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.