Reducing prejudice through brain stimulation

Sellaro, R.; Derks, B.; Nitsche, M.A.; Hommel, B.; van den Wildenberg, W.P.M.; van Dam, K.; Colzato, L.S.

DOI
10.1016/j.brs.2015.04.003

Publication date
2015

Document Version
Final published version

Published in
Brain Stimulation

Brain Stimulation 8 (2015) 891–897

Contents lists available at ScienceDirect

Brain Stimulation

journal homepage: www.brainstimjrnl.com

Original Articles

Reducing Prejudice Through Brain Stimulation

Roberta Sellaro [a,*], Belle Derks [b], Michael A. Nitsche [c], Bernhard Hommel [a], Wery P.M. van den Wildenberg [d,e], Kristina van Dam [a], Lorenza S. Colzato [a]

[a] Cognitive Psychology Unit & Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[b] Social and Organizational Psychology, Utrecht University, Utrecht, The Netherlands
[c] Department of Clinical Neurophysiology, Georg-August University Göttingen, Germany
[d] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[e] Amsterdam Brain & Cognition ABC., University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 24 November 2014
Received in revised form
13 April 2015
Accepted 17 April 2015
Available online 16 May 2015

Keywords:
Medial prefrontal cortex
Transcranial direct current stimulation
Implicit bias
Stereotype
Implicit Association Test
Cognitive control

ABSTRACT

Background: Social categorization and group identification are essential ingredients for maintaining a positive self-image that often lead to negative, implicit stereotypes toward members of an out-group. The medial prefrontal cortex (mPFC) may be a critical component in counteracting stereotypes activation.
Objective: Here, we assessed the causal role of the mPFC in these processes by non-invasive brain stimulation via transcranial direct current stimulation (tDCS).
Method: Participants ($n = 60$) were randomly and equally assigned to receive anodal, cathodal, or sham stimulation over the mPFC while performing an Implicit Association Test (IAT): They were instructed to categorize in-group and out-group names and positive and negative attributes.
Results: Anodal excitability-enhancing stimulation decreased implicit biased attitudes toward out-group members compared to excitability-diminishing cathodal and sham stimulation.
Conclusions: These results provide evidence for a critical role of the mPFC in counteracting stereotypes activation. Furthermore, our results are consistent with previous findings showing that increasing cognitive control may overcome negative bias toward members of social out-groups.

Introduction

The desire to affiliate and the ability to discriminate "us" from "them" are important ingredients for building and maintaining a positive self-image, but are also associated with social discrimination, stereotypes, prejudices and intergroup conflicts [1]. According to Allport [2] stereotyping and prejudice are a normal product of an automatic categorization process – one of the most adaptive and fundamental human cognitive functions. The ability to categorize is an efficient cognitive heuristic that allows to simplify the complexity of the physical and social world [3]. As such, the process of categorizing individuals in different groups is not different from the process of categorizing other events or objects on the basis of their underlying properties [4]. Furthermore, social categorization is essential for defining and maintaining one's social identity (i.e., the self-knowledge and self-esteem that derives from being member of a given group) – a fundamental part of the self-concept [5].

As individuals have an innate tendency to maintain a positive image of themselves and of the group they belong to (and identify with), they tend to maximize the distinction between in-group and out-group, often implying a positive evaluation of the in-group at the expense of the out-group [6,7]. However, the utility that derives from any kind of categorization has a cost: it can lead to irrational, over-generalized stereotypes that may have dramatic consequences when applied to individuals. Crucially, several studies have shown that, although people explicitly report unbiased attitudes toward members of an out-group, they often demonstrate negative implicit attitudes that affect choices, judgments, and nonverbal behaviors toward them [8–11]. As explicit attitudes are regulated more strongly by societal norms and, hence, by motivational factors (e.g., the desire to be politically correct; [12]), implicit attitudes are informative as they are thought to capture a less controllable bias against social out-groups [11].

Given the important role that implicit attitudes play in mediating social discrimination processes, several studies have been devoted to assess the cognitive and neural correlates of implicit

* Corresponding author. 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.brs.2015.04.003
1935-861X/

attitudes [1,13,14]. An extensive line of research has focused on the mechanisms supporting self-regulatory cognitive control processes that may allow overriding the activation of social stereotypes. Social stereotyping concepts suggest that the ability to refrain from biased behaviors resembles the selection of an appropriate response in a context in which another well-learned response is concurrently activated [1,13,15–19]. This process, usually referred to as response conflict, requires effective self-regulation and the implementation of top-down control to select the appropriate response/behavior [20,21].

Consistent with this hypothesis, implicit biased attitudes typically increase after cognitive demanding tasks, thus suggesting that people are less efficient in counteracting the behavioral effects that are driven by activated stereotypes when control-related resources are depleted [22–24]. Similarly, alcohol use and aging – two factors known to be associated with impairments in self-regulation and cognitive control [25,26] – were found to increase stereotype and prejudice [27,28]. Conversely, factors that increase self-regulation and cognitive control by enhancing the motivation to appear unbiased, such as morality [29] and guilty feelings about being prejudiced [30], have been found to attenuate implicit biased attitudes.

Neuroimaging and electrophysiological studies have revealed a consistent pattern of results: the same neural structures typically engaged during cognitive control tasks to implement goal-directed behavior, such as the anterior cingulate cortex (ACC) and the dorsolateral prefrontal cortex (dlPFC; cf. conflict monitoring theory [20,31]), are also recruited to overcome overbearing responses reflecting the automatic activation of implicit biased attitudes [27–29,32–35]. More important for the purpose of the current study, recent findings suggest that the medial prefrontal cortex (mPFC) – an area typically linked to socio-cognitive processes [36,37] – may be implicated in regulating and controlling stereotypes as well [33,36]. For instance, Amodio et al. [33] observed that activity in the mPFC was uniquely related to behavioral control over activation of stereotypes, initiated by external demands to appear non-prejudiced (i.e., participants were told that the experimenter would monitor their performance to assess whether they showed signs of prejudice).

The mPFC is a key area implicated in the representation of an individual's traits, preferences and mental states during the formation of impression about other people [38]. Furthermore, activity in the mPFC is considered to be associated with a humanization process. Specifically, a lack of activation in this region during presentation of social targets has been suggested to be associated with prejudice, reflecting dehumanization and lack of empathy [39,40]. Crucially, the mPFC has important interconnections with the ACC and the dlPFC – areas involved in conflict monitoring and regulation [20] – and several other regions, including the amygdala, and the orbitofrontal cortex (OFC), implicated in the top-down regulation of emotional responses [36,41]. Building on these premises, Amodio and Frith [36] have proposed that the mPFC may be involved in regulating complex behavioral responses associated with the processing of social information on the basis of external social cues (e.g., the external, not internal, pressure to behave without prejudice). Taken together, these functions make the mPFC a prime candidate area to implement cognitive control over stereotypes activation. However, direct evidence supporting this hypothesis is missing.

The present study aimed at providing preliminary evidence supporting the role of the mPFC in counteracting implicit social stereotypes. To this end, we used transcranial direct current stimulation (tDCS; [42,43]) to induce specific changes of excitability of the mPFC and evaluate the behavioral effects of these changes on participants' performance in a task assessing implicit biased attitudes toward social out-groups. tDCS is a non-invasive brain stimulation technique, that polarity-dependently enhances (anodal tDCS) or reduces (cathodal tDCS) cortical excitability. The primary effects depend on sub-threshold membrane polarization, and prolonged stimulation induces neuroplastic alterations of cortical excitability driven by the glutamatergic system. Beyond its physiological effects, tDCS has been demonstrated to be an effective and promising tool to modulate several cognitive functions [43–46]. Interestingly, tDCS over the mPFC was recently found to modulate error monitoring in conflict-inducing tasks [47], and reactions to fairness [48]. Furthermore, bilateral stimulation of the dorsolateral prefrontal cortex with tDCS has been found to reduce food, alcohol and smoking craving [49–51]. Therefore, tDCS is suited to alter prefrontal physiology, including medial prefrontal areas, and stimulation of this area is functionally effective.

Implicit biases were assessed by means of the Implicit Association Test (i.e., IAT; [52]). The IAT is a well-established behavioral measure that has been extensively used to detect and quantify implicit bias and stereotypes about race, gender, age, politics, religion and several other social groups and constructs [14,53]. The task assesses the strength of an association between stimuli representing social groups and positive and negative attributes. This is achieved by confronting participants with a speeded double categorization task requiring them to categorize, using two response buttons, in-group and out-group names and positive and negative attributes. In one block of trials, in-group names are categorized by using the same response button as positive attributes, whereas out-group names are categorized by using the same response button as negative attributes (i.e., congruent block). In the other block of trials, the stimulus-response mapping is reversed, so that out-group names become associated with (i.e., share the same response button as) positive attributes and in-group names with negative attributes (i.e., incongruent block). The underlying idea is that if people hold implicit negative stereotypes toward a social out-group, they would produce slower and less accurate responses to trials that are inconsistent with their implicit associations (i.e., incongruent block), as compared to trials that are consistent with their implicit associations (i.e., congruent block). This is because, when confronted with incongruent associations, people experience a time-consuming response conflict, whereby the selection of the correct response requires to counteract the overbearing one. Therefore, congruent and incongruent trials differ in terms of the degree of cognitive control that is required to perform the task, with incongruent trials requiring higher cognitive control (cf. conflict monitoring theory; [20]). The difference in reaction times (RTs) and/or percentage of errors (PEs) between congruent and incongruent trials is thus indicative of an individual's bias against a social group, which would be more pronounced the larger such a difference is. Importantly, biased attitudes assessed by the IAT have been found to predict several behavioral forms of discrimination [11].

Based on previous evidence, we assumed that tDCS over the mPFC might modulate implicit biased attitudes, as indexed by performance on the IAT. In particular, to the extent to which the mPFC is involved in counteracting social stereotypes, as recent theories have suggested [33,36], increased cortical excitability of the mPFC induced by anodal tDCS should initiate cognitive-control processes aimed to override biased associations, which would be apparent in incongruent trials (cf. conflict monitoring theory; [20]). If so, participants receiving excitatory anodal tDCS should show less pronounced implicit biases (i.e., smaller differences between congruent and incongruent trials due to faster and/or more accurate responses on incongruent trials) as compared to participants receiving cathodal and sham stimulation. The reduced cortical excitability of the mPFC induced by cathodal stimulation should interfere with the implementation of such control processes, and affect performance accordingly.

Method

Participants

Sixty native 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 considered suitable to participate in this study if they fulfilled the following criteria: i) Dutch for at least three generations back; ii) age between 18 and 32 years; iii) no history of neurological or psychiatric disorders; iv) no history of substance abuse or dependence; v) no history of brain surgery, tumor or intracranial metal implantation; vi) no chronic or acute medications; vii) no pregnancy; viii) no susceptibility to seizures or migraine; ix) no pacemaker or other implanted devices.

Once recruited, participants were randomly assigned to one of the three experimental groups, each receiving only one type of stimulation: anodal ($N = 20$; 8 male; mean age = 22.5 years; age range: 18–27 years), cathodal ($N = 20$; 6 male; mean age = 22.10 years; age range: 18–30 years), or sham ($N = 20$; 7 male; mean age = 21.10 years; age range: 18–27 years). Groups did not differ in terms of age, $F = 1.29$, $P = .28$, or gender distribution, $\chi^2 = .44$, $P = .80$.

All participants were naïve to tDCS. Prior to the testing session, they received a verbal and written explanation of the 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.

Procedure

A single-blinded, sham-controlled experiment was conducted to investigate the effect of a single session of tDCS – applied over the medial prefrontal cortex (mPFC) in healthy Dutch students – on implicit biased attitudes, as measured by the Implicit Association Test (IAT [52]). Specifically, we assessed tDCS-induced effects on implicit negative stereotypes towards Moroccans – a prominent social out-group for ethnic Dutch individuals in the Netherlands.

All participants took part in a single session and were tested individually. After having read and signed the informed consent, active (either anodal or cathodal) or sham stimulation was applied for a period of 20 min. Ten minutes after the onset of the stimulation, participants performed the IAT after being tested on another task assessing interpersonal trust (reported elsewhere [54]). The two tasks lasted for 10 min. Thus, tDCS was applied through the whole task.

After completion of the IAT, participants were properly debriefed and asked to complete 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.

Transcranial direct current stimulation

Direct current was induced by a saline-soaked pair of surface sponge electrodes (5 cm × 7 cm; 35 cm^2) and delivered by a DC Brain Stimulator Plus (NeuroConn, Ilmenau, Germany), a device complying with the Medical Device Directive of the European Union (CE-certified). Electrodes were held in place by rubber bands and the stimulator was placed behind the participants. To stimulate the mPFC, the anode or cathode electrode (depending on the group assignment) was centered horizontally over Fpz (individually measured on each participant) – a location atop the mPFC, according to the international 10-20 system for EEG electrode placement [47,48]; the return electrode was placed horizontally over Oz. The rationale of choosing Oz as return electrode position is that previous studies have suggested that increasing electrodes separation on the head is effective in decreasing the current shunted through the head so as to increase the current density in depth [55], and that the primary visual cortex is not critically involved in the task under study. For anodal tDCS, the anode electrode was placed over Fpz and the cathode electrode was placed over Oz. For the cathodal stimulation, the polarity was reversed. For the active stimulation (either anodal or cathodal), a constant current of 1 mA (current density of .029 mA/cm^2) was delivered for 20 min with a linear fade-in/fade-out of 10 s. These parameters are within safety limits established from prior work in humans [44,56,57]. For sham stimulation, the position of the electrodes, current intensity and fad-in/fade-out were the same as in the active tDCS, but 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 s does not induce after-effects [58]. This procedure has been shown to be effective in blinding participants to their stimulation condition [57,59–61]. The placement of the anode electrode for the sham condition either over Fpz or over Oz was counterbalanced across participants.

Implicit Association Test (IAT)

The set of stimuli consisted of 10 target concepts – 5 typically Dutch names (in-group names; Sander, Hendrik, Stefan, Johan, Hans) and 5 typically Moroccan names (out-group names; Habib, Salim, Sharif, Yousef, Hakim) – and 10 attributes – 5 words representing positive attributes [liefde (love), vrede (peace), gelukkig (happiness), plezier (fun), vreugde (joy)] and 5 words representing negative attributes [pijn (pain), kwaad (evil), verdriet (sadness), falen (fail), vreselijk (terrible)].

The task consisted of 5 blocks, that is, 3 training blocks (i.e., blocks 1, 2 and 4) and 2 test blocks (i.e., blocks 3 and 5), as designed by Greenwald et al. [52]. In the first training block, participants were asked to categorize Dutch and Moroccan names by pressing a left and right button (the "q" and "p" buttons on the computer keyboard, respectively). Half of the participants pressed the left button in response to Dutch names and the right button in response to Moroccan names, whereas the remaining participants received the opposite mapping. In (training) block 2, participants used the same response buttons to categorize words as either positive or negative. In (training) block 4, participants were to discriminate between Dutch and Moroccan names, as they did in block 1, but the response buttons assigned to the target concepts were switched. In (test) blocks 3 and 5, the names discrimination and attributes discrimination tasks were combined. In one of the blocks, in-group names shared the same response button as positive attributes, and out-group names the same response button as negative attributes (i.e., congruent block). In the other block, in-group names were associated with negative attributes and out-group names with positive attributes (i.e., incongruent block). The order of the congruent and incongruent blocks was counterbalanced across participants.

In each trial, the target stimulus was presented in the center of the screen, Verdana 20-point font. In each block, category labels

signaling the response button associated with each category were displayed on the left and right sides of the screen, above the central target. Blocks 1, 2 and 4 consisted of 20 trials, blocks 3 and 5 of 40 trials each. Each trial started with a 500-ms fixation cross, followed by stimulus presentation that remained visible until the response was given. Trials were separated by a blank of 750 ms. Participants were instructed to respond as quickly and accurately as possible.

Results and discussion

First, following the classical procedure used to analyze the IAT data [62], for each stimulation type we computed the D-IAT score – a compound score representing the difference in RTs between congruent and incongruent blocks divided by a pooled standard deviation (SD) of all trials [($M_{incongruent} - M_{congruent}$)/$SD_{pooled}$]. According to the improved scoring algorithm suggested by Greenwald et al. [62], we included all trials and replaced error latencies with a replacement value (M + 600 ms). The resulting D-IAT scores are an indication of people's evaluative bias against Moroccans, with higher D-IAT scores associated with a more pronounced implicit bias. For each stimulation type, a one sample *t*-test was performed to verify whether the D-IAT score differed significantly from zero (i.e., no bias). Analyses revealed that the D-IAT scores differed significantly from zero in all conditions: anodal, $t(19) = 4.07$, $P = .001$ ($M = .43$, SD = .48), cathodal, $t(19) = 6.60$, $P < .0001$ ($M = .76$, SD = .52), and sham, $t(19) = 8.80$, $P < .0001$ ($M = .89$, SD = .45). Thus, a negative implicit bias towards the out-group (i.e., Moroccans) was present in all groups regardless of the stimulation condition. To assess the specific effect of tDCS, the D-IAT scores were submitted to a one-way analysis of variance (ANOVA) with stimulation type (anodal vs. cathodal vs. sham) as between-participants factor. As expected, ANOVA revealed that the size of the implicit bias differed significantly between groups, $F(2,57) = 4.71$, $P = .01$, $\eta_p^2 = .14$. Specifically, Newman–Keuls post-hoc analyses showed that participants who underwent anodal stimulation showed a less pronounced implicit bias (i.e., smaller D-IAT score) than participants in the cathodal ($P = .03$) and the sham condition ($P = .01$), with performance in the latter two conditions being comparable in terms of D-IAT scores ($P = .43$; Fig. 1A).

Next, to gain more insight into how anodal tDCS reduced implicit bias, we analyzed RTs and PEs for the congruent and incongruent blocks. Smaller D-IAT scores can be the product of either slower and/or less accurate responses to (stereotypical) congruent associations, or faster and/or more accurate responses to (counter-stereotypical) incongruent associations. As previously mentioned, congruent and incongruent trials are different in terms of cognitive control that is needed to perform the task. Indeed, responses to congruent associations, which can be taken to reflect the result of a naturally occurring social categorization process, do not depend on the availability of cognitive-control resources that, instead, are necessary to respond to incongruent associations in order to counteract the activation of stereotypic associations, thereby enabling proper task performance. Therefore, if and to the extent to which the mPFC is involved in counteracting the activation of stereotypic associations, anodal (excitatory) tDCS of mPFC is expected to promote faster and more accurate responses in reacting to incongruent trials, compared to cathodal and sham tDCS. Conversely, anodal tDCS of mPFC is not expected to affect responses to congruent trials, and comparable RTs and PEs across the three groups should be observed. To verify this hypothesis, mean RTs and

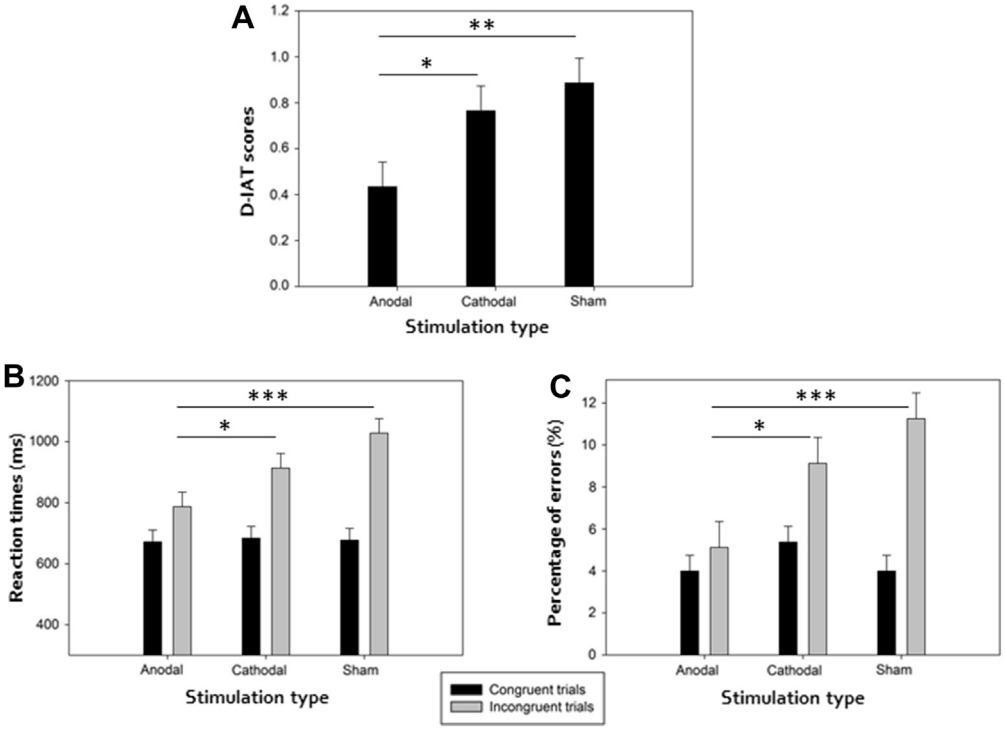

Figure 1. A. D-IAT scores as a function of stimulation type (anodal, cathodal and sham). The D-IAT scores were computed as the difference in reaction times (RTs) on incongruent and congruent trials divided by a pooled standard deviation (SD) of all trials [($M_{incongruent} - M_{congruent}$)/$SD_{pooled}$]. Higher D-IAT scores indicate a more pronounced implicit bias. **B**. Mean reaction times (RTs; in ms) as a function of congruence (congruent vs. incongruent trials) and stimulation type. **C**. Percentage of errors (PEs; in percent) as a function of congruence and stimulation type. Vertical capped lines atop bars indicate standard error of the mean. Asterisks indicate significant differences (Newman–Keuls post-hoc tests; $^{*}P < .05$; $^{**}P < .01$; $^{***}p < .001$). As expected, results showed that anodal tDCS over the mPFC significantly reduced negative implicit biases towards social out-groups (panel A) and that the less pronounced bias shown by participants in the anodal condition was due to better performance (i.e., faster RTs and lower error rates; panel B and C, respectively) of these participants on incongruent trials – a finding that fits with the conflict monitoring theory [20].

PEs for the congruent and incongruent blocks were submitted to separate repeated measures ANOVAs with congruency (congruent vs. incongruent) as within-participants factor and stimulation type as between-participants factor. ANOVAs revealed main effects of congruency, $F(1, 57) = 80.389$, $P < .001$, $\eta_p^2 = .59$ (RT) and $F(1, 57) = 39.247$, $P < .001$, $\eta_p^2 = .41$ (PE). Participants were faster and produced less mistakes in congruent (678 ms and 4.5%) than in incongruent trials (910 ms and 8.5%) – i.e., the IAT effect [11,60]. The main effect of stimulation type was not significant in the RT analysis (mean RTs were 729, 799, and 853 ms in the anodal, cathodal and sham conditions, respectively), $F(2, 57) = 2.8139$, $P = .07$, $\eta_p^2 = .09$, but it was in the PE analysis, $F(2, 57) = 3.8445$, $P = .03$, $\eta_p^2 = .12$. Newman–Keuls post-hoc analyses showed that participants in the anodal condition (4.6%) made less mistakes than participants in the cathodal (7.3%, $P = .03$) and the sham condition (7.6%, $P = .04$), whose performance was comparable ($P = .76$). More importantly, the interactions between congruency and stimulation type were significant, $F(2, 57) = 6.8905$, $P < .005$, $\eta_p^2 = .19$ (RT) and $F(2, 57) = 7.5625$, $P < .005$, $\eta_p^2 = .21$ (PE). Post-hoc tests revealed no significant difference between congruent and incongruent trials for participants who underwent anodal stimulation (672 vs. 787 ms, $P = .06$, and 4.0% vs. 5.1%, $P = .58$, for the RT and PE analyses, respectively; Fig. 1B and C), whereas significant differences were observed for participants who underwent cathodal (684 vs. 914 ms, $P < .001$, and 5.4% vs. 9.1%, $P < .005$, for the RT and PE analyses, respectively) or sham stimulation (677 vs. 1028 ms, $P < .001$, and 4.0% vs. 11.3%, $P < .001$, for the RT and PE analyses, respectively). Crucially, as expected, the three groups of participants did not differ on congruent trials ($P_s \geq .61$), whereas they differed significantly on incongruent trials. Specifically, participants who received anodal tDCS responded faster and made fewer errors on incongruent trials than participants who received cathodal ($P_s \leq .04$) and sham tDCS ($P_s \leq .001$), who were comparable ($P_s \geq .06$). Thus, confirming our expectations, the less pronounced implicit bias (i.e., smaller D-IAT scores) shown by participants in the anodal condition was caused by a smaller difference between responses on incongruent and congruent trials and, specifically, by the fact that these participants responded faster and more accurately on incongruent trials than participants in the cathodal and sham conditions. The present results, therefore, provide evidence favoring the hypothesis that the mPFC plays a critical role in counteracting activated stereotypes, and that increasing cognitive control via anodal (excitatory) tDCS may be functional effective in reducing social stereotyping.

General discussion

Results of recent studies suggest that the mPFC may contribute to self-regulatory and cognitive-control processes implemented to overcome unwanted responses driven by stereotypes activation [33,36]. To explore the causal contribution of this area to the respective processes, we used tDCS [42,43] to alter the cortical excitability of the mPFC and examined the behavioral effects of the induced cortical excitability changes on participants' performance during an IAT [52] evaluating implicit biased attitudes toward Moroccans. Consistent with our expectation, tDCS of the mPFC was effective in modulating implicit biases. Anodal stimulation significantly decreased implicit biases as compared to sham and cathodal tDCS. Because anodal stimulation, as applied here, increases cortical excitability [45], the significant reduction of the D-IAT scores observed provides support for the emerging hypothesis that the mPFC is recruited to control for implicit biased attitudes; this follows the assumption that lower D-IAT scores, as observed in the present study, reflect more efficient cognitive control over stereotype activation [1,13,16–19,22]. Consistent with that, and in line with the conflict-monitoring theory [20], we observed that anodal tDCS of mPFC affected specifically responses to incongruent associations, but not responses to congruent associations. Indeed, results showed that the reduced D-IAT scores in the anodal condition were due to the fact that, compared to participants receiving cathodal or sham stimulation, these participants were faster and more accurate on incongruent trials, that is, on those trials in which cognitive-control resources are needed to overcome the activation of stereotypic associations. By comparison, we did not observe increased D-IAT scores during cathodal stimulation. This might be caused by inefficient stimulation in a spontaneously relatively silent area, as observed for somatosensory cortex stimulation under resting conditions [63]. For an alternative scenario, the failure of cathodal stimulation to produce behavioral effects when non-motor areas are inhibited has been attributed to compensatory processes initiated in the areas surrounding the stimulated one [64]. Finally, it is also possible that the absence of a cathodal modulation is simply due to a ceiling effect, i.e., to the fact that participants who received this stimulation already displayed such a pronounced bias that could not be increased further.

The present results are consistent with previous studies showing that implicit biased attitudes can be controlled in several ways, for example, by emphasizing morality and thus the motivation to suppress them [29], by interfering with the functioning of cerebral areas devoted to the processing of semantic associations [65], or by pharmacological interventions [66]. However, our findings are the first to provide direct evidence for a role of mPFC in counteracting activated stereotypes, presumably in response to increasing cognitive control in the anodal stimulation condition, which again helps to overcome negative bias toward social out-groups.

The present study has some limitations that deserve 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 our parameters of stimulation blinding is reliable [60,61]. Second, although our results provide evidence supporting the relationship between mPFC activity and cognitive control over implicit biased attitudes, the absence of any (electro)physiological measures does not allow to infer the exact mechanism by which mPFC modulation via tDCS reduced implicit attitudes and/or which cognitive-control component was targeted. Thus, it would be valuable for further studies to extend these preliminary findings using event-related potentials (ERPs) to track tDCS-induced real-time changes in the mPFC activity while measuring negative stereotypes. Third, Amodio and Frith [36] observed that the mPFC was uniquely associated with behavioral control driven by external cues and only for those participants who were particularly sensitive to the external pressure to respond without prejudice. This suggests that external vs. internal motivation to respond without prejudice may be critical in determining whether the mPFC is recruited to overcome stereotypes. Given that we did not assess participants' motivation to respond without prejudice, this possibility remains an open question that future studies should address. Fourth, given that the use of relatively large electrodes, as the ones employed in the present study, cannot guarantee selective stimulation of the mPFC, follow-up studies should adopt smaller sized electrodes to increase focality.

Notwithstanding these limitations, the present findings represent an important step in stimulating research to further extend our understanding of the specific role of the mPFC in modulating social and cognitive functioning. For instance, it would be informative for follow-up studies to assess whether anodal tDCS of mPFC modulates performance in other non-social conflict tasks, such as the Stroop [67] and the Simon [68] tasks. The failure to observe any modulation when targeting the mPFC during the execution of non-social conflict tasks would suggest that mPFC is recruited specifically to implement control over stereotypes activation.

Additionally, it would be of interest to compare our findings with possible behavioral effects induced by the stimulation of other areas typically linked with control and self-regulatory processes, such as dlPFC [20,31]. A recent study failed to observed behavioral effects following tDCS of the dlPFC during an insect-flower IAT [69]. However, this failure might be related to the offline tDCS protocol implemented.

To sum up, our findings support the hypothesis that the mPFC is critical for implementing cognitive control over stereotypes activation. This finding has important practical implications as the ability to engage in efficient self-regulation may determine whether activated stereotypes will result in biased behaviors. Moreover, 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.

References

[1] Amodio DM. The neuroscience of prejudice and stereotyping. Nat Rev Neurosci 2014;15:670–82.
[2] Allport GW. The nature of prejudice. Reading, MA: Addison Wesley Publishing Company; 1979.
[3] Hogg MA, Abrams D. Social identifications: a social psychology of intergroup relations and group processes. London: Routledge; 1988.
[4] Medin DL, Smith EE. Concepts and concept formation. Annu Rev Psychol 1984;35:113–38.
[5] Tajfel H, Turner JC. An integrative theory of intergroup conflict. In: Austin WG, Worchel S, editors. The social psychology of intergroup relations. Monterey: Brooks/Cole; 1979. p. 33–47.
[6] Brewer MB. In-group bias in the minimal intergroup situation: a cognitive motivational analysis. Psychol Bull 1979;86:307–24.
[7] Tajfel H, Turner JC. The social identity theory of intergroup behavior. In: Worchel S, Austing WG, editors. Psychology of intergroup relations. Chicago: Nelson-Hall; 1986. p. 7–24.
[8] Greenwald AG, Banaji MR. Implicit social cognition: attitudes self-esteem and stereotypes. Psychol Rev 1995;102:4–27.
[9] Wheeler SC, Petty RE. The effects of stereotype activation on behavior: a review of possible mechanisms. Psychol Bull 2001;127:797–826.
[10] Dovidio JF, Kawakami K, Gaertner SL. Implicit and explicit prejudice and interracial interactions. J Pers Soc Psychol 2002;82:62–8.
[11] Greenwald AG, Poehlman TA, Uhlmann E, Banaji MR. Understanding and using the Implicit Association Test: III Meta-analysis of predictive validity. J Pers Soc Psychol 2009;97:17–41.
[12] Crosby F, Bromley F, Saxe L. Recent unobtrusive studies of black and white discrimination and prejudice. Psychol Bull 1980;87:546–63.
[13] Ito TA, Bartholow BD. The neural correlates of race. Trends Cogn Sci 2009;13:524–31.
[14] Falk E, Lieberman M. The neural bases of attitudes evaluations and behavior change. In: Kruger F, Grafman J, editors. The neural basis of human belief systems. London: Taylor and Francis Psychology Press; 2012. p. 71–94.
[15] Devine PG. Stereotypes and prejudice: their automatic and controlled components. J Pers Soc Psychol 1989;56:5–18.
[16] Conrey FR, Sherman JW, Gawronski B, Hugenberg K, Groom CJ. Separating multiple processes in implicit social cognition: the quad Model of implicit task performance. J Pers Soc Psychol 2005;89:469–87.
[17] Sherman JW, Gawronski B, Gonsalkorale K, Hugenberg K, Allen TJ, Groom CJ. The self-regulation of automatic associations and behavioral impulses. Psychol Rev 2008;115:314–35.
[18] Amodio DM, Devine PG, Harmon-Jones E. Individual differences in the regulation of intergroup bias: the role of conflict monitoring and neural signals for control. J Pers Soc Psychol 2008;94:60–74.
[19] Bartholow BD, Henry EA. Response conflict and affective responses in the control and expression of race bias. Soc Personal Psychol Compass 2010;4:871–88.
[20] Botvinick MM, Braver TS, Carter CS, Barch DM, Cohen JD. Conflict monitoring and cognitive control. Psychol Rev 2001;108:624–52.
[21] Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. Science 2004;303:1023–6.
[22] Richeson JA, Baird AA, Gordon HL, et al. An fMRI examination of the impact of interracial contact on executive function. Nat Neurosci 2003;6:1323–8.
[23] Richeson JA, Shelton JN. When prejudice does not pay effects of interracial contact on executive function. Psychol Sci 2003;14:287–90.
[24] Payne BK. Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. J Pers Soc Psychol 2001;81:181–92.
[25] Easdon CM, Vogel-Sprott M. Alcohol and behavioral control: impaired response inhibition and flexibility in social drinkers. Exp Clin Psychopharmacol 2000;8:387–94.
[26] Hasher L, Zacks RT. Working memory comprehension and aging: a review and a new view. Psychol Learn Motiv 1988;22:193–225.
[27] Bartholow BD, Dickter CL, Sestir MA. Stereotype activation and control of race bias: cognitive control of inhibition and its impairment by alcohol. J Pers Soc Psychol 2006;90:272–87.
[28] Gonsalkorale K, Sherman JW, Allen TJ, Klauer KC, Amodio DM. Accounting for successful control of implicit racial bias the roles of association activation response monitoring and overcoming bias. Pers Soc Psychol Bull 2011;37:1534–45.
[29] van Nunspeet F, Ellemers N, Derks B, Nieuwenhuis S. Moral concerns increase attention and response monitoring during IAT performance: ERP evidence. Soc Cogn Affect Neurosci 2014;9:141–9.
[30] Amodio DM, Devine PG, Harmon-Jones E. A dynamic model of guilt: implications for motivation and self-regulation in the context of prejudice. Psychol Sci 2007;18:524–30.
[31] MacDonald AW, Cohen JD, Stenger VA, Carter CS. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 2000;288:1835–8.
[32] Cunningham WA, Johnson MK, Raye CL, Gatenby JC, Gore JC, Banaji MR. Neural components of conscious and unconscious evaluations of Black and White faces. Psychol Sci 2004;15:806–13.
[33] Amodio DM, Kubota JT, Harmon-Jones E, Devine PG. Alternative mechanisms for regulating racial responses according to internal vs external cues. Soc Cogn Affect Neurosci 2006;1:26–36.
[34] Amodio DM. Coordinated roles of motivation and perception in the regulation of intergroup responses: frontal cortical asymmetry effects on the P2 event-related potential and behavior. J Cogn Neurosci 2010;22:2609–17.
[35] Stanley D, Phelps EA, Banaji MR. The neural basis of implicit attitudes. Curr Dir Psychol Sci 2008;17:164–70.
[36] Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 2006;7:268–77.
[37] Fiske ST, Ames DL, Cikara M, Harris LT. Scanning for scholars: how neuroimaging the MPFC provides converging evidence for interpersonal stratification. In: Derks B, Scheepers D, Ellemers N, editors. Neuroscience of prejudice and intergroup relations. London: Taylor and Francis Psychology Press; 2013. p. 89–109.
[38] Frith CD, Frith U. Interacting minds-a biological basis. Science 1999;286:1692–5.
[39] Harris LT, Fiske ST. Dehumanizing the lowest of the low neuroimaging responses to extreme out-groups. Psychol Sci 2006;17:847–53.
[40] Cikara M, Bruneau EG, Saxe RR. Us and them intergroup failures of empathy. Curr Dir Psychol Sci 2011;20:149–53.
[41] Amodio DM, Ratner K. Mechanisms for the regulation of intergroup responses: a social neuroscience analysis. In: Decety J, Cacioppo JT, editors. Handbook of social neuroscience. New York: Oxford University Press; 2011. p. 729–41.
[42] Paulus W. Transcranial electrical stimulation (tES – tDCS; tRNS tACS) methods. Neuropsychol Rehabil 2011;21:602–17.
[43] Nitsche MA, Paulus W. Transcranial direct current stimulation–update 2011. Restor Neurol Neurosci 2011;29:463–92.
[44] Nitsche MA, Liebetanz D, Lang N, Antal A, Tergau F, Paulus W. Safety criteria for transcranial direct current stimulation tDCS in humans. Clin Neurophysiol 2003;114:2220–2.
[45] Nitsche MA, Cohen LG, Wassermann EM, et al. Transcranial direct current stimulation: State of the art 2008. Brain Stimul 2008;1:206–23.
[46] Kuo MF, Nitsche MA. Effects of transcranial electrical stimulation on cognition. Clin EEG Neurosci 2012;43:192–9.
[47] Bellaïche L, Asthana M, Ehlis A-C, Polak T, Herrmann MJ. The modulation of error processing in the medial frontal cortex by transcranial direct current stimulation. Neurosci J 2013. http://dx.doi.org/10.1155/2013/187692.
[48] Civai C, Miniussi C, Rumiati RI. Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. Soc Cogn Affect Neurosci 2014. pii: nsu154.
[49] Fregni F, Orsati F, Pedrosa W, et al. Transcranial direct current stimulation of the prefrontal cortex modulates the desire for specific foods. Appetite 2008;51:34–41.
[50] Boggio PS, Sultani N, Fecteau S, et al. Prefrontal cortex modulation using transcranial DC stimulation reduces alcohol craving: a double-blind sham-controlled study. Drug Alcohol Depend 2008;92:55–60.
[51] Fregni F, Liguori P, Fecteau S, Nitsche MA, Pascual-Leone A, Boggio PS. Cortical stimulation of the prefrontal cortex with transcranial direct current stimulation reduces cue-provoked smoking craving: a randomized sham-controlled study. J Clin Psychiatry 2008;69:32–40.
[52] Greenwald AG, McGhee DE, Schwartz JLK. Measuring individual differences in implicit cognition: the implicit association test. J Pers Soc Psychol 1998;74:1464–80.
[53] De Houwer J, Teige-Mocigemba S, Spruyt A, Moors A. Implicit measures: a normative analysis and review. Psychol Bull 2009;135:347–68.
[54] Colzato LS, Sellaro R, van den Wildenberg WPM, Hommel B. tDCS of medial prefrontal cortex does not enhance interpersonal trust. J Psychophysiol 2015.
[55] Miranda PD, Lomarev M, Hallett M. Modeling the current distribution during transcranial direct current stimulation. Clin Neurophysiol 2006;117:1623–9.
[56] Nitsche MA, Niehaus L, Hoffmann KT, et al. MRI study of human brain exposed to weak direct current stimulation of the frontal cortex. Clin Neurophysiol 2004;115:2419–23.

[57] Gandiga PC, Hummel FC, Cohen LG. Transcranial DC stimulation tDCS.: a tool for double-blind sham-controlled clinical studies in brain stimulation. Clin Neurophysiol 2006;117:845–50.
[58] Nitsche MA, Paulus W. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 2000;527:633–9.
[59] Poreisz C, Boros K, Antal A, Paulus W. Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. Brain Res Bull 2007;72:208–14.
[60] Ambrus GG, Al-Moyed H, Chaieb L, Sarp L, Antal A, Paulus W. 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 2012;5:499–504.
[61] Palm U, Reisinger E, Keeser D, et al. Evaluation of sham transcranial direct current stimulation for randomized placebo-controlled clinical trials. Brain Stimul 2013;6:690–5.
[62] Greenwald AG, Nosek BA, Banaji MR. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. J Pers Soc Psychol 2003;85:197–216.
[63] Matsunaga K, Nitsche MA, Tsuji S, Rothwell JC. Effect of transcranial DC sensorimotor cortex stimulation on somatosensory evoked potentials in humans. Clin Neurophysiol 2004;115:456–60.
[64] Jacobson L, Koslowsky M, Lavidor M. Review tDCS polarity effects in motor and cognitive domains: a meta-analytical review. Exp Brain Res 2012;216:1–10.
[65] Gallate J, Wong C, Ellwood S, Chi R, Snyder A. Noninvasive brain stimulation reduces prejudice scores on an implicit association test. Neuropsychology 2011;25:185–92.
[66] Terbeck S, Kahane G, McTavish S, Savulescu J, Cowen PJ, Hewstone M. Propranolol reduces implicit negative racial bias. Psychopharmacology 2012;222:419–24.
[67] MacLeod CM. Half a century of research on the Stroop effect: an integrative review. Psychol Bull 1991;109:163–203.
[68] Simon JR, Small Jr AM. Processing auditory information: Interference from an irrelevant cue. J Appl Psychol 1969;53:433–5.
[69] Gladwin TE, den Uyl TE, Wiers RW. Anodal tDCS of dorsolateral prefontal cortex during an Implicit Association Test. Neurosci Lett 2012;517:82–6.