates in our analyses, including trait questionnaires, ingroup identification score and payment in the PDG. Crucially, all reported findings hold, irrespective of whether we controlled for the mentioned covariates (see 'Results' section for details).

The SPSS macro 'Process' (Preacher and Hayes, 2008; Hayes, 2013) was used for all mediation analyses. It investigates whether an independent variable (*X*, in our case, the treatment group variable) affects a dependent variable (*Y*, in our case, the parochial punishment pattern) through one or more intervening variables, or mediators (*M*, in our case, the item on perspective-taking and/or the punishment-related emotions/motivations). Significance of the mediated, indirect effect through *M* was tested with a bootstrapping strategy (Preacher and Hayes, 2008). Bootstrapping involves the repeated extraction of samples from the data set and the estimation of the indirect effect in each resampled data set. By repeating this process thousands of times, an empirical approximation of the sampling distribution of the indirect effect is built and used to construct confidence intervals. We used 10 000 bootstrap samples to generate bootstrap confidence intervals (90%, 95% and 99%) for the indirect effects.

RESULTS

Parochial punishment patterns

In clear agreement with the literature, we found on average, irrespective of treatment groups, a strong impact of the perpetrator's group membership on third-parties' invested punishment points in the intergroup constellation, i.e. a much more harsh punishment of an outgroup perpetrator in OUT/IN compared with an ingroup perpetrator in IN/OUT, which was most pronounced in the DC trials (mean punishment differences DC trials \pm s.e.m. = 4.55 ± 0.55 , dependent *t*-test: $t_{(35)} = 8.1$, $P < 0.001$), and less pronounced in the DD trials (mean punishment difference DD trials \pm s.e.m. = 2.48 ± 0.50 , dependent *t*-test: $t_{(35)} = 4.9$, $P < 0.001$). Thus, findings indicate that an outgroup member who defects against a cooperating (or defecting) ingroup member is more severely punished than an ingroup member who defects against a cooperating (or defecting) outgroup member. Furthermore, we found a similar parochial punishment pattern on average, irrespective of treatment groups, in the intragroup constellation, albeit as expected at a lower level and only in DC trials, i.e. an ingroup perpetrator in IN/IN was more strongly punished than an ingroup perpetrator in IN/OUT (mean punishment difference DC trials \pm s.e.m. = 0.84 ± 0.28 , dependent *t*-test: $t_{(35)} = 2.93$, $P = 0.006$). Thus, an ingroup member who defects against a cooperating ingroup member is punished more strongly than an ingroup member who defects against a cooperating outgroup member.

Of primary interest for this study is the question whether rTMS of right and/or left TPJ might be able to affect these parochial punishment patterns. Accordingly, we calculated a repeated-measures ANOVA with the within-subject factor behavioral trials (CC, CD, DC and DD) and the between-subject factor treatment group (left TMS, right TMS and sham) for the two parochial punishment patterns (OUT/IN–IN/OUT and IN/IN–IN/OUT).

Results provide evidence that rTMS in fact modulates the stronger punishment of an outgroup perpetrator in OUT/IN compared with an ingroup perpetrator in IN/OUT, indicated by a significant main effect of treatment group ($F_{(2,32)} = 3.38$, $P = 0.046$, $\eta^2 = 0.17$) and a significant interaction effect of treatment group \times behavioral trials ($F_{(6,96)} = 2.65$, $P = 0.020$, $\eta^2 = 0.14$). *Post hoc* independent *t*-tests revealed that the right TMS group exhibits significantly diminished parochial punishment behavior than either the left TMS group ($P = 0.009$) or the sham group ($P = 0.041$). This was unique to the DC trials (Figure 2A). A similar pattern emerged in the DD trials, but only the difference between the right TMS and left TMS group

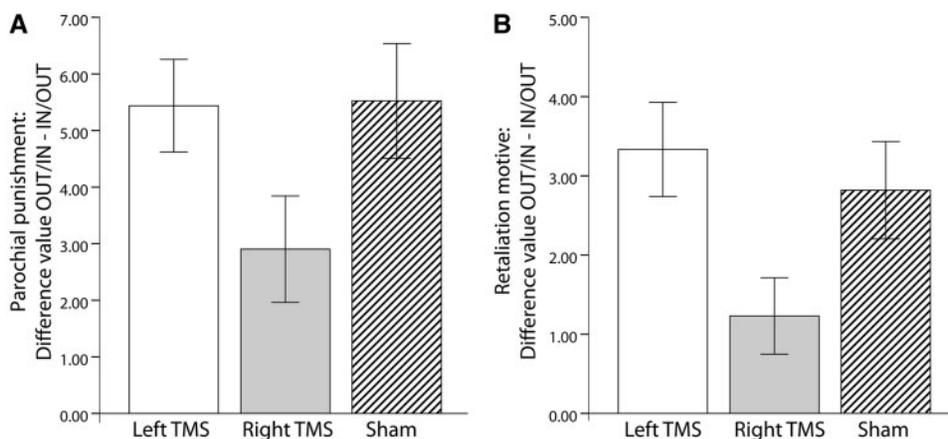


Fig. 2 Treatment group differences in parochial punishment and retaliation motive. Treatment group differences (means \pm s.e.m.) in (A) parochial punishment pattern and (B) retaliation motive in a conflicting intergroup constellation (OUT/IN vs IN/OUT). (A) Depicted is third-parties' stronger punishment of an outgroup perpetrator (OUT/IN) compared with an ingroup perpetrator (IN/OUT) and its modulation by rTMS. Subjects whose right TPJ is disrupted exhibit a much lower parochial punishment (smaller difference score) than those in the other two treatment groups (all $P < 0.05$). This suggests that the right TPJ causally affects third-parties' parochial punishment in intergroup conflict. (B) Depicted is third-parties' stronger retaliation motive toward an outgroup perpetrator (OUT/IN) compared with an ingroup perpetrator (IN/OUT) and its modulation by rTMS. Subjects whose right TPJ is disrupted exhibit a much lower parochial retaliation motivation (smaller difference score) than those in the other two treatment groups (all $P < 0.05$).

was significant ($P = 0.011$), while the difference between the right TMS and sham group did not reach statistical significance ($P = 0.260$). A closer inspection of the observed TMS effect revealed that the diminished parochial punishment is mainly caused by a reduction in OUT/IN, i.e. third-parties with disrupted right TPJ show a less severe punishment of an outgroup perpetrator who defects against a cooperating ingroup member (see Supplementary Figure 2A). Finally, as expected, there were no TMS effects with regard to the CC and CD trials, where punishment was virtually absent in all treatment groups (all $P > 0.18$).

Next, we checked whether we find evidence in our data that rTMS causally modulates the stronger punishment of an ingroup perpetrator in IN/IN compared with an ingroup perpetrator in IN/OUT. Results of the repeated-measures ANOVA revealed neither a main effect of treatment group ($F_{(2, 33)} = 1.553, P = 0.23, \eta^2 = 0.08$) nor an interaction effect of treatment group \times behavioral trials ($F_{(6, 99)} = 1.473, P = 0.21, \eta^2 = 0.08$), suggesting that the stronger punishment in the intragroup constellation, i.e. the stronger punishment of an ingroup perpetrator in IN/IN compared with an ingroup perpetrator in IN/OUT, is neither causally modulated by rTMS of the right nor the left TPJ.

Punishment-related emotions and motivations as mediators

We next investigated whether punishment-related emotions or motivations (anger, retaliation, improvement of future normative behavior and justice) might provide an explanation for the parochial punishment patterns and whether they might mediate the observed TMS effect of disrupting the right TPJ on parochialism. Due to time limitations of the TMS after-effect, we only obtained these emotions and motives for the DC trials, where we expected and found the strongest parochial punishment patterns. Interestingly, only one of the four obtained emotions/motives, namely retaliation, shows a treatment group effect and is capable of mediating the observed TMS effect.

Third-parties report a more severe retaliation motive in OUT/IN vs IN/OUT (dependent t -test: $t_{(35)} = 6.90, P < 0.001$), i.e. the motive to retaliate against an outgroup perpetrator who defected against an ingroup member is much stronger compared with an ingroup member who defected against an outgroup member. In contrast, we find no difference with regard to this motive in the intragroup

constellation (i.e. IN/IN vs IN/OUT; dependent t -test: $t_{(35)} = 0.00, P = 1.00$). Moreover, we found a clear TMS effect on the retaliation motive (but not on the other emotion/motivations; univariate ANOVA with between-subject factor treatment group: $F_{(2, 33)} = 4.00, P = 0.028, \eta^2 = 0.20$, all other univariate ANOVAs regarding the other emotion/motives: $P > 0.10$). In other words, the perpetrator's group membership has a stronger impact on the retaliation motive in both the left TMS and sham group compared with the right TMS group (independent t -tests: left TMS vs right TMS: $P = 0.01$, sham vs right TMS: $P = 0.05$, Figure 2B). This means that the disruption of the right TPJ by TMS seems to narrow the difference in motivation to retaliate against out- and ingroup perpetrators. A closer inspection of this TMS effect revealed that an attenuated retaliation motive in OUT/IN mainly causes this narrowed difference, i.e. third-parties whose right TPJ was disrupted reported a reduced retaliation motive toward an outgroup perpetrator who defected against a cooperating ingroup member compared with the sham group or left TMS group (see Supplementary Figure 2B).

In order to test whether the retaliation motive mediates the TMS effect on parochial punishment behavior, we conducted a mediation analysis (Preacher and Hayes, 2008). For this analysis, we merged the left TMS and sham groups into one group, because as shown above, these two groups demonstrated both an undistinguishable parochial punishment pattern and retaliation motive. The path diagram of the mediation analysis (Figure 3) shows that disrupting the right TPJ reduces the difference in the retaliation motive toward an outgroup compared with an ingroup perpetrator which in turn causes a smaller difference in the punishment meted out to an outgroup or ingroup perpetrator in DC trials. Please note that if we conduct two separate analyses for the left TMS and sham group, both mediation analyses are also significant (see Supplementary Figure 3 for details). Moreover, if we conduct mediation analyses with the other emotions/motives as mediators (anger, justice and improvement of future behavior), none of the meditation analyses becomes significant (bootstrapping statistics comparing Path c' vs Path c : all $P > 0.20$).

Possible process that might explain the observed TMS effect on retaliation motive and punishment behavior

In order to investigate potential functional processes that might be affected by the disruption of the TPJ, we obtained an item on

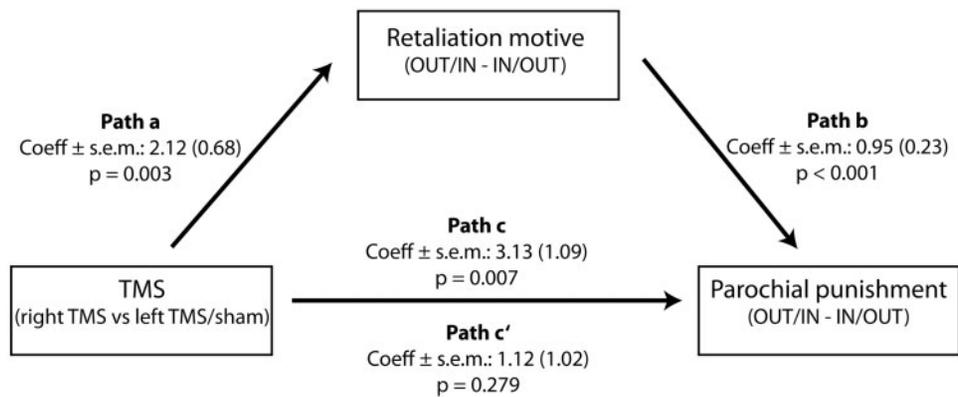


Fig. 3 Mediation analysis. Depicted is the path diagram (including regression coefficients \pm s.e.m. and P -values) of the mediation analysis demonstrating that the retaliation motive mediates the impact of disrupting the right TPJ on parochial punishment pattern in a conflicting intergroup constellation. All four requirements for a mediation effect are satisfied: Path a, Path b, and Path c are significant, and Path c' is significantly smaller than Path c. In detail, Path a represents the effect of TMS (right TPJ vs left TPJ/Sham) on the retaliation motive. Path b represents the impact of the retaliation motive on parochial punishment controlling for the TMS effect. Together, Path a and Path b represent the indirect (mediated) effect of TMS on parochial punishment through the retaliation motive. Path c' represents the direct effect of TMS on parochial punishment and is calculated controlling for the indirect, mediated effect. Path c represents the total (mediated and direct) effect of TMS on parochial punishment. Finally, the decisive statistical test to examine whether mediation occurs is the statistical test between Path c and Path c' . Bootstrapping statistics (see 'Materials and Methods' section for details) revealed that Path c' is significantly smaller than Path c ($P < 0.01$), providing evidence that the retaliation motive is indeed a significant mediator.

perspective-taking processes (see 'Materials and Methods' section) because one of the key social cognitive processes associated with the TPJ is perspective-taking, also known as mentalizing (Gallagher and Frith, 2003; Van Overwalle, 2009; Young *et al.*, 2007, 2010). With this item, we asked subjects how well they were able to put themselves into the shoes of an ingroup or outgroup perpetrator when they decided to punish. TMS over the right TPJ (but not over the left TPJ or sham) induced a change in mentalizing processes. More specifically, both the left TMS and the sham group reported that it was easier to put themselves into the position of an ingroup perpetrator than in an outgroup perpetrator (dependent t -test left TMS group: $t_{(12)} = 3.83$, $P = 0.003$; sham group: $t_{(11)} = 3.13$, $P = 0.011$), whereas the right TMS group reported no differences in regard to how well they were able to put themselves into the position of an ingroup and outgroup perpetrator (dependent t -tests right TMS group: $t_{(13)} = 1.32$, $P < 0.211$).

Moreover, we created a 'mentalizing bias' by subtracting perspective-taking with outgroup perpetrators from perspective-taking with ingroup perpetrators. Thus, high values on this mentalizing bias indicate that participants better understood the behavior of ingroup perpetrators compared with that of outgroup perpetrators, whereas low values mean that individuals did not differ in mentalizing toward ingroup and outgroup perpetrators. We specified a mediation model with both the mentalizing bias and the differences in the motive to retaliate as serial mediators. This mediation model was significant (at $P < 0.05$) and suggested that disrupting the right TPJ (compared with rTMS of left TPJ or sham stimulation) reduces the mentalizing bias, which in turn reduces the differences in the motive to retaliate, which then leads to a smaller difference in the punishment level meted out to an outgroup and ingroup perpetrator (see Supplementary Figure 4).

No treatment group differences in dispositional traits, parochial patterns in the PDG and ingroup identification scores

In a final set of analyses, we checked whether we can provide evidence that the observed treatment group effect is not due to pre-existing differences in dispositional characteristics across treatment groups. We measured a series of trait questionnaires (at a time when the TMS effect can be expected to have worn off) with regard to the sensitivity toward injustice (injustive sensitivity scale; Schmitt *et al.*, 2005), propensity to reciprocate kind or hostile acts (personal norm of reciprocity, Perugini *et al.*, 2003), moral foundations (moral foundations

questionnaire, Graham *et al.*, 2011) and social dominance orientation (Jost and Thompson, 2000). Univariate ANOVAs revealed no difference in any scale across treatment groups (all $P > 0.11$, please see Supplementary Table 1 for details). Furthermore, if we use all these scales as covariates in the two reported analyses showing treatment group effects in DC trials, both treatment group effects remain significant (parochial punishment in intergroup constellation: $F_{(2,22)} = 3.92$, $P = 0.035$, $\text{ETA}^2 = 0.26$; retaliation motive in intergroup constellation: $F_{(2,23)} = 5.95$, $P = 0.01$, $\text{ETA}^2 = 0.34$).

A few weeks before subjects took part in the TMS session, they also played the described PDG. Since this paradigm also evokes strong parochial patterns (i.e. stronger cooperation rate with ingroup members, see for example Goette *et al.*, 2006; Koopmans and Rebers, 2009), we have the possibility to check whether there were pre-existing differences with regard to parochial behavior across treatment groups. Findings revealed that all treatment groups demonstrated a similar parochial behavior, qualified by a much higher cooperation rate when interacting with ingroup members compared with outgroup members (difference in cooperation rate, i.e. ingroup $>$ outgroup: left TMS group: 75%, right TMS group: 70% and sham group: 73%). Chi-square tests revealed no difference across treatment groups regarding the cooperation rates in the PDG (all $P > 0.89$). In other words, we found no pre-existing differences across treatment groups in the PDG.

Finally, because studies on ingroup and outgroup evaluations showed that ingroup biases in judgments depend on how strongly subjects identify with ingroup members (Aberson *et al.*, 2000), we also checked for potential pre-existing differences with regard to ingroup identification. We used an ingroup identification questionnaire (Wann and Branscombe, 1993; 5-point Likert scale) and found no differences across treatment groups (univariate ANOVA with between-subject factor treatment group: $F_{(2,33)} = 0.15$, $P = 0.86$, $\text{ETA}^2 = 0.01$). Moreover, if we use this identification score as covariate, the observed TMS effects on parochial punishment (repeated measures ANOVA: $F_{(2,31)} = 3.69$, $P = 0.036$, $\text{ETA}^2 = 0.19$) and retaliation motive (repeated measures ANOVA: $F_{(2,32)} = 4.19$, $P = 0.024$, $\text{ETA}^2 = 0.21$) remain significant.

Taken together, these findings provide strong evidence that the observed TMS effect can be explained neither by pre-existing differences in dispositional characteristics nor by the strength of ingroup identification.

DISCUSSION

Our results indicate a lateralized effect of a disrupted function of the TPJ on parochialism in intergroup conflict. We found that disrupting the function of the right, but not left TPJ, significantly diminishes the impact of the perpetrator's group membership on punishment decisions. Third-parties whose right TPJ was disrupted by rTMS treat an outgroup perpetrator who defected against a cooperating ingroup member and an ingroup perpetrator who committed the very same norm violation toward an outgroup member more similarly than do third-parties with a disrupted left TPJ or sham stimulation. Our findings further show that third-parties' punishment behavior in the context of an intragroup constellation is not affected by TMS of the right or left TPJ. Thus, the findings of our study suggest that the TPJ plays a decisive role in modulation of parochialism in intergroup conflict, but not in intragroup conflict.

Furthermore, we find that the retaliation motive mediates the impact of the disrupted right TPJ on parochial punishment. Disrupting the function of the right TPJ seems to narrow the difference in motivation to retaliate against outgroup and ingroup perpetrators, which in turn reduces the parochial punishment magnitude. Finally, we find that disruption of the right TPJ induced a change in mentalizing processes, i.e. the right TMS group reported no differences in regard to how well they were able to put themselves into the position of an ingroup and outgroup perpetrator, i.e. they process the intentions and goals behind the defective act more similarly. As a consequence of this reduced mentalizing bias, the strong retaliation motive toward an outgroup perpetrator is attenuated, which in turn reduces the punishment magnitude.

This study is based on the results of a previous functional neuroimaging study on parochialism (Baumgartner *et al.*, 2012), in which we found increased activity in the left TPJ, right TPJ and DMPFC when third-parties were confronted with defecting ingroup members compared with defecting outgroup members. This study was set up to draw causal conclusions about the necessity of the TPJ in parochial punishment. Our findings suggest that this is the case for the processes in the right, but not left TPJ. The fact that no TMS effect is observed after disruption of the left TPJ indicates that this area does not play a causal role in parochial punishment. Since our previous fMRI study on parochial punishment linked the functional connectivity between left TPJ and DMPFC to the punishment of ingroup perpetrators (i.e. the stronger the connectivity, the lower the punishment of ingroup perpetrators), we might speculate that the DMPFC alone is 'capable' of implementing the processes required for the punishment of ingroup perpetrators and does not require the additional contribution of the left TPJ. Future research could investigate how disrupting the function of the DMPFC impacts parochial punishment behavior.

The diminished parochialism in subjects with a disrupted right TPJ was evident in the fact that outgroup perpetrators were treated more like ingroup perpetrators, i.e. punished less strongly and not due to treating ingroup perpetrators more like outgroup perpetrators. A closer look at data of our previous fMRI study on parochialism (Baumgartner *et al.*, 2012) might explain the observed direction of the TMS finding. With a less stringent significance threshold, we found correlative evidence (at $P < 0.05$, $r = 0.498$, $ETA^2 = 25\%$) that the functional connectivity between right TPJ and DMPFC is linked to the punishment of outgroup perpetrators: The stronger the functional connectivity between these two areas, the more severely outgroup perpetrators are punished (see Supplementary Figure 6). This might explain the direction of the observed TMS effect, namely that TMS over the right TPJ reduced the punishment of outgroup perpetrators.

Certain regions nearby the TPJ (e.g. in the inferior parietal cortex) are known to be involved in attention and processing of more general task

performance (Corbetta *et al.*, 2000; Mitchell, 2008; Scholz *et al.*, 2009; Van Overwalle, 2011). Thus, the question arises whether our findings are rather unspecific and due to diminished attention or generally diminished task performance. We believe that this is unlikely for the following reasons: *First*, the rTMS effect is only observable in the condition where perpetrators defected against cooperating interaction partners (DC), but not in conditions where there was no reason to punish (i.e. in the case of cooperative behavior in the PDG). Moreover, we only find a rTMS effect in the intergroup constellation, i.e. when the social category of the perpetrator to whom punishment could be assigned is different, but not in the intragroup constellation, i.e. where both perpetrators belong to the same social category, namely the ingroup. Such highly specific findings would be difficult to explain with unspecific attentional or cognitive processes because they would be more likely to affect all punishment patterns and conditions in a similar way. *Second*, we find no treatment group effects on response times. Repeated-measures ANOVA with within-subject factors behavioral trials (CC, CD, DC and DD), group constellations (OUT/IN, IN/OUT, IN/IN) and between-subject factor treatment groups (left TMS, right TMS, sham) revealed neither a main effect of treatment groups nor interaction effects of treatment groups with one or both other factors (all $P > 0.487$). These findings are not consistent with the assumption of an unspecific effect of attention or more general task performance because such an effect would rather manifest itself in slower response times.

Summing up, this is the first study providing causal evidence that the right TPJ plays a pivotal role in modulating parochialism in intergroup conflict. Based on the evidence of decreased parochialism after right TPJ disruption together with the observation that the differential behaviors toward members of different social groups often underlie acts such as discrimination, it is tempting to conclude that a disrupted TPJ could lead to the world being a better place. Such diminished parochialism is only evident, however, if the brain is not functioning properly. Therefore—although parochialism in its extreme can lead to severe and detrimental outcomes (Haushofer *et al.*, 2010)—treating an ingroup member differently than an outgroup member appears to have a necessary function that has developed through evolution. As a social species, individuals have evolved living in different social groups and cognitive adaptations must therefore have developed to differentially treat individuals according to group membership (Efferson *et al.*, 2008). It thus seems plausible that along with the cognitive adaptations, evolution has engineered neural adaptations to distinguish the behavior of ingroup and outgroup members and to differentiate the punishment of norm transgressions depending on the perpetrator's social category. Our study suggests that these neural adaptations have developed in the right TPJ.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

Conflict of Interest

None declared.

REFERENCES

- Aberson, C.L., Healy, M., Romero, V. (2000). Ingroup bias and self-esteem: a meta-analysis. *Personality and Social Psychology Review*, 4(2), 157–73.
- Abler, B., Walter, H., Wunderlich, A., et al. (2005). Side effects of transcranial magnetic stimulation-biased task performance in a cognitive neuroscience study. *Brain Topography*, 17(4), 193–6.
- Baumgartner, T., Gotte, L., Gugler, R., Fehr, E. (2012). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*, 33(6), 1452–69.
- Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11), 1468–74.

- Ben-Ner, A., McCall, B.P., Stephane, M., Wang, H. (2009). Identity and in-group/out-group differentiation in work and giving behaviors: experimental evidence. *Journal of Economic Behavior & Organization*, 72(1), 153–70.
- Bendor, J., Swistak, P. (2001). The evolution of norms. *American Journal of Sociology*, 106(6), 1493–545.
- Bernhard, H., Fischbacher, U., Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105), 912–5.
- Bowles, S. (2009). Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science*, 324(5932), 1293–8.
- Brewer, M. (1979). In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307–24.
- Brewer, M. (1999). The psychology of prejudice: ingroup love or outgroup hate? *Journal of Social Issues*, 55(3), 429–44.
- Bruneau, E.G., Saxe, R. (2010). Attitudes towards the outgroup are predicted by activity in the precuneus in Arabs and Israelis. *Neuroimage*, 52(4), 1704–11.
- Chen, Y., Li, S.X. (2009). Group identity and social preferences. *American Economic Review*, 99(1), 431–57.
- Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, 3(3), 292–7.
- De Dreu, C.K., Greer, L.L., Handgraaf, M.J., et al. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science*, 328(5984), 1408–11.
- Efferson, C., Lalive, R., Fehr, E. (2008). The coevolution of cultural groups and ingroup favoritism. *Science*, 321(5897), 1844–9.
- Eisenegger, C., Treyer, V., Fehr, E., Knöch, D. (2008). Time-course of “off-line” prefrontal rTMS effects—a PET study. *Neuroimage*, 42(1), 379–84.
- Falk, E.B., Spunt, R.P., Lieberman, M.D. (2012). Ascribing beliefs to ingroup and outgroup political candidates: neural correlates of perspective-taking, issue importance and days until the election. *Philosophical Transactions of the Royal Society B—Biological Sciences*, 367(1589), 731–43.
- Fehr, E., Bernhard, H., Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454(7208), 1079–83.
- Fehr, E., Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4), 185–90.
- Fischbacher, U. (2007). z-Tree: zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2), 171–8.
- Freeman, J.B., Schiller, D., Rule, N.O., Ambady, N. (2010). The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Human Brain Mapping*, 31(1), 150–9.
- Gallagher, H.L., Frith, C.D. (2003). Functional imaging of ‘theory of mind’. *Trends in Cognitive Sciences*, 7(2), 77–83.
- Goette, L., Huffman, D., Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: evidence using random assignment to real social groups. *American Economic Review*, 96(2), 212–6.
- Graham, J., Nosek, B.A., Haidt, J., Iyer, R., Koleva, S., Ditto, P.H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101(2), 366–85.
- Harris, L.T., Fiske, S.T. (2006). Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychological Sciences*, 17(10), 847–53.
- Haushofer, J., Biletzki, A., Kanwisher, N. (2010). Both sides retaliate in the Israeli–Palestinian conflict. *Proceedings of the National Academy of Sciences of the United States of America*, 107(42), 17927–32.
- Hayes, A.F. (2013). *An Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach*. New York: Guilford Press.
- Hein, G., Silani, G., Preuschoff, K., Batson, C.D., Singer, T. (2010). Neural responses to ingroup and outgroup members’ suffering predict individual differences in costly helping. *Neuron*, 68(1), 149–60.
- Hewstone, M., Rubin, M., Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575–604.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Jost, J.T., Thompson, E.P. (2000). Group-based dominance and opposition to equality as independent predictors of self-esteem, ethnocentrism, and social policy attitudes among African Americans and European Americans. *Journal of Experimental Social Psychology*, 36(3), 209–32.
- Keller, L.B., Oswald, M.E., Stucki, I., Gollwitzer, M. (2010). A closer look at an eye for an eye: Laypersons’ punishment decisions are primarily driven by retributive motives. *Social Justice Research*, 23(2–3), 99–116.
- Koopmans, R., Rebers, S. (2009). Collective action in culturally similar and dissimilar groups: an experiment on parochialism, conditional cooperation, and their linkages. *Evolution and Human Behavior*, 30(3), 201–11.
- Levine, M., Prosser, A., Evans, D., Reicher, S. (2005). Identity and emergency intervention: how social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, 31(4), 443–53.
- Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, 58, 259–89.
- Mathur, V.A., Harada, T., Lipke, T., Chiao, J.Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *Neuroimage*, 51(4), 1468–75.
- Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, 18(2), 262–71.
- Perugini, M., Callucci, M., Presaghi, F., Ercolani, A.P. (2003). The personal norm of reciprocity. *European Journal of Personality*, 17(4), 251–83.
- Preacher, K.J., Hayes, A.F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–91.
- Rossi, S., Hallett, M., Rossini, P.M., Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, 120(12), 2008–39.
- Sack, A.T., Cohen Kadosh, R., Schuhmann, T., Moerel, M., Walsh, V., Goebel, R. (2009). Optimizing functional accuracy of TMS in cognitive studies: a comparison of methods. *Journal of Cognitive Neuroscience*, 21(2), 207–21.
- Schmitt, M., Gollwitzer, M., Maes, J., Arbach, D. (2005). Justice sensitivity: assessment and location in the personality space. *European Journal of Psychological Assessment*, 21(3), 202–11.
- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One*, 4(3), e4869.
- Tajfel, H., Billig, M., Bundy, R., Flament, C. (1971). Social categorization in intergroup behaviour. *European Journal of Social Psychology*, 1, 149–78.
- Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, 30(3), 829–58.
- Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage*, 54(2), 1589–99.
- Wann, D.L., Branscombe, N.R. (1993). Sports fans—measuring degree of identification with their team. *International Journal of Sport Psychology*, 24(1), 1–17.
- Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences of the United States of America*, 107(15), 6753–8.
- Young, L., Cushman, F., Hauser, M., Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), 8235–40.