Neuropsychologia 77 (2015) 400–408

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Increasing the role of belief information in moral judgments by stimulating the right temporoparietal junction

Roberta Sellaro [a,*], Berna Güroğlu [b], Michael A. Nitsche [c,d,e], Wery P.M. van den Wildenberg [f], Valentina Massaro [a], Jeffrey Durieux [a], Bernhard Hommel [a], Lorenza S. Colzato [a]

[a] Cognitive Psychology Unit & Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[b] Developmental and Educational Psychology Unit & Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[c] Department of Clinical Neurophysiology, Georg-August University Göttingen, Göttingen, Germany
[d] Leibniz Research Centre for Working Environment and Human Resources, Dortmund, Germany
[e] Department of Neurology, University Medical Hospital Bergmannsheil, Bochum, Germany
[f] Department of Psychology & Amsterdam Brain & Cognition (ABC), University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 25 February 2015
Received in revised form
28 August 2015
Accepted 10 September 2015
Available online 13 September 2015

Keywords:
Right temporoparietal junction
Moral judgment
Belief
Transcranial direct current stimulation

ABSTRACT

Morality plays a vital role in our social life. A vast body of research has suggested that moral judgments rely on cognitive processes mediated by the right temporoparietal junction (rTPJ), an area thought to be involved in belief attribution. Here we assessed the role of the rTPJ in moral judgments directly by means of transcranial direct current stimulation (tDCS) – a non-invasive brain stimulation technique that, by applying a weak current to the scalp, allows modulating cortical excitability of the area being stimulated. Participants were randomly and equally assigned to receive anodal stimulation (to increase cortical excitability), cathodal stimulation (to decrease cortical excitability), or sham (placebo) stimulation over the rTPJ before completing a moral judgment task. Participants read stories in which protagonists produced either a negative or a neutral outcome based on either a negative or a neutral belief that they were causing harm or no harm, respectively. Results revealed a selective group difference when judging the moral permissibility of accidental harms (belief neutral, outcome negative), but not intentional harms (belief negative, outcome negative), attempted harms (belief negative, outcome neutral), or neutral acts (belief neutral, outcome neutral). Specifically, participants who received anodal stimulation assigned less blame to accidental harms compared to participants who received cathodal or sham stimulation. These results are consistent with previous findings showing that the degree of rTPJ activation reflects reliance on the agent's innocent intention. Crucially, our findings provide direct evidence supporting the critical role of the rTPJ in mediating belief attribution for moral judgment.

1. Introduction

A fundamental aspect of social cognition is the ability to infer what is going on inside other people's head, what motivates behaviors and actions of others. The ability to attribute minds to others and to infer their mental states – including thoughts, beliefs, desires, intentions, and motivations (i.e., mental state reasoning or theory of mind, ToM; Premack and Woodruff, 1978) – plays a vital role in social interactions (Baron-Cohen, 1995). Mental state reasoning enables us to explain people's past actions, to predict people's future behavior, and drives our moral judgments (Baron-Cohen et al., 2013).

Morality is a building block of human societies and moral cognition is the product of a complex process resulting from the interplay between genes, environment, and the brain (Fumagalli and Priori, 2012). In recent years, several studies have been carried out to assess the cognitive and neural correlates that underlie human moral cognition (for reviews, see Young and Dungan (2012), Fumagalli and Priori (2012), Moll et al. (2005)). Neuroscientists have shed light on several factors that seem to play a crucial role in mediating morality, such as emotion (Nadelhoffer, 2006; Young et al., 2006; Nichols, 2002; Greene et al., 2001), desires (e.g., Cushman, 2008), the magnitude of an action's consequences (Greene et al., 2001; Cushman et al., 2006), situational constrains (Woolfolk et al., 2006), prior record (Kliemann et al., 2008), the means used to cause harms (Cushman et al., 2006; Greene et al., 2004), luck (Young et al. 2010c), and beliefs (Koster-

* Correspondence to: Cognitive Psychology Unit, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.
E-mail address: r.sellaro@fsw.leidenuniv.nl (R. Sellaro).

http://dx.doi.org/10.1016/j.neuropsychologia.2015.09.016
0028-3932/

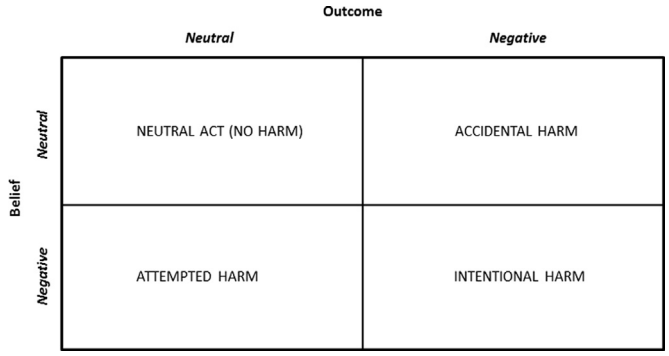

Fig. 1. Schematic representation of the experimental design employed by Young and colleagues (e.g., Young et al., 2010b, 2007; Young and Saxe, 2009b) as well as in the present study. As shown by the picture, participants are confronted with 4 types of moral actions resulting from the Belief (neutral vs. negative) by Outcome (neutral vs. negative) combination.

Hale et al., 2013; Young et al., 2010b; Cushman, 2008; Young and Saxe, 2009a, 2009b, 2008; Young et al., 2007; Cushman et al. 2006).

An interesting and extensive line of research has focused on the neural mechanisms supporting mental state reasoning and belief attribution for third party moral judgments (for reviews, see Young (2013) and Young and Waytz (2013)). In a series of studies, Young and colleagues (Koster-Hale et al., 2013; Young et al., 2010b; Young and Saxe, 2009b, 2008; Young et al., 2007; Cushman et al., 2006) examined the role of agents' beliefs vs. the consequences of agents' actions for moral judgments. These studies employed a study design where participants were confronted with scenarios in which protagonists produce either a negative outcome (harm to another person; e.g., poisoning someone and causing his/her death) or a neutral outcome (no harm), based on the belief that they would cause the negative outcome or the neutral outcome (negative belief vs. neutral belief, respectively; e.g., putting sugar in someone's coffee believing it to be poison vs. sugar). This design results in four (Belief × Outcome) potential moral actions that participants rate on a forbidden–permissible scale (Fig. 1). Across different studies (Koster-Hale et al., 2013; Young et al., 2010b, 2007; Young and Saxe, 2009b, 2008), behavioral results were consistent in showing that, when evaluating a moral action, people consider not just the consequences of that action but the beliefs behind it as well. Indeed, participants condemned actions resulting in negative outcomes and those driven by negative beliefs at a significantly higher rate than actions resulting in neutral outcomes and performed under neutral beliefs. Crucially, when conflicting information about beliefs and outcomes was presented (i.e., negative beliefs–neutral outcomes, and neutral beliefs–negative outcomes), participants' moral judgments were determined primarily by belief information: Judgments were harsher when rating attempted harms (i.e., negative beliefs–neural outcomes; e.g., trying but failing to poison another person) than when rating accidental harms (i.e., neutral beliefs–negative outcomes; e.g., accidentally poisoning another person; e.g., Young et al. (2007) and Cushman et al. (2006)).

Neuroimaging data from moral judgment tasks revealed activation in a specific neuronal network, including sub-regions of medial prefrontal cortex, precuneus, and right and left temporoparietal junction (TPJ), which has previously been associated with mental state reasoning (Perner et al., 2006; Ruby and Decety, 2003; Saxe and Kanwisher, 2003; Vogeley et al., 2001). Among these regions, the pattern of activation observed in the right TPJ (rTPJ) is particularly interesting. Young and Saxe (2008) observed that rTPJ activity correlated with moral judgments, responding selectively to the different types of moral actions, with higher response for the attempted harms (see also Young et al. (2007)). In a follow up study, Young and Saxe (2009b) showed that activation of the rTPJ also correlated with individual differences in making moral judgments when evaluating accidental harms: Participants with higher activation of the rTPJ were more likely to exculpate accidents, showing greater consideration for the agent's innocent intentions (see also Koster-Hale et al. (2013)).

Taken together these findings suggest that recruitment of the rTPJ for moral judgment reflects reliance on belief information: neutral beliefs allow to mitigate blame in case of accidents, whereas negative beliefs allow to assign blame in absence of an actual harm. Further evidence supporting the assumed role of the rTPJ in moral judgment comes from developmental and clinical studies. For instance, research has shown that young children tend to prioritize outcomes over intentions, and assign more blame for accidental than attempted harms (Cushman et al., 2013; Zelazo et al., 1996; Yuill and Perner, 1988; Shultz et al., 1986; Piaget, 1965/1932). As children grow up, they become more sensitive to the information about the intentions and their moral judgments change accordingly (Baird and Astington, 2004; Saxe et al., 2004), probably reflecting the maturation of (the processes underlying) ToM (Chandler et al., 2001; Killen et al., 2011). A developmental study investigating the neural networks involved in fairness related decisions has also shown increasing recruitment of rTPJ with increasing age, which was associated with age-related changes in understanding intentionality (Güroğlu et al., 2011). Similar to young children, high-functioning individuals with autism (ASD) – a disorder characterized by ToM impairments (Baron-Cohen, 1995; Baron-Cohen et al., 1985) – judge accidents more harshly on the basis of the bad outcome rather than the neutral intent (Moran et al., 2011; see also Koster-Hale et al. (2013)). These findings suggest that forgiving an agent for causing an accidental harm requires strong mental state representations and, thus, increased recruitment of the rTPJ (Koster-Hale et al., 2013; Young and Saxe, 2009b).

To sum up, the available evidence favors the idea that moral judgments depend on mental state reasoning and highlights the pivotal role of the rTPJ in mediating this process. However, most of what we know about the role of rTPJ in moral judgment comes from functional magnetic resonance (fMRI) studies that, given their correlational nature, provide no direct information about the functional and causal contribution of rTPJ in moral judgment (Poldrack, 2008; Page, 2006). Brain stimulation techniques, such as transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS), represent promising tools, as they allow one to directly modulate cortical excitability instead of relying on correlations between brain activation and behavior. By doing so, researchers can infer causal relationship between activity of a particular brain region and a specific cognitive function (Pascual-Leone et al., 2000; George and Aston-Jones, 2009; Dayan et al., 2013; Filmer et al., 2014).

To the best of our knowledge, only one study applied brain stimulation techniques to study the role of rTPJ for moral judgment. Specifically, Young et al. (2010b) showed that temporarily disrupting rTPJ activity with repetitive TMS (rTMS; see Walsh et al. (2003)) led participants to rely less on the agent's mental states, judging attempted harms (e.g., a failed murder attempt) as less morally forbidden and thus more morally permissible. Interestingly, the authors did not observe any effect of TMS on judgments of accidental harms. Therefore, disrupting rTPJ activity only altered judgments of moral actions associated with higher response of the rTPJ (see Young and Saxe (2008)). That being said, it would be of interest to assess possible behavioral changes in moral judgments that might result from increasing spontaneously present, instead of disrupting, rTPJ activation.

To this end, we employed tDCS (Paulus, 2011; Nitsche and

Paulus, 2011) to induce specific changes in the cortical excitability of the rTPJ and evaluate the behavioral effects of these changes on participants' moral judgments.

tDCS is a non-invasive brain stimulation technique that, by applying a weak current to the scalp via surface electrodes, polarity-dependently enhances (anodal tDCS) or reduces (cathodal tDCS) cortical excitability, and spontaneous cortical activity. The primary effects depend on sub-threshold membrane polarization, and prolonged stimulation induces neuroplastic alterations of cortical excitability driven by the glutamatergic system (Stagg and Nitsche, 2011). Several studies have provided converging evidence showing that tDCS is suited to alter cognitive functions (Kuo and Nitsche, 2012) and to ameliorate symptoms of several neurological and psychiatric disorders (Brunoni et al., 2012). Interestingly, anodal stimulation over the rTPJ was recently found to improve the ability to switch between representations of the self and other in both control of imitation and perspective-taking tasks (Santiesteban et al., 2012). In the current study, moral judgments were assessed by means of the well-established moral judgment task employed by Young and colleagues (e.g., Young et al., 2010b, 2007; Young and Saxe, 2009b): Participants were asked to rate, on a scale ranging from forbidden (1) to permissible (7), four classes of moral actions (Belief × Outcome design). In a single session, participants performed this task before (baseline assessment) and after (critical post-tDCS assessment) having received anodal (excitatory), cathodal (inhibitory), or sham (placebo) tDCS. Based on previous evidence, we predicted that the increased cortical excitability of the rTPJ induced by anodal tDCS would enhance the influence of belief information on moral judgments. This should affect judgments for moral actions in which belief and outcome conflict: either attempted harms or accidental harms, or both. By comparison, cathodal tDCS of the rTPJ was expected to reduce the influence of belief information and affect moral judgments accordingly.

2. Experiment 1

2.1. Materials and methods

2.1.1. Participants

Sixty Dutch students of the University of Amsterdam took part in the study. Participants were recruited via an on-line recruiting system and offered course credits or a financial reward (10 €) for participating in a study on the effects of brain stimulation on decision-making. Participants were screened individually via a phone interview by the same lab-assistant using the Mini International Neuropsychiatric Interview (M.I.N.I.). The M.I.N.I. is a short, structured, interview of about 15 min that screens for several psychiatric disorders and drug use, often used in clinical and pharmacological research (Sheehan et al., 1998; Colzato et al., 2008; Colzato et al., 2011). Participants were considered suitable to participate in this study if they fulfilled the following criteria: (i) age between 18 and 32 years; (ii) no history of neurological or psychiatric disorders; (iii) no history of substance abuse or dependence; (iv) no history of brain surgery, tumor or intracranial metal implantation; (v) no chronic or acute medications; (vi) no pregnancy; (vii) no susceptibility to seizures or migraine; (viii) no pacemaker or other implanted devices.

Once recruited, participants were randomly assigned to one of the three following experimental groups, each receiving only one type of stimulation: anodal ($N=20$; 8 male; mean age=22.2, SD=3.3), cathodal ($N=20$; 7 male; mean age=21.6, SD=3.3), or sham ($N=20$; 6 male; mean age=22.5, SD=3.3). Groups did not differ in terms of age, $F<1$, $p=.71$, or gender, $\chi^2=.44$, $p=.80$.

All participants were naïve to tDCS. Prior to the testing session, participants received a verbal and written explanation of the tDCS procedure and of the typical adverse effects (i.e., itching and tingling skin sensation, skin reddening, and headache). No information was provided about the different types of stimulation (active vs. sham) or about the hypotheses concerning the outcome of the experiment. All participants gave their written informed consent to participate to the study. The study conformed to the ethical standards of the declaration of Helsinki and the protocol was approved by the local Ethics Review Board of the University of Amsterdam.

2.1.2. Procedure

A single-blinded, sham-controlled design was used to assess the effect of tDCS – applied over the rTPJ in healthy young volunteers – on moral judgment. All participants took part in a single session and were tested individually. After having read and signed the informed consent, participants performed the first part of the moral judgment task, which served as a baseline measurement (pre-tDCS task). Next, active (either anodal or cathodal) or sham stimulation was applied for 20 min while at rest. After this phase, participants completed the second part of the moral judgment task (post-tDCS task). As the physiological effects of tDCS have been found to outlast the stimulation period by more than 60 min (Nitsche and Paulus, 2001), we can be sure that the effects of tDCS lasted throughout the entire critical task. After completion of the post-tDCS task, participants were properly debriefed and asked to fill in a tDCS adverse effects questionnaire requiring them to rate, on a five-point scale, how much they experienced: (1) headache, (2) neck pain, (3) nausea, (4) muscles contraction in face and/or neck, (5) stinging sensation under the electrodes, (6) burning sensation under the electrodes, (7) uncomfortable (generic) feelings, (6) other sensations and/or adverse effects. None of the participants reported major complains or discomfort during or after tDCS.

2.1.3. Transcranial direct current stimulation (tDCS)

Direct current was induced with two saline-soaked surface sponge electrodes (5 cm × 7 cm; 35 cm^2) and was delivered by means of a DC Brain Stimulator Plus (NeuroConn, Ilmenau, Germany). Electrodes were held in place by rubber bands and the stimulator was placed behind the participants. To stimulate the rTPJ, the target electrode (either the anode or the cathode, depending on the group assignment) was centered over CP6 (individually measured on each participant) – a location atop the rTPJ (cf. Santiesteban et al., 2012), according to the international 10–20 system for EEG electrode placement; the return electrode was placed over the left supraorbital area. The distance between the two electrodes was large enough to decrease the current shunted through the head and to increase the current density in depth (Miranda et al., 2006). For the active stimulation (either anodal or cathodal), a constant current of 1 mA (current density of 0.029 mA/cm^2) was delivered for 20 min with a linear fade-in/fade-out of 10 s, in conformity with safety criteria (Nitsche et al., 2003; Poreisz et al., 2007). For the sham stimulation, the position of the electrodes, current intensity and fade-in/fade-out were the same as in the active tDCS, but the stimulation was automatically turned off after 35 s, without the participants' awareness. Hence, participants felt the initial short-lasting skin sensation (i.e., itching and/or tingling) associated with tDCS without receiving any active current for the rest of the stimulation period. Stimulation for 35 seconds does not induce after-effects (Nitsche and Paulus, 2000) and is quite reliable in blinding participants to their stimulation condition (see Gandiga et al. (2006), Poreisz et al. (2007), Ambrus et al. (2012) and Palm et al. (2013)).

2.1.4. Moral judgment task

The task was adapted from Young and colleagues (Young et al.,

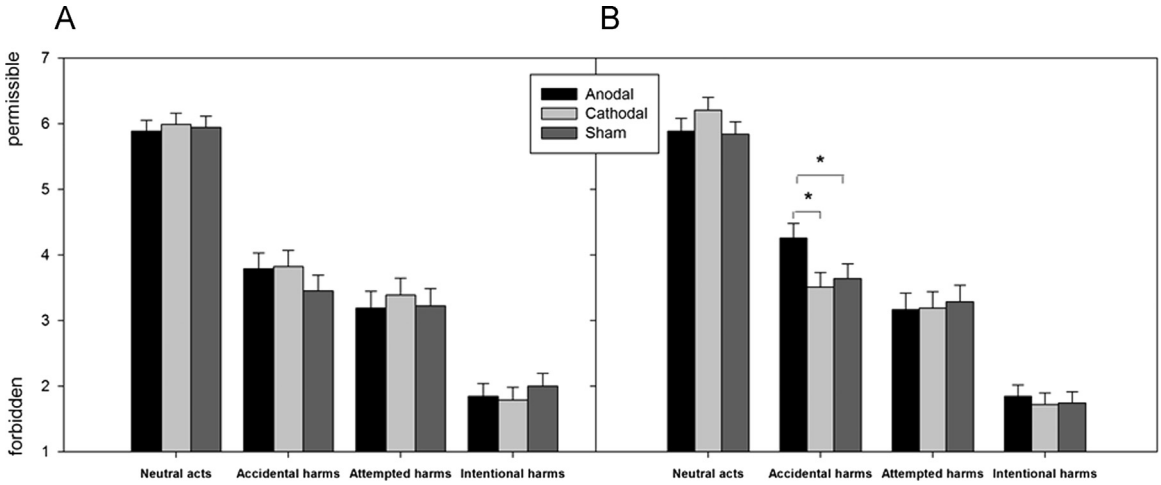

Fig. 2. Mean moral judgments on a seven-point scale (1=*morally forbidden*, 7=*morally permissible*), as a function of the type of moral action (resulting from the Belief by Outcome combination) and Stimulation type (anodal, cathodal, and sham) for the pre-tDCS (panel A) and the post-tDCS (panel B) tasks. Vertical capped lines atop bars indicate standard error of the mean.

2010b). To create the Dutch version of the task, the original scenarios were translated from English into Dutch and then back-translated from Dutch to English for comparison. Forty hypothetical moral scenarios were selected and randomly distributed among the pre- (16 scenarios) and post-tDCS (24 scenarios) tasks.[1] In both pre- and post-tDCS tasks, experimental stimuli were created by combining the information about the protagonist's belief (neutral vs. negative) and the outcome (neutral vs. negative). This resulted in a total number of 160 stories (64 for the pre-tDCS task and 96 for the post-tDCS task) distributed among 4 conditions (see Fig. 1): (i) Neutral act/No harm (neutral belief, neutral outcome); (ii) Accidental harm (neutral belief, negative outcome); (iii) Attempted harm (negative belief, neutral outcome); (iv) Intentional harm (negative belief, negative outcome). Each participant was confronted with 16 moral stories (4 per condition) in the pre-tDCS task and with 24 moral stories (6 per condition) in the post-tDCS task. Each participant read only one version of each scenario and across participants every scenario occurred in each of the four conditions. On average each story consisted of about 97.58 words, and the number of words was matched across conditions and tasks ($F_s < 1$, $p_s \geq .36$).

Stories were presented on a computer screen in a sequence of four segments, each presented for 8 s, describing in a fixed order: (1) background (i.e., information to set the scene), (2) foreshadow (i.e., information foreshadowing whether the action will result in a neutral or negative outcome), (3) the protagonist's belief, (4) the protagonist's action and its outcome. The background was identical across conditions. Following the presentation of each story, participants were asked to rate the moral permissibility of the action on a seven-point Likert scale (1=morally forbidden, 7=morally permissible), by pressing the corresponding button on a computer QWERTY keyboard. The time limit for responding was 5 s. Stories were interleaved by a blank screen presented for 2 s, followed by a 2 s screen warning participants that a new story was going to be presented. The pre-tDCS task lasted no more than 10 min; the post-tDCS task took no more than 14 min.

2.2. Results

Data from the baseline (pre-tDCS) and the post-tDCS tasks were analyzed separately by means of repeated measures analyses of variance (ANOVAs) treating either each participant (by-subjects analyses; F_1) or each scenario (by-item analyses; F_2) as a case.

In the by-subject analyses (F_1), data were submitted to ANOVAs with Belief (neutral vs. negative) and Outcome (neutral vs. negative) as within-subjects factors and Stimulation type (Anodal, Cathodal and Sham) as a between-subjects factor. In the by-item analyses (F_2), data were submitted to ANOVAs with Belief, Outcome and Stimulation type as within-items factors.

First, we analyzed the data from the baseline task to verify whether the three groups of participants showed comparable performance before tDCS was applied. ANOVAs performed on the data of the baseline task revealed significant main effects of Belief [$F_1(1,57)=186.56$, $p<.001$, $\eta_p^2=0.77$, $F_2(1,15)=86.67$, $p<.001$, $\eta_p^2=.85$], and Outcome [$F_1(1,57)=219.06$, $p<.001$, $\eta_p^2=.79$, $F_2(1,15)=205.70$, $p<.001$, $\eta_p^2=.93$]. As usually observed, actions performed with the belief of causing harms ($M=2.6$) were judged to be less morally permissible than actions performed with neutral beliefs ($M=4.8$). Also, actions resulting in harmful outcomes ($M=2.8$) were rated as less morally permissible than those resulting in neutral outcomes ($M=4.6$). Moreover, significant interactions involving Belief and Outcome were observed [$F_1(1,57)=22.44$, $p<.001$, $\eta_p^2=.28$, $F_2(1,15)=17.12$, $p<.005$, $\eta_p^2=.53$]. Newman-Keuls post-hoc analyses showed that intentional harms ($M=1.9$) were rated as more blameworthy than attempted harms ($M=3.3$, $p<.001$), accidental harms ($M=3.7$, $p<.001$), and neutral acts ($M=5.9$, $p<.001$)—all conditions differed significantly from each other ($p_s \leq .002$). Importantly, the main effect of Stimulation type was not significant, nor did it interact with any factor, $F_s \leq 1.6$, $p_s \geq .21$. Thus, participants' performance in the pre-tDCS task was comparable across the three experimental groups (Fig. 2A).

Next, we analyzed the data from the post-tDCS task in the same way. One scenario was excluded from the analyses because of a

[1] The original task employed by Young et al. (2010b) comprised of 48 scenarios. However, in our version of the task we made use of only 40 scenarios as we excluded scenarios that once translated in Dutch resulted in a larger number of words and whose translation sounded awkward, compared to the original ones.

The pre-tDCS task included the following scenarios: bike, bouncy ball, bridge, CPR, hunt, coffee, latex, Logan airport, malaria pond, motorboat, mushrooms, spinach, spring break, sushi, veterinarian, wet floor. The post-tDCS task included the following scenarios: alarm, asthma, cayo, chairlift, fraternity, ham, harness, jellyfish, iron, laptop, meatloaf, mother, parachutes, pool, porridge, rabies, river, safety cord, seatbelt, sesame, teenagers, tree house, vitamin, zoo. For full text of scenarios see Young et al. (2010b) – supporting information online at www.pnas.org/cgi/content/full/0914826107/DCSupplemental. The original scenarios' proper names were changed into Dutch proper names. Also, other names were modified to match the Dutch ones (e.g., Logan airport was translated into Schiphol airport).

mistake in the sentence describing the protagonist's belief. ANOVAs revealed significant main effects of Belief [$F_1(1,57)=270.69$, $p<.001$, $\eta_p^2=.83$, $F_2(1,22)=203.15$, $p<.001$, $\eta_p^2=.90$], and Outcome [$F_1(1,57)=235.51$, $p<.001$, $\eta_p^2=.81$, $F_2(1,22)=206.63$, $p<.001$, $\eta_p^2=.90$]. Actions performed with negative beliefs ($M=2.5$) and those resulting in negative outcomes ($M=2.8$) were judged as more blameworthy than actions performed with neutral belief ($M=4.9$) and those resulting in neutral outcomes ($M=4.6$). The interactions between Belief and Outcome were significant too [$F_1(1,57)=23.45$, $p<.001$, $\eta_p^2=.29$, $F_2(1,22)=13.97$, $p<.005$, $\eta_p^2=.39$]. Post-hoc analyses (Newman-Keuls) revealed that intentional harms ($M=1.8$) were judged to be less morally permissible than attempted harms ($M=3.2$, $p<.001$), accidental harms ($M=3.8$, $p<.001$), and neutral acts ($M=6.0$, $p<.001$). The latter three conditions differed significantly from each other ($p_s\leq.001$). A significant interaction involving Outcome and Stimulation type was observed in the by-item analysis [$F_2(2,44)=5.53$, $p<.01$, $\eta_p^2=.20$], indicating that participants relied less on the outcome after having received anodal than cathodal or sham stimulation. In particular, post-hoc analyses showed that the three groups of participants showed no difference when judging actions resulting in neutral outcomes (mean judgments were 4.5, 4.7, and 4.6, in anodal, cathodal, and sham conditions, respectively, $p_s\geq.35$). In contrast, they differed significantly when judging actions resulting in negative outcomes, with participants in the anodal condition rating these actions as more morally permissible ($M=3.1$) as compared to participants in the cathodal ($M=2.6$, $p=.003$) and sham ($M=2.7$, $p=.007$) conditions, whose judgments did not differ ($p=.52$). However, this interaction was not significant in the by-subjects analysis [$F_1(2,57)=2.28$, $p=.11$, $\eta_p^2=.07$].

Crucially, a significant three-way interaction involving Belief, Outcome, and Stimulation Type was observed [$F_1(2,57)=3.20$, $p<.05$, $\eta_p^2=.10$, $F_2(2,44)=3.20$, $p=.05$, $\eta_p^2=.10$], indicating selective differences between participant groups depending on the specific type of moral action. Specifically, post-hoc analyses showed that judgments were comparable across the three groups of participants when judging neutral acts (anodal: $M=5.9$, cathodal: $M=6.2$, sham: $M=5.8$, $p_s\geq.25$), attempted harms (anodal: $M=3.2$, cathodal: $M=3.2$, sham: $M=3.3$, $p_s\geq.72$), and intentional harms (anodal: $M=1.8$, cathodal: $M=1.7$, sham: $M=1.7$, $p_s\geq.71$). In contrast, significant differences across groups were observed when judging accidental harms: Participants who underwent anodal stimulation rated accidental harms ($M=4.3$) as less blameworthy than participants who received cathodal ($M=3.5$, $p=.019$) and sham ($M=3.6$, $p=.025$) stimulation, whose judgments were comparable ($p=.63$; Fig. 2B). No other significant sources of variance were found, $F_s\leq1.37$, $p_s\geq.26$.

2.3. Discussion

Consistent with previous findings (Koster-Hale et al., 2013; Young et al., 2010b, 2007; Young and Saxe, 2009b, 2008), in both pre- and post-tDCS tasks, we observed that to evaluate the moral permissibility of an agent's action, participants made use of both the information about the agent's beliefs and the information about the consequences of the agent's action. Indeed, they judged actions performed under the belief to cause harms and those causing negative consequences as more morally forbidden than actions performed under neutral beliefs and those causing no harm. Moreover, results indicated that participants integrated the two sources of information (belief and outcome) to determine the degree of blame to assign to the different moral actions. First, participants assigned substantial blame to accidental harms, thus showing little consideration of the agent's innocent beliefs in evaluating the moral permissibility of these actions. This was previously reckoned to reflect a particularly challenging aspect of these actions, which oppose salient information about a bad outcome against a neutral (non-salient) belief: Forgiving an accident implies to override the pre-potent emotional response to the salient bad outcome in favor of the less salient belief information (Young and Saxe, 2009b), which requires robust mental state representations (e.g., Cushman, 2008). Second, participants assigned more blame to attempted than accidental harms, and judged both moral actions as less blameworthy than intentional harms. This pattern reflects the fact that participants took into consideration not just the outcomes but also the agent's beliefs and weighted blame accordingly (Young et al., 2007).

More importantly, by applying tDCS over rTPJ we were able to alter selectively participants' moral judgments in the task performed after the stimulation period. Specifically, we observed that following excitatory (anodal) stimulation participants endorsed accidental harms at a significantly higher rate than participants who received inhibitory (cathodal) or sham (placebo) stimulation. This is consistent with previous fMRI data showing that increased activity in the rTPJ is associated with increased influence of belief information on moral judgments, and that individual differences in the the rTPJ activation predict the extent to which people make use of belief information to mitigate blame for accidents (Young and Saxe, 2009b; Koster-Hale et al., 2013). Conversely, the three groups of participants showed comparable judgments when rating neutral, attempted and intentional harms. Therefore, the results of the present study corroborate the hypothesis that the rTPJ plays a crucial role in mediating mental state reasoning during moral judgments. However, it is worth noting that, although placing the return electrode over supraorbital regions is quite common in tDCS studies, the use of this bipolar cortical electrode montage (cf. Nasseri et al., 2015), where anodal tDCS over rTPJ was combined with cathodal tDCS of the left supraorbital area, might have not been appropriate in our case. Indeed, supraorbital areas are located over frontal poles and orbitofrontal cortices and previous studies have found that moral judgments critically depend on the functioning of frontopolar and ventromedial frontal areas as well (Moll et al., 2011; Karim et al. 2010; Fumagalli et al., 2010; Young et al., 2010a; Moll et al. 2005), which are reckoned to play a pivotal role in emotional processing when making moral judgments (Greene et al. 2001). For instance, patients with focal demage to the ventromedial prefrontal cortex, who exhibit reduced emotional responsivity and lessened social emotions, have been found to produce an abnormally high rate of utilitarian responses when confronting with emotionally salient moral dilemmas (Koenigs et al., 2007). A similar outcome was observed in a recent study where female participants exhibited higher rate of utilitarian judgments following a tDCS-induced reduction of the ventromedial prefrontal cortex activity (Fumagalli et al., 2010).

Building on the aforementioned link between frontal pole areas, emotional processing and moral judgments, one may argue that the higher indulgence shown by participants in the anodal group when rating accidents was due to a reduced emotional response to harmful outcomes caused by cathodal tDCS of the prefrontal cortex. That being said, our experimental design does not allow one to ascertain whether the observed outcome was due to anodal stimulation of the rTPJ or to cathodal stimulation over the left pre-frontal region (lPFC). To shed light on this issue, we ran a control experiment (i.e., Experiment 2) to verify whether decreasing the cortical excitability (through cathodal tDCS) of the lPFC alone is sufficient to produce the behavioral effect we found (i.e., reduced blame for accidents observed in the anodal condition). More specifically, we tested a new sample of participants who underwent monopolar cathodal stimulation of the lPFC: A 35 cm^2 target (cathode) electrode was placed over the left

supraorbital area, and a 100 cm^2 return electrode was centered over the rTPJ (hereafter referred as "cathodal lPFC (or control) stimulation/group"). The use of differently sized electrodes and specifically, of a larger return electrode, has been shown to be an effective and easy way to allow a functional monopolar montage because of smaller current density, when current strength is kept constant (Nitsche et al., 2007). We then compared moral judgments of this new sample of participants with judgments of those participants who, in Experiment 1, received anodal tDCS over the rTPJ combined with cathodal tDCS over the left supraorbital area (hereafter referred as "anodal rTPJ stimulation/group"). To the extent to which the reduced blame assigned to accidents observed in the anodal group of Experiment 1 was due specifically to increased activation of the rTPJ, and not to decreased activity of the lPFC, participants undergoing cathodal lPFC stimulation should rate, in the post-tDCS assessment, accidental harms as more blameworthy than participants who received anodal rTPJ tDCS. By contrast, if the reduced blame assigned to accidents found for the anodal group of Experiment 1 was due to reduced cortical excitability of the lPFC, then comparable rating in the two groups should be observed.

3. Experiment 2

3.1. Materials and methods

3.1.1. Participants

A new sample of twenty Dutch students of the Leiden University (mean age=22.1 years, SD=2.8; 5 males) participated in the experiment for partial fulfillment of course credit or a financial reward. As in Experiment 1, all participants were prescreened by a phone interview using the M.I.N.I. (Sheehan et al., 1998) and were selected on the basis of the same inclusion criteria. Participants received verbal and written explanation of the tDCS procedure and of the typical adverse effects, but no information about the type of stimulation or the experimental hypotheses. All participants gave their written informed consent. The protocol was approved by the local ethical committee (Leiden University, Faculty of Social and Behavioral Sciences).

3.1.2. Apparatus, tasks, and procedure

The apparatus, tasks, and procedure were as in the Experiment 1 with the following exception. All participants underwent the same type of stimulation with the following montage: the target (cathode) electrode (5 cm × 7 cm; 35 cm^2 current density of 0.029 mA/cm^2) was placed over the left supraorbital area (lPFC), and the return electrode (10 cm × 10 cm; 100 cm^2 current density of 0.01 mA/cm^2) was centered over the rTPJ (i.e., over CP6).

3.2. Results and discussion

Table 1 shows mean moral judgments of the control group (i.e., cathodal lPFC stimulation) as a function of the to-be-rated moral action in the pre- and the post-tDCS tasks. Given that this second experiment served as a control to verify that anodal stimulation of the rTPJ, but not cathodal stimulation of the lPFC produced the behavioral effect observed in Experiment 1, moral judgments of this new group of participants were compared directly with those observed in the group of participants who, in Experiment 1, received anodal rTPJ tDCS. As in Experiment 1, data from the baseline (pre-tDCS) and the post-tDCS tasks were analyzed separately by means of repeated measures ANOVAs, treating either each participant (by-subjects analyses; F_1) or each scenario (by-item analyses; F_2) as a case. In the by-subject analysis (F_1), Belief (neutral vs. negative) and Outcome (neutral vs. negative) served as within-subjects factors, whereas Stimulation type (cathodal lPFC vs. anodal rTPJ) was entered as between-subjects factor. In the by-item analysis (F_2), Belief, Outcome and Stimulation type were treated as within-subjects factors.

Table 1
Mean moral judgments of the cathodal lPFC group (Experiment 2) as a function of the type of moral action (resulting from the Belief by Outcome combination), for the pre-tDCS and the post-tDCS tasks. Standard error of the mean are shown within parentheses.

Moral action	Pre-tDCS task	Post-tDCS task
Neutral acts (neutral belief, neutral outcome)	5.9 (.2)	5.9 (.2)
Accidental harms (neutral belief, negative outcome)	3.6 (.2)	3.6 (.2)
Attempted harms (negative belief, neutral outcome)	3.3 (.3)	3.0 (.3)
Intentional harms (negative belief, negative outcome)	1.8 (.2)	1.7 (.2)

ANOVAs performed on the data of the baseline task revealed that, before tDCS was applied, the two groups of participants showed comparable rating. Indeed, only three significant sources of variance were observed: main effects of Belief [$F_1(1,38)=160.289$, $p<.001$, $\eta_p^2=.81$, $F_2(1,15)=81.668$, $p<.001$, $\eta_p^2=.84$], and Outcome [$F_1(1,38)=194.169$, $p<.001$, $\eta_p^2=.84$, $F_2(1,15)=256.370$, $p<.001$, $\eta_p^2=.94$], and significant interactions involving the two factors [$F_1(1,38)=10.565$, $p<.005$, $\eta_p^2=.23$, $F_2(1,15)=8.754$, $p<.01$, $\eta_p^2=.37$]. Actions performed with negative beliefs ($M=2.5$) and those resulting in negative outcomes ($M=2.8$) were judged as more blameworthy than actions performed with neutral belief ($M=4.8$) and those resulting in neutral outcomes ($M=4.6$). Newman–Keuls post-hoc analyses performed to disentangle the interaction showed that intentional harms ($M=1.8$) were judged to be less morally permissible than attempted harms ($M=3.2$, $p<.001$), accidental harms ($M=3.7$, $p<.001$), and neutral acts ($M=5.9$, $p<.001$), and that the latter three conditions differed significantly from each other too ($p_s\leq.05$). No other significant sources of variance were found [$F_s<1$, $p_s\geq.62$].

ANOVAs performed on the data of the post-tDCS task showed significant main effects of Belief [$F_1(1,38)=180.410$, $p<.001$, $\eta_p^2=.83$, $F_2(1,22)=186.413$, $p<.001$, $\eta_p^2=.89$], and Outcome [$F_1(1,38)=148.545$, $p<.001$, $\eta_p^2=.80$, $F_2(1,22)=171.768$, $p<.001$, $\eta_p^2=.89$]. Mirroring the data of the baseline task, participants condemned at a significantly higher rate actions performed with the belief of causing harms ($M=2.4$) and those resulting in harmful outcomes ($M=2.8$) than actions performed with neutral belief ($M=4.9$) and those resulting in neutral outcomes ($M=4.5$). A significant main effect of Stimulation type was found in the by-item analysis [$F_2(1,22)=7.901$, $p<.05$, $\eta_p^2=.26$], but not in the by-subjects analysis [$F_1(1,38)=2.61$, $p=0.11$, $\eta_p^2=.06$]: participants who received cathodal lPFC stimulation were less indulgent than those who received anodal rTPJ stimulation ($M=3.60$ vs. $M=3.80$). Again, significant Belief × Outcome interactions were found [$F_1(1,38)=14.11$, $p<.001$, $\eta_p^2=.27$, $F_2(1,22)=9.729$, $p<.005$, $\eta_p^2=.31$]: Newman–Keuls post-hoc analyses showed that intentional harms ($M=1.8$) were rated as more blameworthy than attempted harms ($M=3.1$, $p<.001$), accidental harms ($M=3.9$, $p<.001$), and neutral acts ($M=5.9$, $p<.001$) – all conditions differed significantly from each other ($p_s\leq.001$). The interaction between Outcome and Stimulation type tended to be significant in the by-item analysis [$F_2(1,22)=4.215$, $p=.05$, $\eta_p^2=.16$], but not in the by-subjects analysis [$F_1(1,38)=1.78$, $p=.19$, $\eta_p^2=.04$]: the two groups did not differ when judging actions resulting in neutral

outcomes (mean judgments were 4.5 and 4.5 in the cathodal lPFC and anodal rTPJ conditions, respectively, $p=.64$), whereas when judging actions resulting in harmful outcomes, participants in the cathodal lPFC condition were less indulgent than participants in the anodal rTPJ condition ($M=2.6$ vs. $M=3.06$, $p<0.001$). More importantly, the three-way interactions involving Belief, Outcome, and Stimulation Type were significant too [$F_1(1,38)= 4.32$, $p<.05$, $\eta_p^2=.10$, $F_2(1,22)=6.64$, $p<.05$, $\eta_p^2=.23$]. Post-hoc analyses showed that judgments were comparable across the two groups when judging neutral acts (anodal rTPJ: $M=5.9$, cathodal lPFC: $M=5.9$, $p=.82$), attempted harms (anodal rTPJ: $M=3.2$, cathodal lPFC: $M=3.0$, $p=.44$), and intentional harms (anodal rTPJ: $M=1.8$, cathodal lPFC: $M=1.7$, $p=.47$). Notably, a significant difference was observed for accidental harms: Participants who underwent anodal rTPJ tDCS rated accidental harms as less blameworthy than participants who received cathodal lPFC stimulation ($M=4.3$ vs. $M=3.6$, $p<.01$). No other significant sources of variance were found, $F_s \leq 1.78$, $p_s \geq .19$.

The results of this experiment rule out the possibility that the higher indulgence in rating accidental harms shown by participants in the anodal group was due, in fact, to cathodal stimulation of the lPFC. As such, these results provide straightforward evidence that the observed outcome was mediated specifically by tDCS-induced changes of rTPJ activity.

4. Conclusions

Previous studies have indicated that an agent's beliefs about whether his/her actions will cause harm dominate moral judgments, revealing the key role of mental state reasoning – a set of processes linked to rTPJ activity (Perner et al., 2006; Ruby and Decety, 2003; Saxe and Kanwisher, 2003; Vogeley et al., 2001) – for moral judgment (for reviews, see Young (2013) and Young and Waytz (2013)). The present study sought to extend these findings by providing direct evidence for the critical involvement of the rTPJ in moral judgment. To this end, we used tDCS (Paulus, 2011; Nitsche and Paulus, 2011) to alter the cortical excitability of the rTPJ and we examined the behavioral after-effects of the induced cortical changes on participants' moral judgments. Moral judgments were assessed before and after tDCS was applied by means of a well-established task that confronts participants with different moral actions in a 2 (Belief: neutral vs. negative) by 2 (Outcome: neutral vs. negative) design and requires them to judge the moral permissibility of these actions (cf. Young et al., 2010b). Besides replicating the main findings observed in previous studies using the same moral judgment task (Koster-Hale et al., 2013; Young et al., 2010b, 2007; Young and Saxe, 2009b, 2008), here we provided direct evidence favoring the hypothesis that the impact of belief information during third party moral judgments critically depends on rTPJ activity. Indeed, we observed that increasing cortical excitability of the rTPJ via tDCS enhanced the role of belief information, leading participants to mitigate blame when rating the moral permissibility of accidental harms.

A seemingly oddity in our results is the absence of any effect of tDCS on judgments of attempted harms. In Young et al. (2010b), temporarily disrupting rTPJ activity with rTMS reduced participants' consideration of the belief information. Indeed, compared to rTMS targeting a control area, rTPJ rTMS led participants to assign less blame to attempted harms, thus reflecting a sort of “no harm, no fault mentality” (Young et al., 2010b). Building on this finding, it was reasonable to expect anodal rTPJ tDCS, compared to cathodal and sham tDCS, to produce the complementary pattern: higher blame for attempted harms due to increased consideration of belief information. To account for the null effect, it might be useful to consider a critical difference that exists between accidental and attempted harms. As already mentioned, accidental and attempted harms differ in the degree of recruitment of the rTPJ, which is higher for attempted than accidental harms (Koster-Hale et al., 2013; Young et al., 2010b, 2007; Young and Saxe, 2009b). This is probably because belief information has a different weight in the two types of actions: it is salient for attempted harms that confront a negative belief with a neutral outcome, whereas it is less salient in the case of accidental harms where a neutral belief is confronted with a particularly salient bad outcome. To mitigate blame, accidental harms require a stronger mental state representation than attempted harms such to override the salient negative outcome and to prioritize the less salient belief information. Based on these premises, it makes sense to expect that increasing (vs. decreasing) the role of belief information may affect differentially moral judgments for accidental and attempted harms. Specifically, increasing the contribution of mental state representations is more likely to affect judgments of those moral actions in which beliefs matter less and the activation of the rTPJ is less pronounced (i.e., accidental harms). In contrast, decreasing the role of mental state representation should affect mainly judgments of moral actions in which this information is weighted more and the activation of the rTPJ is more pronounced (i.e., attempted harms). Consistently, we observed that increased rTPJ activity induced by anodal stimulation affected judgments of accidental harms, but not judgments of attempted harms. In this view, our results fit perfectly with those reported by Young et al. (2010b) who observed the opposite pattern: disruption of rTPJ activity affected moral judgments for attempted harms, but not for accidental harms. Contrary to Young et al. (2010b), we did not observe any modulation following cathodal (inhibitory) tDCS. This might be due to aspects related to our tDCS protocol, such as the intensity of the current we applied and the use of offline stimulation (see Nozari et al. (2014), and Pirulli et al. (2014) for considerations about the importance of factors like intensity, duration, and timing of application for cathodal tDCS). Future studies might consider to extend our findings by varying these parameters.

A final consideration pertains to the electrode montage employed in Experiment 2, which was aimed to make rTPJ stimulation functionally inert so as to verify whether the results observed in the anodal condition of Experiment 1 could be ascribed to a cortical excitability reduction of the lPFC rather than to a tDCS-induced increase in rTPJ activity. As previously mentioned, increasing the size of the return electrode in relation to the target electrode has been shown to be an effective way to make the stimulation of the non-target area functionally inert, because of the lower current density beneath the return electrode (Nitsche et al., 2007; see Knoch et al. (2008), Fregni et al. (2008), and Klein et al. (2013), for examples of behavioral studies using this approach). Following this logic, in Experiment 2, while the size of the lPFC electrode was kept identical to the one used in Experiment 1 (i.e., 35 cm^2), the size of the rTPJ electrode was increased to cover an area of 100 cm^2, which results in a maximal current density of 0.01 mA/cm^2 – a value that is unlikely to induce any physiological effect (Nitsche & Paulus, 2000; Nitsche et al., 2007). Importantly, given that the location of the two electrodes was the same as in the anodal condition of Experiment 1, this control condition made it possible to assess the specific influence of the lPFC without modifying current flow direction, which is critical for neuronal effects to be found and/or to be compared with each other (Kabakov et al., 2012). Consistent with our expectations, cathodal lPFC stimulation alone was not sufficient to affect moral judgment ratings – a finding that allows us to conclude with a certain confidence that the pattern of results observed in Experiment 1 cannot be ascribed to the position of the return electrode (i.e., to a cortical excitability reduction in the lPFC), but was specifically due

to increased cortical excitability in the rTPJ. Moreover, the failure to observe any tDCS-induced effects in Experiment 2 suggests that the use of a larger electrode applied over rTPJ was in fact effective in making this area functionally inert. This is not to deny that, however, it would be advisable for future studies to run a computer simulation of the tDCS montage we employed to assess more precisely the current density beneath each electrode and the resulting field intensity, which would also allow one to predict whether excitability changes are likely to occur in other close and/or distant cortical regions (cf. Klein et al., 2013).

The current study has some limitations that warrant discussion. First, we did not assess explicitly participants' blinding by asking them if they could guess the stimulation received. However, previous studies have shown that with the chosen parameters of stimulation blinding is quite reliable (Ambrus et al., 2012; Palm et al., 2013). Second, the use of relatively large electrodes, as the ones employed in the present study, cannot guarantee that tDCS was focalized only under the electrodes (Miranda et al., 2006; Wagner et al., 2007). Thus, follow-up studies using smaller sized electrodes to increase focality would be essential as further confirmation of these findings. Third, and related to the previous point, to gain a better understanding of how tDCS, as it was applied in the present study, influences rTPJ activity and how tDCS-induced changes in cortical excitability affect moral judgments, it would be interesting to combine tDCS with fMRI. This would enable a more detailed interpretation of our results while reducing uncertainty about the neural substrate that was in fact modulated by our stimulation protocol (Shafi et al., 2012).

To sum up, our findings show that anodal tDCS over the rTPJ can alter selectivity moral judgments for accidental harms, leading participants to rely more on the agent's (innocent) mental states when judging the moral permissibility of these actions. It is worth noting that anodal tDCS did not alter judgments of accidental harms to such an extent to make participants extremely indulgent. Abnormally lenient judgments of accidents are typically shown by psychopaths, probably reflecting the absence of an emotional response to the harmful outcome (Young et al., 2012), and by individuals with alexithymia, probably because of their emphatic deficits (Patil and Silani, 2014).

What implications might these findings have for the field of cognitive neuroscience? First of all, our findings support the hypothesis that mental state reasoning is critical for moral judgments and provide direct evidence that mental state reasoning during moral judgments depends critically on the neural activity in the rTPJ. Second, our results provide additional evidence supporting the efficacy of the tDCS in modulating cognitive and social functions that are assumed to rely on the targeted area. Third, the finding that anodal stimulation over the rTPJ can enhance mental state reasoning suggests that tDCS may represent a promising and effective tool to mitigate mental state impairments that typically characterize ASD individuals.

Acknowledgments

This work was supported by a research grant from the Netherlands Organization for Scientific Research (NWO; www.nwo.nl) awarded to LSC (Vidi Grant: #452-12-001). The NWO had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

References

Ambrus, G.G., Al-Moyed, H., Chaieb, L., Sarp, L., Antal, A., Paulus, W., 2012. The fade-in – short stimulation – fade out approach to sham tDCS–reliable at 1 mA for naive and experienced subjects, but not investigators. Brain Stimul. 5, 499–504.
Baird, J.A., Astington, J.W., 2004. The role of mental state understanding in the development of moral cognition and moral action. New Dir. Child Adolesc. Dev. 103, 37–49.
Baron-Cohen, S., 1995. Mindblindness: An Essay on Autism and Theory of Mind. MIT Press, Cambridge, MA.
Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21, 37–46.
Baron-Cohen, S., Lombardo, M., Tager-Flusberg, H., Cohen, D., 2013. Understanding Other Minds: Perspectives From Developmental Social Neuroscience. Oxford University Press, New York, NY.
Brunoni, A.R., Nitsche, M.A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., Fregni, F., 2012. Clinical research with transcranial direct current stimulation (tDCS): challenges and future directions. Brain Stimul. 5, 175–195.
Chandler, M.J., Sokol, B.W., Hallett, D., 2001. Moral responsibility and the interpretive turn: Childrens changing conceptions of truth and rightness. In: Malle, B.F., Moses, L.J., Baldwin, D. (Eds.), Intentions and Intentionality: Foundations of Social Cognition. MIT Press, Cambridge, MA, pp. 345–365.
Colzato, L.S., Kool, W., Hommel, B., 2008. Stress modulation of visuomotor binding. Neuropsychologia 46, 1542–1548.
Colzato, L.S., Ruiz, M.J., van den Wildenberg, W.P., Hommel, B., 2011. Khat use is associated with impaired working memory and cognitive flexibility. PloS One 6, e20602.
Cushman, F., 2008. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. Cognition 108, 353–380.
Cushman, F., Sheketoff, R., Wharton, S., Carey, S., 2013. The development of intent-based moral judgment. Cognition 127, 6–21.
Cushman, F., Young, L., Hauser, M., 2006. The role of conscious reasoning and intuition in moral judgment testing three principles of harm. Psychol. Sci. 17, 1082–1089.
Dayan, E., Censor, N., Buch, E.R., Sandrini, M., Cohen, L.G., 2013. Noninvasive brain stimulation: from physiology to network dynamics and back. Nat. Neurosci. 16, 838–844.
Filmer, H.L., Dux, P.E., Mattingley, J.B., 2014. Applications of transcranial direct current stimulation for understanding brain function. Trends Neurosci. 37, 742–753.
Fregni, F., Liguori, P., Fecteau, S., Nitsche, M.A., Pascual-Leone, A., Boggio, P.S., 2008. Cortical stimulation of the prefrontal cortex with transcranial direct current stimulation reduces cue-provoked smoking craving: a randomized, sham-controlled study. J. Clin. Psychiatry 69, 32–40.
Fumagalli, M., Priori, A., 2012. Functional and clinical neuroanatomy of morality. Brain 135, 2006–2021.
Fumagalli, M., Vergari, M., Pasqualetti, P., Marceglia, S., Mameli, F., Ferrucci, R., Priori, A., 2010. Brain switches utilitarian behavior: does gender make the difference. PLoS One 5, e8865.
Gandiga, P.C., Hummel, F.C., Cohen, L.G., 2006. Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. Clin. Neurophysiol. 117, 845–850.
George, M.S., Aston-Jones, G., 2009. Noninvasive techniques for probing neurocircuitry and treating illness: vagus nerve stimulation (VNS), transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS). Neuropsychopharmacology 35, 301–316.
Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D., 2004. The neural bases of cognitive conflict and control in moral judgment. Neuron 44, 389–400.
Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293, 2105–2108.
Güroğlu, B., van den Bos, W., van Dijk, E., Rombouts, S.A., Crone, E.A., 2011. Dissociable brain networks involved in development of fairness considerations: understanding intentionality behind unfairness. Neuroimage 57, 634–641.
Kabakov, A.Y., Muller, P.A., Pascual-Leone, A., Jensen, F.E., Rotenberg, A., 2012. Contribution of axonal orientation to pathway-dependent modulation of excitatory transmission by direct current stimulation in isolated rat hippocampus. J. Neurophysiol. 107, 1881–1889.
Karim, A.A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., Birbaumer, N., 2010. The truth about lying: inhibition of the anterior prefrontal cortex improves deceptive behavior. Cereb. Cortex 20, 205–213.
Killen, M., Mulvey, K.L., Richardson, C., Jampol, N., Woodward, A., 2011. The accidental transgressor: morally-relevant theory of mind. Cognition 119, 197–215.
Klein, E., Mann, A., Huber, S., Bloechle, J., Willmes, K., Karim, A.A., Moeller, K., 2013. Bilateral bi-cephalic tDCS with two active electrodes of the same polarity modulates bilateral cognitive processes differentially. PloS One 8, e71607.
Kliemann, D., Young, L., Scholz, J., Saxe, R., 2008. The influence of prior record on moral judgment. Neuropsychologia 46, 2949–2957.
Knoch, D., Nitsche, M.A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., Fehr, E., 2008. Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. Cereb. Cortex 18, 1987–1990.
Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A., 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446, 908–911.
Koster-Hale, J., Saxe, R., Dungan, J., Young, L., 2013. Decoding moral judgments from neural representations of intentions. Proc. Natl. Acad. Sci. USA 110, 5648–5653.
Kuo, M.F., Nitsche, M.A., 2012. Effects of transcranial electrical stimulation on cognition. Clin. EEG Neurosci. 43, 192–199.
Miranda, P.C., Lomarev, M., Hallett, M., 2006. Modeling the current distribution

during transcranial direct current stimulation. Clin. Neurophysiol. 117, 1623–1629.
Moll, J., Zahn, R., de Oliveira-Souza, R., Bramati, I.E., Krueger, F., Tura, B., Grafman, J., 2011. Impairment of prosocial sentiments is associated with frontopolar and septal damage in frontotemporal dementia. Neuroimage 54, 1735–1742.
Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., Grafman, J., 2005. The neural basis of human moral cognition. Nat. Rev. Neurosci. 6, 799–809.
Moran, J.M., Young, L.L., Saxe, R., Lee, S.M., O'Young, D., Mavros, P.L., Gabrieli, J.D., 2011. Impaired theory of mind for moral judgment in high-functioning autism. Proc. Natl. Acad. Sci. USA 108, 2688–2692.
Nadelhoffer, T., 2006. Bad acts, blameworthy agents, and intentional actions: some problems for juror impartiality. Philos. Explor. 9, 203–219.
Nasseri, P., Nitsche, M.A., Ekhtiari, H., 2015. A framework for categorizing electrode montages in Transcranial Direct Current Stimulation. Front. Hum. Neurosci. 9, 54.
Nichols, S., 2002. Norms with feeling: towards a psychological account of moral judgment. Cognition 84, 221–236.
Nitsche, M.A., Doemkes, S., Karakoese, T., Antal, A., Liebetanz, D., et al., 2007. Shaping the effects of transcranial direct current stimulation of the human motor cortex. J. Neurophysiol. 97, 3109–3117.
Nitsche, M.A., Liebetanz, D., Antal, A., Lang, N., Tergau, F., Paulus, W., 2003. Modulation of cortical excitability by weak direct current stimulation—technical, safety and functional aspects. Suppl. Clin. Neurophysiol. 56, 255.
Nitsche, M.A., Paulus, W., 2000. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J. Physiol. 527, 633–639.
Nitsche, M.A., Paulus, W., 2001. Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology 57, 1899–1901.
Nitsche, M.A., Paulus, W., 2011. Transcranial direct current stimulation–update 2011. Restor. Neurol. Neurosci. 29, 463–492.
Nozari, N., Woodard, K., Thompson-Schill, S.L., 2014. Consequences of cathodal stimulation for behavior: when does it help and when does it hurt performance? PLoS One 9, e84338.
Page, M., 2006. What can't functional neuroimaging tell the cognitive psychologist? Cortex 42, 428–443.
Palm, U., Reisinger, E., Keeser, D., Kuo, M.F., Pogarell, O., Leicht, G., et al., 2013. Evaluation of sham transcranial direct current stimulation for randomized, placebo-controlled clinical trials. Brain Stimul. 6, 690–695.
Pascual-Leone, A., Walsh, V., Rothwell, J., 2000. Transcranial magnetic stimulation in cognitive neuroscience–virtual lesion, chronometry, and functional connectivity. Curr. Opin. Neurobiol. 10, 232–237.
Patil, I., Silani, G., 2014. Alexithymia increases moral acceptability of accidental harms. J. Cogn. Psychol. 26, 597–614.
Paulus, W., 2011. Transcranial electrical stimulation (tES-tDCS; tRNS, tACS) methods. Neuropsychol. Rehabil. 21, 602–617.
Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Thinking of mental and other representations: the roles of left and right temporo-parietal junction. Soc. Neurosci. 1, 245–258.
Piaget, J., 1965/1932. The Moral Judgment of the Child. Harcourt, New York, NY.
Pirulli, C., Fertonani, A., Miniussi, C., 2014. Is neural hyperpolarization by cathodal stimulation always detrimental at the behavioral level? Front. Behav. Neurosci. 8, 226.
Poldrack, R.A., 2008. The role of fMRI in cognitive neuroscience: where do we stand? Curr. Opin. Neurobiol. 18, 223–227.
Poreisz, C., Boros, K., Antal, A., Paulus, W., 2007. Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. Brain Res. Bull. 72, 208–214.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 1, 515–526.
Ruby, P., Decety, J., 2003. What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. Eur. J. Neurosci. 17, 2475–2480.
Santiesteban, I., Banissy, M.J., Catmur, C., Bird, G., 2012. Enhancing social ability by stimulating right temporoparietal junction. Curr. Biol. 22, 2274–2277.
Saxe, R., Carey, S., Kanwisher, N., 2004. Understanding other minds: linking developmental psychology and functional neuroimaging. Annu. Rev. Psychol. 55, 87–124.
Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". NeuroImage 19, 1835–1842.
Shafi, M.M., Westover, M.B., Fox, M.D., Pascual-Leone, A., 2012. Exploration and modulation of brain network interactions with noninvasive brain stimulation in combination with neuroimaging. Eu. J. Neurosci. 35, 805–825.
Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., 1998. The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J. Clin. Psychiatry 59, 22–23.
Shultz, T.R., Wright, K., Schleifer, M., 1986. Assignment of moral responsibility and punishment. Child Dev. 57, 177–184.
Stagg, C.J., Nitsche, M.A., 2011. Physiological basis of transcranial direct current stimulation. Neuroscientist 17, 37–53.
Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Zilles, K., 2001. Mind reading: neural mechanisms of theory of mind and self-perspective. NeuroImage 14, 170–181.
Wagner, T., Fregni, F., Fecteau, S., Grodzinsky, A., Zahn, M., Pascual-Leone, A., 2007. Transcranial direct current stimulation: a computer-based human model study. Neuroimage 35, 1113–1124.
Walsh, V., Pascual-Leone, A., Kosslyn, S.M., 2003. Transcranial Magnetic Stimulation: a Neurochronometrics of mind. MIT Press, Cambridge, MA.
Woolfolk, R.L., Doris, J.M., Darley, J.M., 2006. Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. Cognition 100, 283–301.
Young, L., 2013. Moral thinking. In: Reisberg, D. (Ed.), The Oxford Handbook of Cognitive Psychology. Oxford University Press, New York, NY.
Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M., Damasio, A., 2010a. Damage to ventromedial prefrontal cortex impairs judgment of harmful intent. Neuron 65, 845–851.
Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., 2010b. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc. Natl. Acad. Sci. USA 107, 6753–6758.
Young, L., Cushman, F., Adolphs, R., Tranel, D., Hauser, M.D., 2006. Does emotion mediate the relationship between an action's moral status and its intentional status? Neuropsychological evidence. J. Cognit. Cult. 6, 265–278.
Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. Proc. Natl. Acad. Sci. USA 104, 8235–8240.
Young, L., Dungan, J., 2012. Where in the brain is morality? Everywhere and maybe nowhere. Soc. Neurosci. 7, 1–10.
Young, L., Koenigs, M., Kruepke, M., Newman, J.P., 2012. Psychopathy increases perceived moral permissibility of accidents. J. Abnorm. Psychol. 121, 659.
Young, L., Nichols, S., Saxe, R., 2010c. Investigating the neural and cognitive basis of moral luck: it's not what you do but what you know. Rev. Philos. Psychol. 1, 333–349.
Young, L., Saxe, R., 2008. The neural basis of belief encoding and integration in moral judgment. NeuroImage 40, 1912–1920.
Young, L., Saxe, R., 2009a. An fMRI investigation of spontaneous mental state inference for moral judgment. J. Cogn. Neurosci. 21, 1396–1405.
Young, L., Saxe, R., 2009b. Innocent intentions: a correlation between forgiveness for accidental harm and neural activity. Neuropsychologia 47, 2065–2072.
Young, L., Waytz, A., 2013. Morality. In: Baron-Cohen, S., Tager-Flusberg, H., Lombardo, M. (Eds.), Understanding Other Minds. Oxford University Press, New York, NY.
Yuill, N., Perner, J., 1988. Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. Dev. Psychol. 24, 358–365.
Zelazo, P.D., Helwig, C.C., Lau, A., 1996. Intention, act, and outcome in behavioral prediction and moral judgment. Child Dev. 67, 2478–2492.