Trends in Cognitive Sciences

CellPress
REVIEWS

Review

Studying Implicit Social Cognition with Noninvasive Brain Stimulation

Maddalena Marini,[1,2,*] Mahzarin R. Banaji,[1] and Alvaro Pascual-Leone[3,4]

Given that globalization has brought different sociocultural groups together on an unprecedented scale, understanding the neurobiology underlying intergroup social behavior has never been more urgent. Social and cognitive scientists are increasingly using noninvasive brain-stimulation techniques (NBS) to explore the neural mechanisms underlying implicit attitudes and stereotyping. NBS methods, such as transcranial magnetic stimulation (TMS) and transcranial direct-current stimulation (tDCS), can interfere with ongoing brain activity in targeted brain areas and distributed networks, and thus offer unique insights into the mechanisms underlying how we perceive, understand, and make decisions about others. NBS represents a promising tool to promote knowledge about the social minds of humans.

Implicit Social Cognition: Attitudes and Stereotypes

Until recently in human history, social groups lived in small units that were genetically and culturally homogeneous [1]. These living conditions supported specific adaptations that are still a part of our nature today. We show strong tendencies to cooperate with people who we perceive to be 'like us', and we are suspicious of strangers, judging and discriminating against those who are not 'like us' [2,3]. Importantly, research conducted on **attitudes** (see Glossary) and **stereotypes** has revealed that these processes can operate implicitly – that is, without intentional and direct control [4–6]. In addition, it has been shown that implicit attitudes and stereotypes can predict behaviors above and beyond explicit attitudes and stereotypes [4,7–9] (a meta-analytical comparison of the predictive power of implicit and explicit measures is given in [10]).

Despite these advances, the mechanisms and the brain areas causally involved in these processes are still not clear. We focus here on the small but growing number of studies that use **noninvasive brain stimulation** (NBS). Unlike traditional imagining techniques, NBS allows researchers to interfere with ongoing brain activity (creating a 'virtual lesion') in targeted cortical areas and distributed brain networks [11]. By directly interfering with brain activity, NBS can provide powerful evidence that specific brain regions are causally related to specific sociocognitive behaviors. In addition, NBS, by modulating the activity of these areas, can yield insights of high relevance into interventions in multiethnic and multinational societies because these social contexts provide natural settings in which primitive in-group allegiances are often in conflict with one's own standards of equal opportunity, fairness, and justice. Over the past few decades, social scientists have proposed several behavioral interventions to produce changes in implicit cognition. However, even the best strategies have produced only limited results [12–14].

In this review, in addition to highlighting the insights provided by NBS techniques, we also strive to point out the limitations of existing studies and of NBS methods more generally. In doing so, we hope to bring attention to the unique possibilities of NBS in advancing research on how

Highlights

Neural mechanisms underlying implicit attitudes and stereotypes can be investigated by means of NBS, a unique technique that allows researchers to detect causal relationships between brain and behavior.

The anterior temporal lobe is a crucial area for the representation of implicit stereotypes, namely conceptual association between attributes (e.g., terrorist and law-abiding) and social groups (e.g., Arab and non-Arab).

The processing of the implicit attitudes, such as religious beliefs, requires the activity of the inferior parietal lobe, a brain area involved in theory of mind and moral decisions.

Modulation of the medial prefrontal cortex can change the expression of implicit stereotypes and attitudes by operating on its control and regulation mechanisms.

Implicit stereotypes and attitudes are mediated by perception processes, such as those related to the physical characteristics (e.g., body) of individuals, that are carried out in the extrastriate body area.

[1]Department of Psychology, Harvard University, Cambridge, MA, USA
[2]Center for Translational Neurophysiology, Istituto Italiano di Tecnologia, Ferrara, Italy
[3]Berenson-Allen Center for Noninvasive Brain Stimulation, and Division of Cognitive Neurology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA
[4]Institut Guttmann de Neurorehabilitació, Universitat Autònoma, Barcelona, Spain

*Correspondence:
Maddalena.Marini@iit.it (M. Marini).

 https://doi.org/10.1016/j.tics.2018.07.014

human beings think about other social beings, especially as members of social groups, and to promote future research that overcomes current methodological constraints.

Mechanisms Underlying Implicit Social Cognition

Investigating the mechanisms underlying **implicit social cognition** (e.g., cognitive processes that form, shape, and maintain attitudes and stereotypes) is crucial to understand how unintentional and indirect thoughts and feelings influence human behavior. To this end, in the past three decades scientists have conducted a considerable number of studies using one of the most common measures of implicit social cognition, the implicit association test (IAT; Box 1) [15]. Behavioral research using the IAT has helped to establish a cognitive model of implicit attitudes and stereotypes [16–22]. For example, it has been shown that implicit attitudes and stereotypes can reflect representations at the level of categories rather than those at the level of the individual exemplars [19].

However, only with the application of neuroscientific techniques have researchers been able to produce a more comprehensive model of potential neural mechanisms underlying implicit attitudes and stereotypes. For example, **functional magnetic resonance imaging** (fMRI) has identified a network of brain regions that track implicit intergroup processes. This network includes the anterior cingulate cortex (ACC), the ventrolateral prefrontal cortex (VLPFC), and the dorsolateral PFC (DLPFC), regions associated with inhibition, conflict resolution, and control processes [23,24]. In addition, the amygdala, an area that is differentially responsive to faces of ingroup and outgroup members [25,26], is correlated with implicit but not explicit attitudes, indicating its role in automatic evaluations of social groups [27]. **Brain lesion** studies have shown that the PFC plays a role in implicit stereotyping [28,29]. For example, lesions in the ventromedial PFC have been associated with an increase in the gender stereotype 'female + weak/male + strong', whereas lesions in the ventrolateral PFC have been related to a reduction of such bias [29]. Finally, **event-related potential** (ERP) studies have elucidated

Box 1. Brief Description of the Implicit Associations Test (IAT)

The IAT [15] is a measure to assess associations between mental representations existing in memory that operate without intentional and direct control [5].

The IAT assesses attitudes and stereotypes by measuring how quickly and accurately a person can categorize and associate stimuli related to two conceptual categories with stimuli belonging to two evaluative attributes. Stimuli representing categories and attributes are presented one at a time in the center of the computer screen, and participants categorize each of them in two different sorting conditions.

For example, in a typical IAT assessing racial bias (i.e., race IAT; Figure I), participants are asked to categorize stimuli representing the two conceptual categories – White people and Black people – and the two evaluative attributes – good and bad. In one condition (congruent condition with the racial bias), participants categorized pictures representing White people and good words (e.g., joy, love, and peace) with one response key, while categorizing pictures representing Black people and bad words (e.g., agony, terrible, and horrible) by using another response key. In the other condition (incongruent condition with the racial bias), participants categorize the same stimuli but with a different key configuration: this time pictures of White people and bad words are categorized with one key, whereas pictures of Black people and good words are categorized with the other. Faster categorization in the congruent condition compared to the reverse is an indicator of an implicit preference for White people compared to Black people.

The IAT has provided relevant insights into social cognition by showing that, even when weak or no social preferences are apparent on explicit measures (i.e., self-reports that assess intentional and controllable responses), substantial degrees of intergroup bias can be detected on implicit measures. For example, it has been shown that, although White Americans report only slight pro-White preferences on self-report measures, they reveal strong pro-White preference on the IAT [120]. In addition, research has shown that the IAT can predict behaviors, judgments, and physiological responses [8]. To experience the IAT first hand, please visit https://implicit.harvard.edu/implicit/takeatest.html.

Glossary

Attitudes: psychological tendencies that are expressed by evaluating a particular entity (e.g., objects, situations, or people) with some degree of favor or disfavor (e.g., good/bad).
Brain lesions: damage to an area of the brain as a result of trauma (e.g., injury) or disease. Lesions result in 'holes' or 'cavities' in the brain and can entail loss of function. Depending on their size and location, lesions generate a blockage or interruption of neural transmission, with minor or major behavioral effects.
Cognitive dissonance: the aversive feeling generated by any inconsistency between preferences or choices (e.g., not choosing a preferred item) that results in a tendency to change original preferences to justify the past behavior and reduce the mental discomfort.
Event-related potential (ERP): a noninvasive technique to evaluate brain functioning and study psychophysiological correlates of mental processes. It measures electrical potentials generated by the brain in response to specific internal or external events (e.g., sensory, cognitive, or motor stimuli). Electrical activity is detected by means of electrodes placed onto various locations on the scalp and amplified through an EEG machine.
Functional magnetic resonance imaging (fMRI): a method to measure brain activity by detecting regional and time-varying changes in brain metabolism and blood oxygenation. This technique relies on the fact that the change of oxygenated versus deoxygenated blood flow increases when a brain area is active. These alterations are detected by using a magnetic field.
Implicit bias: a term referring to 'unidentified or inaccurately identified' attitudes or stereotypes.
Implicit social cognition: social psychological constructs (e.g., attitudes and stereotypes) that influence performance and occur outside of intentional control.
Noninvasive brain stimulation (NBS): a method that allows to alter the neural activity in given brain areas and distributed networks. It

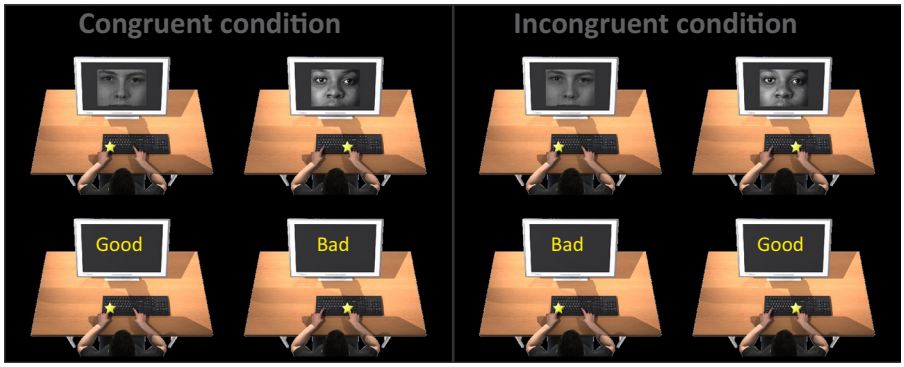

Figure I. Example of an IAT. In a race IAT, participants categorize four types of stimuli (i.e., pictures of White and Black people and words representing concepts of good and bad). Note: in the figure, congruent and incongruent labels are used to define the conditions in which motor responses required in the task are, respectively, compatible or incompatible with the racial bias. We do not use the terms congruent and incongruent normatively – in other words, to imply moral goodness or desirability.

the dynamics and timing of brain processes underlying **implicit bias** [30–32]. For example, it has been shown that implicit bias elicits a characteristic ERP component that peaks at 170 ms (N170) after the presentation of a face [33,34], and that is larger when the stimulus represents an ingroup rather than an outgroup face [35].

Although studies using traditional neuroscientific techniques have provided insights into the neural mechanisms underlying implicit social cognition, they are limited by specific methodological constrains. ERP and fMRI studies are useful for detecting temporal and spatial changes in brain activity that are correlated with performance on a behavioral task, but they cannot provide causal information about such activations *per se*. For example, fMRI techniques can inform about the 'causality' of the brain activity under study (i.e., whether an area is causally involved in a behavior), but only by means of specific analyses and models [36,37]. The causal inferences that can be drawn between brain and behavior based on lesion studies are also limited owing, for example, to adaptation and plasticity of the brain or different location and size of lesions across patients [38].

NBS techniques offer unique advantages to overcome these limitations. By disrupting ongoing neural activity in targeted brain areas or networks at specific times, NBS can directly reveal causality in brain–behavior relations. That is, these techniques can provide information about the biological mechanisms and chronometry (i.e., the timing of the contribution of a given brain region to a specific behavior) of implicit social cognition. In addition, NBS is uniquely able to modulate cognitive processes and produce changes in behavior. Thus, it can be used to interfere with mechanisms underlying implicit attitudes and stereotypes, as well as to modify their expression.

NBS in Social Neuroscience Research

Interest in NBS techniques, particularly **transcranial magnetic stimulation** (TMS) and **transcranial direct-current stimulation** (tDCS; Box 2), is rapidly growing in social neuroscience research (e.g., [26–40])(e.g., [39–53]). According to the PsycINFO database (http://www.apa.org/pubs/databases/psycinfo/index.aspx), the percentage of NBS publications in

includes different techniques such as transcranial magnetic stimulation (TMS) and transcranial direct current stimulation (tDCS).

Parochial behavior: a psychological phenomenon in which outgroup members are punished more severely than ingroup transgressors for violation of social norms.

Repetitive transcranial magnetic stimulation (rTMS): a transcranial magnetic stimulation paradigm that induces repeated single magnetic pulses in the brain to modulate cortical activity. Its effects last beyond the stimulation time.

Stereotypes: specific beliefs (e.g., smart/dumb, strong/weak, hardworking/lazy) about humans that show reliance on their group membership.

Stroop task: a behavioral paradigm used to assess interference in verbal responses. In a classic Stroop task, participants are presented with the name of a color printed in colored characters and are instructed to name the color of the letters as fast and accurately as possible. Faster responses are obtained when the word presented denotes the same color as the characters used to spell it (e.g., the word 'Red' written in red characters) compared to when it denotes a different color (e.g., the word 'Red' written in green characters).

Theory of mind (ToM): the human ability to attribute mental states (e.g., feelings, desires, wishes and goals) to self and others.

Theta-burst stimulation (TBS): a repetitive transcranial magnetic stimulation protocol that uses patterned sequences of magnetic pulses (i.e., bursts) to induce lasting changes in cortical activity (up to 50 minutes).

Transcranial direct-current stimulation (tDCS): a noninvasive brain-stimulation technique that allows changes in cortical activity to be generated by inducing a direct low-intensity current in the brain.

Transcranial magnetic stimulation (TMS): a noninvasive brain-stimulation technique that induces magnetic pulses in the brain to change cortical excitability and neuronal depolarization.

Box 2. Brief Description of Noninvasive Brain Stimulation (NBS)

Different forms of NBS have been developed to alter brain activity. One of the most popular NBS tools is TMS. TMS uses an electric current traveling through a coil to create brief magnetic pulses. The pulses traverse the skull and other matter overlaying the brain, inducing an electric current pulse in the brain which is able to depolarize neurons and influence cortical excitability (Figure IA) [121,122]. TMS can be applied either online, to affect the brain during a task, or offline, to compare task performance before and after stimulation. TMS paradigms include single-pulse TMS (spTMS), paired-pulse TMS (ppTMS), paired associative stimulation (PAS), and repetitive TMS (rTMS). Each paradigm has different applications: spTMS can be used for spatially and temporally mapping behavior-related neurocircuitry and for studying brain–behavior relationships; ppTMS involves the use of two TMS pulses to examine intracortical excitation and inhibition; PAS is used to study plasticity within the sensorimotor system; and rTMS can be utilized to modulate cortical plasticity and track dynamic changes in reactivity [123]. rTMS is thought to either enhance (rTMS $\geq$5 Hz: high-frequency stimulation) or suppress (rTMS $\leq$1 Hz: low-frequency stimulation) cortical activity and modulate excitability in the target area for a period of time that extends beyond the end of the stimulation [124]. Recently, a new rTMS protocol has been introduced, known as theta-burst stimulation (TBS), which can have opposing effects on excitability depending on the temporal pattern of the bursts. In most individuals, intermittent theta-burst (iTBS) induces excitatory effects on brain activity, while continuous theta-burst (cTBS) suppresses cortical activity [125–127].

Unlike TMS, tDCS generates changes in cortical activity by means of a direct low-intensity current flowing between a pair of electrodes applied to the scalp. The current flows from an anodal to a cathodal electrode typically placed over a target brain area (Figure IB). The effect depends on several factors, including the duration of stimulation, its strength, and the polarity of the electrode over the target area. In general, the anodal electrode is associated with an increase in excitability, while an inhibitory effect is observed with the cathodal electrode [128,129]. The tDCS effect is thought to be due to a small change in the membrane potential of cortical neurons [130]. Because of the large size of the electrodes used (typically between 25 to 35 cm^2), tDCS is used to target large areas or cognitive processes that are not highly localized.

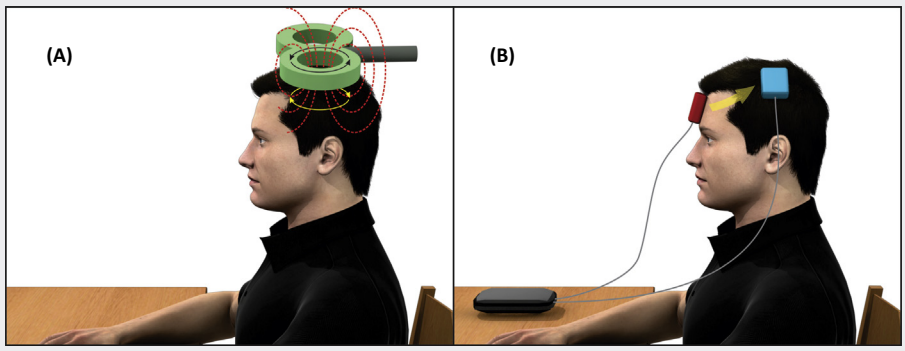

Figure I. Visual Sketch of NBS. (A) TMS induces electrical currents (yellow arrows) in the brain by means of a coil positioned above the head that generates magnetic pulses (red). (B) tDCS induces a direct current in the brain that passes from an anodal (red) to a cathodal (blue) electrode placed on the scalp and that changes the resting electrical charge of a target brain area.

social neuroscience from 2000 to 2017 grew from 1% to 43% (Figure 1). Furthermore, the application of NBS has begun to yield crucial insights into specific areas of social neuroscience (Figure 2).

For example, NBS in research on action perception has allowed researchers to overcome the limitations of correlational studies involving behavioral and brain imaging techniques, and to show not only that there is a relationship between action execution and action perception but also to identify regions where these two processes interact [54]. Indeed, recent studies [55–57] have relied on TMS paradigms that influence the functional state of the neurons by means of perceptual (or motor) adaptation or priming to detect the presence of neurons encoding

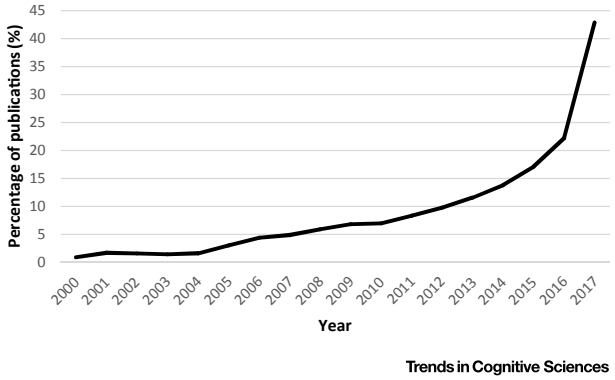

Figure 1. Schematic Illustration of Noninvasive Brain Stimulation (NBS) Publications in Social Neuroscience. Percentages were obtained by dividing the number of publications using the terms 'transcranial magnetic stimulation' or 'transcranial direct current stimulation' and 'social' by the number of publications using the terms 'transcranial magnetic stimulation' or 'transcranial direct stimulation.' The search was limited to the title, abstract, and keywords of publications included in the PsycINFO database from 2000 to 2017.

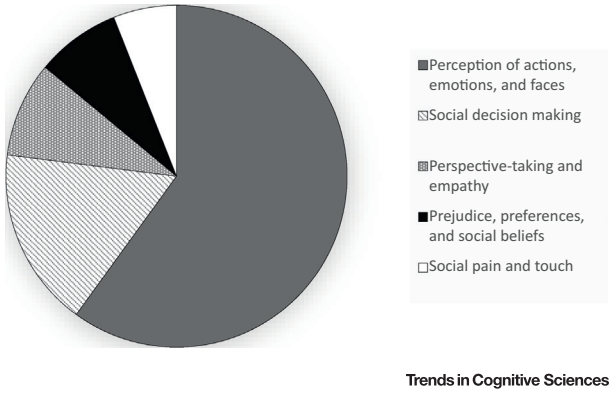

Figure 2. Noninvasive Brain Stimulation (NBS) Publications in Social Neuroscience by Main Topic. Publications were obtained by using the terms 'transcranial magnetic stimulation' or 'transcranial direct current stimulation' and 'social'. The search was limited to the title, abstract, and keywords of publications included in the PsycINFO database from 2000 to 2017.

adapted/primed features in the stimulated area and their relevance to perceptual processing. These paradigms are based on the concept of state-dependency, according to which TMS effects depend on the initial state of stimulated neurons [58–61]. NBS has also provided evidence that motor processes, carried out in areas such as the inferior frontal cortex, (IFC), are sensitive to higher-order aspects of others' actions (e.g., goals and intentions of the actor) [62,63]. For example, in an offline TMS study it has been found that disruption of the IFC abolished the motor facilitation induced by observed actions [62].

Similarly, NBS has led to insights into the brain areas that are causally involved in social decision-making process. For example, it has been shown that the DLPFC is necessary for the implementation of fairness rules (i.e., sharing or distributing resources in an equitable and efficient way) [64], reputation formation [65], strategic decision making [66], and prosocial behaviors [67]. In particular, it has been found that the posterior medial frontal cortex (pMFC) and medial prefrontal cortex (mPFC) are causally involved in inducing preference change after **cognitive dissonance** [68] and in mediating social evaluations [69], respectively. In addition, in studies on social decision making, NBS has identified the right temporoparietal junction (TPJ) as a crucial area for mental state attribution (i.e., inferring the beliefs and intentions of the actor) in moral judgments or decisions [70] and in the implementation of **parochial behaviors** [71].

A promising, but still relatively underinvestigated, application of NBS in social neuroscience research concerns intergroup perception and cognition, specifically attitudes and stereotypes towards groups that vary in social characteristics such as race and ethnicity.

New Insights into Implicit Social Cognition Using NBS

We review here NBS studies published to date on implicit attitude and stereotyping with the goal of illustrating the unique contributions and possibilities that NBS can offer. These studies suggest that NBS has the potential to provide new insights in the field of implicit social cognition; however, it is important to note that this research is still at its beginning. Thus, in addition to highlighting the insights that these studies provide, we also give readers a sense of the limitations of the current work. In doing so, we hope to help to shape the next generation of studies. All the reviewed studies assess implicit cognition by using the IAT, and focus on a brain network that includes the anterior temporal lobe (ATL), inferior parietal lobe (IPL), DLPFC, mPFC, and extrastriate body area (EBA; Table 1 and Figure 3).

Implicit Stereotype Representation

The ATL is known to support semantic memory, including social knowledge about objects, people, words, and facts [72]. Because stereotypes reflect conceptual associations between social groups and attributes that are thought to be located in semantic memory [73], emerging research has suggested that the ATL is involved in the processing of stereotype representation [74].

For example, an fMRI study [75] examined the brain activity representing judgments of Black and White individuals on the basis of stereotype traits (athleticism) versus evaluations (potential for friendship). Results showed that the ATL activity correlated with the implicit stereotypes and attitudes (as measured by the IAT) of the participants when they made trait or evaluative judgments respectively. Similar results were also found in another fMRI study [74] in which participants were asked to consider either social or non-social categories (e.g., men versus women, or violins versus guitars) and judge which category was more likely to be characterized by a particular feature (e.g., enjoys romantic comedies or has six strings). In particular, results showed that, when comparing brain activity between social and non-social conditions, the ATL was uniquely activated during stereotype-relevant judgments of social categories.

Even though these studies indicate that the ATL is involved in knowledge of social stereotypes, they do not provide evidence that this brain region is necessary for their representation. Establishing causal relations requires disrupting ATL activity and assessing its impact on behavior. NBS allows researchers to do exactly that. Indeed, it has been found that 1 Hz **repetitive transcranial magnetic stimulation** (rTMS) (that is thought to produce an inhibitory effect; Box 2) over the right and left ATLs reduced the 'Arab + terrorist/non-Arab + law-abiding' stereotypical association [76]. Similarly, it has been shown that 1 Hz rTMS over both the left and right ATLs decreased the 'male + science/female + humanities' association in a gender IAT [77]. It is interesting to note that, in this latter study, no effect was observed in a control IAT assessing the associations between non-social concepts (i.e., living versus non-living associations) or in a non-semantic control task (i.e., the **Stroop task**).

These results support previous fMRI studies showing that the ATL is specifically involved in some aspects of social stereotyping but not in a more general process of semantic associations [74] or in executive demands. In particular, they provide the first evidence that the ATL is

Table 1. NBS and Behavioral Protocols in Implicit Social Cognition[a]

N	Stimulation protocol	Region stimulated	Task	Results	Refs.
26	Online ppTMS, 65% rMT, ISI 100 ms, active/control	L DLPFC, R aDMPFC, vertex (control)	Gender IAT (i) Male (e.g., Gabriele) (ii) Female (e.g., Francesca) (iii) Strength (e.g., power) (iv) Weakness (e.g., fear)	Stimulation over L DLPFC and R aDMPFC increased D-scores (error rates in the incongruent condition increased) compared to the vertex stimulation	[98]
25	tDCS, anodal stimulation (anodal electrode over right/left EVC and cathodal electrode over Cz), cathodal stimulation (anodal electrode over Cz and cathodal electrode over right/left EVC), EVC localized between O2 and PO8, 2 mA, active/sham	R/L EVC	Weight-valence IAT (i) Thin (ii) Fat (iii) Good (e.g., affable) (iv) Bad (e.g., evil) Weight-esthetic IAT (control) (i) Thin (ii) Fat (iii) Beautiful (e.g., charming) (iv) Ugly (e.g., repulsive)	Cathodal stimulation over L EVC reduced D-scores in the weight-valence IAT	[111]
24	Online ppTMS at 10 Hz, 110% rMT, ISI 100 ms, active/control	R/L IPL, R/L DLPFC, vertex (control)	Religious IAT (i) Religious/spiritual (e.g., soul) (ii) Non-religious/non-spiritual (e.g., agnostic) (iii) Self (e.g., I) (iv) Other (e.g., you) Self-esteem IAT (i) Good (e.g., skillful) (ii) Bad (e.g., ugly) (iii) Self (e.g., I) (iv) Other (e.g., you)	Religious IAT Stimulation over R/L IPL reduced accuracy in the incongruent condition compared to control stimulation. Stimulation over L/R DLPFC reduced accuracy both in the congruent and incongruent conditions compared to control stimulation Self/other IAT Increased error rates in the congruent and incongruent conditions when the L/R DLPFC was stimulated compared to control stimulation	[86]
14	Offline cTBS, three-pulse bursts at 50 Hz every 200 ms (5 Hz) for 20 s, 80% aMT, 300 pulses, active/sham Offline iTBS, three-pulse bursts at 50 Hz every 200 ms, trains for 2 s, and repeated every 10 s for 192 s, 20 trains, 600 pulses, 80% aMT, active/sham	R IPL	Religious IAT (i) Self (e.g., I) (ii) Other (e.g., you) (iii) Religious/spiritual (e.g., soul) (iv) Non-religious/spiritual (e.g., agnostic) Self-esteem IAT (i) Self (e.g., I) (ii) Other (e.g., you) (iii) Good (e.g., skillful) (iv) Bad (e.g., ugly)	iTBS over the R IPL decreased the IAT effect in the religious/spiritual IAT compared to cTBS and sham stimulation. No change on the self-esteem IAT effect	[87]
48	tDCS, anodal electrode over the left DLPFC (F3) or the right IFG, cathodal electrode on the contralateral supraorbital region, 1 mA, active/sham	L DLPFC, R IFG	Affective alcohol IAT (i) Positive words (ii) Negative words (iii) Alcoholic drinks (iv) Regular drinks Motivation IAT (i) Approach words (ii) Avoidance words (iii) Alcoholic drinks (iv) Regular drinks	Decrease of RTs for positive and negative words in the affective IAT after stimulation over L DLPFC compared to R IFG and sham. No effect was observed in the motivational IAT	[105]
40	Offline rTMS at 1 Hz, 90% rMT, 15 minutes, active/sham/control	R/L ATL, Cz (control)	Arab/terrorist IAT (i) Terror words (e.g., hijacker) (ii) Law-abider words (e.g., taxpayer)	Stimulation over L/R ATL reduced D-scores compared to sham and control stimulation	[76]

Table 1. (continued)

N	Stimulation protocol	Region stimulated	Task	Results	Refs.
			(iii) Arab names (e.g., Habib) (iv) Non-Arab names (e.g., Benoit)		
20	tDCS, anodal electrode over DLPFC (F3), cathodal electrode over the right orbit, 1 mA, active/sham	L DLPFC	Insect/flower IAT (i) Insects (e.g., cockroach) (ii) Flowers (e.g., sunflower) (iii) Positive (e.g., friends) (iv) Negative (e.g., filth)	tDCS over the L DLPFC reduced the RTs for words belonging to the target categories 'flowers' and 'insects' in the congruent condition of the IAT compared to sham stimulation	[104]
36	Online ppTMS at 7 Hz, ISI 143 ms, 60% rMT, active/control	mPFC, IPA (control)	Food IAT (i) Tasty (e.g., pizza) (ii) Tasteless (e.g., tofu) (iii) Positive (e.g., love) (iv) Negative (e.g., killer) Self/other IAT (i) Self (e.g., I) (ii) Others (e.g., you) (iii) Positive (e.g., love) (iv) Negative (e.g., killer) Flower/insect IAT (i) Flowers (e.g., rose) (ii) Insects (e.g., bee) (iii) Positive (e.g., love) (iv) Negative (e.g., killer)	Simulation over mPFC increased D-scores in the tasty/tasteless food IAT compared to IPA stimulation and no-TMS condition	[99]
60	tDCS, anodal stimulation (anodal electrode over FPz and cathodal electrode over Oz), cathodal stimulation (anodal electrode over Oz, and cathodal electrode over FPz), 1 mA, active/sham	mPFC	Dutch/Moroccan IAT (i) Dutch names (e.g., Sander) (ii) Moroccan names (e.g., Habib) (iii) Positive words (e.g., love) (iv) Negative words (e.g., pain)	Anodal stimulation reduced the D-scores compared to cathodal and sham stimulation	[102]
45	Offline rTMS at 1 Hz, 90% rMT, 15 minutes of rTMS, active/sham	R/L ATL	Gender IAT (i) Male (e.g., John) (ii) Female (e.g., Emma) (iii) Science (e.g., chemistry) (iv) Humanities (e.g., history) Living/non-living IAT (i) Fish (e.g., salmon) (ii) Birds (e.g., canary) (iii) Boats (e.g., canoe) (iv) Aircraft (e.g., helicopter) Stroop task Four colors, manual response	Stimulation to R/L ATL reduced D-scores on the gender IAT compared to the sham. No change on the non-social IAT and the Stroop task	[77]

[a]Abbreviations: aDMPFC, anterior dorsomedial prefrontal cortex; aMT, active motor threshold; ATL, anterior temporal lobe; EVC, extrastriate visual cortex; cTBS, continuous theta-burst stimulation; DLPFC, dorsolateral prefrontal cortex; D-score, mean difference between incongruent and congruent conditions divided by the overall SD [131]; IAT effect, reaction-time difference between incongruent and congruent conditions; IFG, inferior frontal gyrus; IPA, parietal cortex; IPL, inferior parietal lobe; iTBS, intermittent theta-burst stimulation; L, left; mPFC, medial prefrontal cortex; ppTMS, paired-pulse transcranial magnetic stimulation; R, right; rMT, resting motor threshold; rTMS, repetitive transcranial magnetic stimulation; RT, reaction time; tDCS, transcranial direct-current stimulation.

causally involved in processing stereotypical social associations, and that modulation of its activity can lead to changes in stereotype representation, such as modifying the strength with which a particular set of attributes are associated with a social group. However, more research will be necessary to confirm these findings. Specifically, additional studies clarifying the reasons why rTMS did not affect non-social associations (e.g., in a living versus non-living categories IAT) [77] would be desirable to elucidate the potential social role of the ATL.

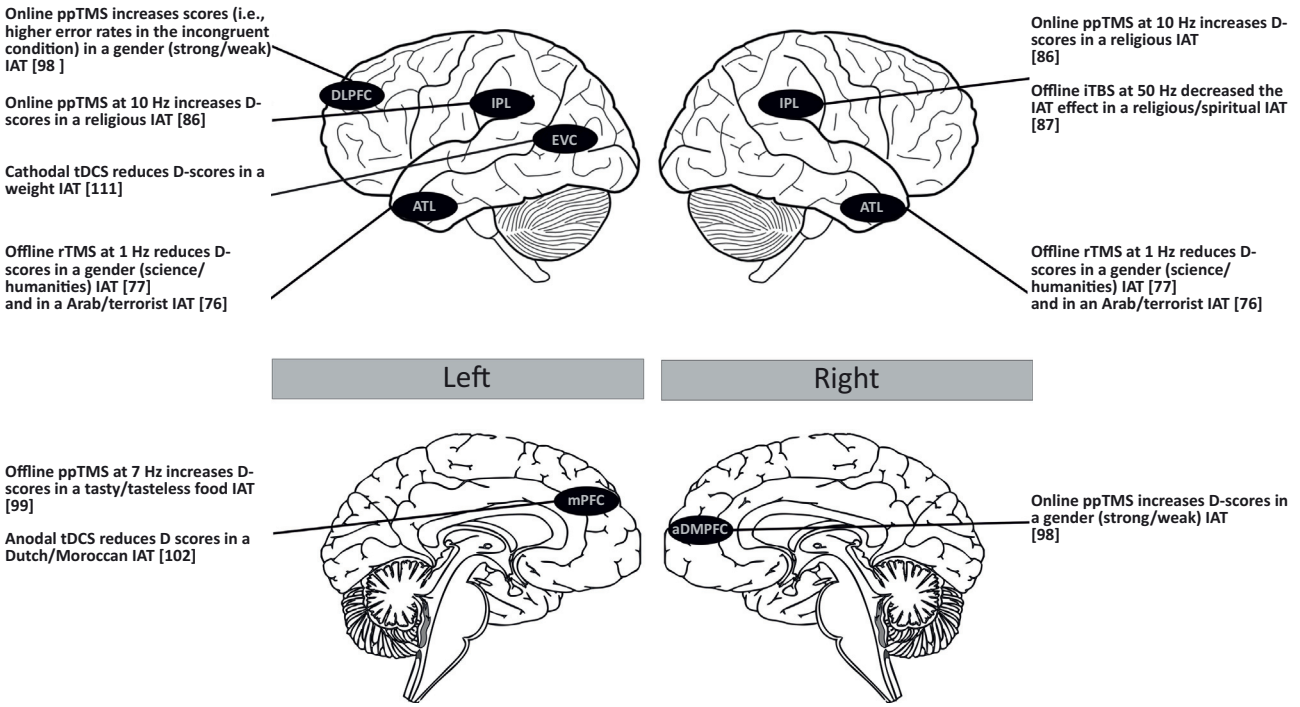

Figure 3. Target-Areas and Main Findings of Noninvasive Brain Stimulation (NBS) Studies. In implicit social cognition, NBS studies report a network of brain areas assumed to be causally involved in attitudes and stereotypes as measured by the IAT. These regions include the anterior temporal lobe (ATL), inferior parietal lobe (IPL), dorsolateral prefrontal cortex (DLPFC), medial prefrontal cortex (mPFC) and a subpart of the extrastriate visual cortex (EVC; i.e., extrastriate body area, EBA). Abbreviations: D-score, mean difference between incongruent and congruent conditions divided by the overall SD; IAT, implicit association test; ppTMS, paired-pulse transcranial magnetic stimulation; tDCS, transcranial direct-current stimulation.

Implicit Attitude Representation

The temporal–parietal junction (TPJ) [78] is an area that is known to play a crucial role in **theory of mind** (ToM) [79–82] and in the ability of individuals to make moral decisions [70,83]. Interestingly, it has been shown that, although the ventral part of the TPJ – namely the posterior superior temporal sulcus (STS) – is commonly associated with social processing [84], attribution of beliefs also engages the dorsal part of the TPJ [85], which includes portions of the inferior parietal lobule (IPL). This was revealed in an fMRI study in which participants were asked to perform a verbal task that included sets of single sentences. All the sentences were identical except for the type of mental state described (i.e., belief, emotion, perception). When participants were required to attribute a belief to either themselves or others, both the STS and IPL showed stronger activation compared to when participants needed to attribute either emotions or perceptions [85].

Given the role of IPL in social beliefs, neuroscientists have hypothesized that it is also involved in the processing of implicit social attitudes. However, only recently, with the development of NBS techniques, has research been able to demonstrate the crucial role of this region in mediating implicit associations underlying social attitudes. In an online TMS study, it was found that disrupting the activity of the left and right IPLs during a religious IAT, which involved assessing automatic associations between the categories ‘self/others’ and the ‘attributes religious/non-religious’, led to a decrease in performance, namely increased error rates. Interestingly, no changes were observed when participants performed a self-esteem IAT (i.e., assessing

associations between categories 'self/other' and attributes 'positive/negative'), suggesting that the left and right IPLs are specifically implicated in the processing of religious attitudes [86]. Similar results have also been found [87] in an offline **theta-burst stimulation** (TBS) paradigm. In this study, TBS was applied over the right IPL before participants performed a religious IAT and a self-esteem IAT. Each participant underwent three different TBS protocols: continuous TBS (cTBS), intermittent TBS (iTBS), and sham stimulation. Results showed that iTBS on the right IPL produced a reduction of the religious attitudes compared to cTBS and sham stimulation. Self-esteem attitudes were unchanged by TBS.

These studies provide causal evidence of the role of the IPL in mediating and processing religious self-representations [86,87]. Future NBS studies investigating and comparing different social attitudes will be necessary to support these results and clarify whether the IPL, in the context of implicit attitudes, is involved in the strengthening of religious self-representations or in a more general cognitive process associated with the processing of implicit ideologies or attitudes.

Control and Regulation of Implicit Attitude and Stereotype

The DLPFC is a brain region assumed to be associated with executive control, goal maintenance, and inhibition of prepotent responses [88]. Research has suggested that the DLPFC plays a crucial role in the control of social attitude and stereotyping [89,90]. In particular, it has been hypothesized that the DLPFC works in combination with the ACC, a neural structure involved in monitoring response competition and in engaging executive control [91]. According to this view, the ACC detects a conflict between intentions and automatic social evaluations, and the DLPFC controls the bias [92]. For example, in a fMRI study it was found that implicit pro-White attitudes were correlated both with the performance of participants on a cognitive task involving control mechanisms of automatic processes (i.e., Stroop task) as well as with their neural activity in ACC and DLPFC. Interestingly, only DLPFC activation mediated the relationship between implicit race preference and Stroop interference, suggesting that this brain region is specifically engaged in a regulatory mechanism to control implicit attitudes [93].

Similarly, it has also been suggested that the mPFC plays a crucial role in regulating and controlling implicit attitudes and stereotypes [90]. The mPFC has been implicated in the processing of social information (e.g., the formation of impressions about other people, attribution of mental states [82,94], and attribution of human qualities [95]) with a prominent number of interconnections, including the ACC and the DLPFC [96]. In particular, it has been proposed that this brain region plays a prominent role in the regulation of behavioral responses associated with social cues (e.g., external pressure to respond without prejudice). In an ERP study, it was shown that responses regulated by external cues to 'respond without prejudice' involved electrophysiological components of error-perception (i.e., error-related negativity, ERN), a process associated with mPFC and rostral ACC (rACC) activity, whereas the regulation of intergroup attitudes by internal cues was associated with conflict-monitoring electrophysiological components (i.e., error-positivity, Pe) and activation of the dorsal ACC (dACC) [97]. This result is remarkable because it suggests that specific brain regions subserve the representation of higher-order goals, such as internal and external incentives, to be unbiased.

Although these studies have highlighted the potential brain structures and mechanisms involved in regulating implicit intergroup responses, they provide no evidence that either DLPFC or mPFC are causally engaged in these social processes. NBS techniques, on the other hand, offer the possibility of investigating the role of these areas in controlling implicit

social biases [86,98,99]. In an online TMS study it has been shown that interfering with the activity of the left DLPFC and right anterior dorsomedial prefrontal cortex (aDMPFC) led to lower performance in a gender IAT. In particular, results showed increased error rates when stereotype-incongruent responses (i.e., 'female + strong/male + weak') were required [98]. Similarly, it has been found that disrupting the activity of the left and right DLPFC during a religious IAT (i.e., assessing associations between categories 'self/other' and attributes 'religious/non-religious') and in a self-esteem IAT (i.e., assessing associations between categories 'self/other' and attributes 'positive/negative') produced higher error rates in both tasks [86]. Analogous results have been found also in a TMS study aimed at investigating the role of the mPFC in controlling food preferences. Results showed that the online stimulation of the mPFC worsened performance in a tasty–tasteless food IAT (i.e., increased response times, RTs, in the incongruent condition) [99].

Taken together, these studies indicate that the DLPFC and mPFC are causally involved in controlling implicit attitudes and stereotypes because interference with their activity by means of NBS significantly affected the IAT performance. However, additional studies will be necessary to clarify the specific role of these brain areas in implicit social cognition. The IAT is a task that relies on cognitive interference between automatic (i.e., respond according to the stereotypical associations) and controlled (i.e., respond according to the task instructions) processes [100,101] that require the activity of the prefrontal areas (e.g., DLPFC and mPFC). The results presented here thus cannot be considered as conclusive in establishing whether the causal involvement of DLPFC and mPFC observed using the IAT may be specifically imputed to control processes elicited by implicit attitudes and stereotypes or by the task itself. Future studies using cognitive conflict tasks that are not related to implicit attitudes and stereotypes (e.g., Stroop task) may be useful to address and clarify this issue.

Interestingly, a recent study has shown that disruption of activity in mPFC modulates intergroup stereotypical associations. This study found that tDCS with anode over the mPFC (a protocol that is thought to have an excitatory effect on the target area; Box 2), decreased implicit bias towards outgroup members (i.e., 'Dutch + positive/Moroccan + negative') [102]. In particular, faster and more accurate responses were observed when negative ingroup associations were required (i.e., 'Dutch + negative/Moroccan + positive'). These results therefore suggest that modulating the activity of the mPFC by NBS may increase or decrease its role in controlling implicit intergroup attitudes.

In addition, recent NBS studies have allowed researchers to empirically test the control exerted by the DLPFC in processing implicit non-social associations and show, as suggested by previous fMRI studies [23,103], that the role of this area can differ on basis of the nature of the implicit associations assessed by the IAT (e.g., social versus non-social associations). Specifically, it has been shown that tDCS with anode over the left DLPFC produced no reduction of the stereotypical associations 'flowers + good/insects + bad', but only produced a decrease in RTs in response to specific stimuli (i.e., words belonging to the target categories 'flowers' and 'insects' in the congruent condition) [104]. Similarly, the same authors [105] in a subsequent study found that tDCS with anode over the left DPLFC did not modulate the scores in an alcohol IAT that assessed associations between the categories 'alcoholic drinks/regular drinks' and the attributes 'positive/negative', but led to a general reduction of RTs in the task. These findings thus showed no specific effect when the IAT was used to assess non-social associations, revealing a different causal involvement of the DLPFC in controlling social and non-social associations that it was only suggested at a correlation level by previous fMRI studies.

Physical Perception and Implicit Attitude and Stereotype

The EBA is a region of visual cortex that is selectively activated when images of human bodies and body parts are presented [106]. Interestingly, fMRI studies have suggested that the EBA is also involved in mediating higher-order cognitive processes [107–110]. For example, a recent study suggested [110] that the EBA is functionally linked with brain areas involved in representing the mental states of others (i.e., the ToM network). In this experiment, participants were asked to observe bodies that had previously been associated with social information (i.e., positive or negative trait-based information, such as 'she donated to charity' or 'he lied on his CV') and non-social information (i.e., neutral information, such as 'she sharpened her pencil'). Results showed a greater coupling between the EBA and the temporal pole of the ToM network when participants observed bodies associated with trait-based information compared to neutral information. Similarly, it has been demonstrated [108] that EBA is sensitive to the stereotype-related status of individuals. In this study, participants were asked to make judgments about pictures of men and women portrayed in gender-stereotypical occupations (e.g., female hairdressers or male airline pilots) and non-stereotypical occupations (e.g., male hairdressers or female airline pilots). Specifically, they categorized stimuli based on of the gender of each target or on the color of a dot located on the picture. Results showed that when participants sorted pictures that violated stereotypical beliefs, activity increased not only in cortical areas associated with executive control (i.e., DLPFC) but also in areas related to person perception, such as the EBA.

With the use of NBS techniques, it has recently been possible to further investigate the role of the EBA in social expectations and provide evidence of its causal role in stereotypical beliefs. For example, a study [111] has used tDCS to directly interfere with cortical excitability in the EBA and investigated its effects on social attitudes and stereotypes based on body size. This experiment consisted of two separated sessions in which tDCS electrodes were placed over the left or right extrastriate visual cortex. In each session, participants underwent three different stimulations (i.e., anodal, cathodal, and sham stimulation). After each stimulation they performed two IATs: a weight IAT measuring the implicit attitude 'fat + bad/thin + good', and an esthetic IAT evaluating the implicit stereotype 'ugly + fat/beauty + thin' as a control condition. Results showed that tDCS with cathode over the left extrastriate visual cortex (a protocol that is thought to have an inhibitory effect on the target area; Box 2) reduced implicit weight attitude 'fat + bad/thin + good' as compared to sham stimulation over the same hemisphere. This result was observed only in male participants, who showed a significant implicit weight attitude in the sham condition, but not in female participants, who showed no significant implicit weight attitude in the sham condition.

These findings show that areas implicated in perception are also causally involved in processes that mediate thoughts and beliefs about others, and that by modulating the neural excitability of these areas it is possible to change the expression of attitudes based on physical characteristics. NBS thus, compared to previous fMRI studies suggesting that attitudes are sensitive to early perceptual processes, provides the first causal evidence that perception actively contributes to the cognitive formation and expression of implicit attitudes. Importantly, this result suggests that implicit attitudes do not entail only the activity of so-called higher-level brain areas (e.g., DLPFC, ATL) but also involve lower-level regions (e.g., EBA) that are associated with the representation of perceptual information.

Methodological Challenges of NBS

The research reviewed here suggests that NBS is a useful method to advance the knowledge of neurobiological mechanisms of implicit social cognition, specifically the role of attitudes and

stereotypes. However, it is important to point out that targeting a given brain region with NBS does not necessarily prove that the observed effects at the behavioral level are in fact due to modification of activity in the targeted brain area. NBS exerts an effect both on the stimulated area and distributed networks associated with it [112]. This implies that care must be taken when designing NBS experiments and interpreting their results. In addition, directly and precisely targeting a given brain region may not be always possible. For example, the ATL is a brain region surrounded by cerebrospinal fluid (CSF), and CSF can shunt the electric current induced by TMS [113], thereby reducing the spatial accuracy of the stimulation and making targeting unreliable even if neuronavigation systems are used.

Therefore, NBS should be ideally used in real-time combination with brain imaging methods (e.g., fMRI and electroencephalography, EEG) which provide insights into the physiological impact produced by NBS on the brain areas that mediate the effect on behavior. Only the simultaneous monitoring of the physiological impact and the behavioral consequences of NBS, and the examination of the relation between them, can provide conclusive evidence concerning the neurobiological substrates of social cognitive processes.

Concluding Remarks and Future Directions

In the increasingly multicultural societies that characterize the world today, understanding intergroup attitudes and stereotypes is especially important both for theory and praxis. For almost a century, intergroup attitudes have been studied using behavioral methods that have provided great insight into the beliefs and attitudes humans have about their own groups and those of others, as well as into how they can operate and implicitly influence behavior.

Recently, with the developmental of neuroscience techniques (e.g., ERP and fMRI) it has also been possible to investigate the neurobiological substrates of implicit social cognition. However, these studies have been limited in offering causal relationship between these social processes and brain. Advances in NBS techniques have fundamentally changed this. According to the studies we have reviewed, we are now able to define a potential neural network for implicit attitudes and stereotypes. This network includes the EBA, IPL, ATL, DLPFC, and mPFC. The EBA is involved in the processing of perceptual information that contributes to the shaping and expression of implicit attitudes. That is, the EBA may activate early social beliefs and thoughts about others on the basis of their physical characteristics. Furthermore, it appears that activity in the IPL and ATL supports the representation of implicit attitudes and stereotypes. The IPL may play a role in strengthening implicit beliefs (and perhaps making them resistant to change), and the ATL may sustain stereotypical associations between attributes and social groups. Finally, control and regulation processes appear to be carried out by the DLPFC and mPFC, in the service of regulating and monitoring the shaping, processing, and maintaining implicit attitudes and stereotypes, as well as of their expression.

More research is necessary to confirm the role of these areas in implicit social cognition as well as to identify potential brain regions and processes that may be additionally involved in such a phenomenon. For example, it would be of interest to investigate whether motor and pre-motor areas are causally engaged in the processing of implicit attitudes and stereotypes to explore whether implicit social cognition is shaped by information from the physical body of the organism. Recent research has suggested that it is possible to modulate implicit attitudes and stereotypes by exposing individuals to bodily illusions that induce ownership over a body different from their own with respect to race, gender, or age [114–117]. These results have been interpreted as an indication that implicit attitudes and stereotypes may occur via a process of self-association that starts in the physical body (i.e., with the perception of physical similarity

Outstanding Questions

Can we create NBS protocols that are able to modulate different attitudes, stereotypes, and beliefs?

Can we use NBS to create new positive social associations?

How long can the modulation of implicit cognition using NBS last? Can we generate longlasting modulation of implicit attitudes and stereotypes?

What is the contribution of different brain areas to implicit cognition? Does it differ on the basis of the bias evaluated (e.g., race, weight, gender, etc.)? Is it influenced by the nature of the implicit associations assessed by the IAT (e.g., social versus non-social associations)?

Is the IPL associated with processing of religious (and perhaps other ideological) beliefs in particular, or with the more general processing of implicit social cognition?

Are all implicit attitudes and stereotypes influenced by perceptual processes? Do perceptual processes only mediate biases that involve specific physical characteristics (e.g., body)?

Does the modulation of the DLPFC and mPFC reflect a change in the control mechanisms underlying the implicit social cognition, or a general effect on cognitive interference elicited in the task?

between our body and the other body) and extends to the conceptual domain (i.e., with the generalization of positive self-associations and negative other-associations) [118].

In this review we have discussed the unique benefits that NBS techniques can offer to understanding the underlying processes of implicit attitudes and stereotypes, and have summarized the main insights that have been achieved about implicit social cognition using NBS methods. However, the application of NBS in social cognitive research is still at the starting line, and more research will be necessary to clarify how individuals process social information and interact with others (see Outstanding Questions). Research would benefit from combining NBS with brain imaging methods and greater control over measures at the behavioral level to test the specificity of the effects induced by NBS, such as by using IATs assessing a wider range of social cognition domains, comparisons to non-social cognition, and additional tasks that reflect cognitive conflict similar to that of the IAT [119] (e.g., Stroop task).

Acknowledgments

This review was supported by the 2016-2017 Postdoctoral Fellow Award, Harvard Mind Brain Behavior Interfaculty Initiative (MBB).

References

1. Dunbar, R.I.M. (1992) Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* 22, 469–493
2. Dovidio, J.F. *et al.* (2010) *Prejudice, Stereotyping and Discrimination: Theoretical and Empirical Overview,* Sage
3. Liberman, Z. *et al.* (2017) The origins of social categorization. *Trends Cogn. Sci.* 21, 556–568
4. Fazio, R.H. *et al.* (1995) Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol.* 69, 1013–1027
5. Greenwald, A.G. and Banaji, M.R. (1995) Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol. Rev.* 102, 4–27
6. Nosek, B.A. *et al.* (2012) Implicit social cognition. In *Handbook of Social Cognition* (Fiske, S.T. and Macrae, C.N., eds), pp. 31–53, Sage
7. Dovidio, J.F. *et al.* (1997) On the nature of prejudice: automatic and controlled processes. *J. Exp. Soc. Psychol.* 33, 510–540
8. Greenwald, A.G. *et al.* (2009) Understanding and using the implicit association test. III. Meta-analysis of predictive validity. *J. Pers. Soc. Psychol.* 97, 17–41
9. Green, A.R. *et al.* (2007) Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *J. Gen. Intern. Med.* 22, 1231–1238
10. Kurdi, B. *et al.* (2018) Relationship between the implicit association test and intergroup behavior: a meta-analysis. *Am. Psychol.* (in press)
11. Rotenberg, A. *et al.*, eds (2014) *Transcranial Magnetic Stimulation*, Springer New York
12. Lai, C.K. *et al.* (2014) Reducing implicit racial preferences. I. A comparative investigation of 17 interventions. *J. Exp. Psychol. Gen.* 143, 1765–1785
13. Lai, C.K. *et al.* (2016) Reducing implicit racial preferences. II. Intervention effectiveness across time. *J. Exp. Psychol. Gen.* 145, 1001–1016
14. Marini, M. *et al.* (2012) The role of self-involvement in shifting IAT effects. *Exp. Psychol.* 59, 348–354
15. Greenwald, A.G. *et al.* (1998) Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480
16. Brendl, C.M. *et al.* (2001) How do indirect measures of evaluation work? Evaluating the inference of prejudice in the implicit association test. *J. Pers. Soc. Psychol.* 81, 760–773
17. Greenwald, A.G. *et al.* (2005) Validity of the salience asymmetry interpretation of the implicit association test: comment on Rothermund and Wentura (2004). *J. Exp. Psychol. Gen.* 134, 420-5-30
18. Hall, G. *et al.* (2003) Acquired equivalence and distinctiveness in human discrimination learning: evidence for associative mediation. *J. Exp. Psychol. Gen.* 132, 266–276
19. De Houwer, J. (2001) A structural analysis of indirect measures of attitudes. *J. Exp. Soc. Psychol.* 37, 443–451
20. Mierke, J. and Klauer, K.C. (2003) Method-specific variance in the Implicit Association Test. *J. Pers. Soc. Psychol.* 85, 1180–1192
21. Olson, M.A. and Fazio, R.H. (2003) Relations between implicit measures of prejudice. *Psychol. Sci.* 14, 636–639
22. Rothermund, K. and Wentura, D. (2004) Underlying processes in the Implicit Association Test: dissociating salience from associations. *J. Exp. Psychol. Gen.* 133, 139–165
23. Chee, M.W. *et al.* (2000) Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *Neuroreport* 11, 135–140
24. Cunningham, W.A. *et al.* (2004) Separable neural components in the processing of black and white faces. *Psychol. Sci.* 15, 806–813
25. Stanley, D.A. *et al.* (2012) Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 744–753
26. Ronquillo, J. *et al.* (2007) The effects of skin tone on race-related amygdala activity: an fMRI investigation. *Soc. Cogn. Affect. Neurosci.* 2, 39–44
27. Phelps, E.A. *et al.* (2000) Performance on indirect measures of race evaluation predicts amygdala activation. *J. Cogn. Neurosci.* 12, 729–738
28. Milne, E. and Grafman, J. (2001) Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J. Neurosci.* 21, RC150
29. Gozzi, M. *et al.* (2009) Dissociable effects of prefrontal and anterior temporal cortical lesions on stereotypical gender attitudes. *Neuropsychologia* 47, 2125–2132
30. Schiller, B. *et al.* (2016) Clocking the social mind by identifying mental processes in the IAT with electrical neuroimaging. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2786–2791

31. Forbes, C.E. *et al.* (2012) Identifying temporal and causal contributions of neural processes underlying the implicit association test (IAT). *Front. Hum. Neurosci.* 6, 320
32. Hilgard, J. *et al.* (2015) Characterizing switching and congruency effects in the implicit association test as reactive and proactive cognitive control. *Soc. Cogn. Affect. Neurosci.* 10, 381–388
33. Eimer, M. (2000) Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clin. Neurophysiol.* 111, 694–705
34. Carmel, D. and Bentin, S. (2002) Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* 83, 1–29
35. Ratner, K.G. and Amodio, D.M. (2013) Seeing 'us vs: them': minimal group effects on the neural encoding of faces. *J. Exp. Soc. Psychol.* 49, 298–301
36. Friston, K.J. *et al.* (2005) Modeling brain responses. *Int. Rev. Neurobiol.* 66, 89–124
37. Goebel, R. *et al.* (2003) Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* 21, 1251–1261
38. Pascual-Leone, A. *et al.* (2000) Transcranial magnetic stimulation in cognitive neuroscience – virtual lesion, chronometry, and functional connectivity. *Curr. Opin. Neurobiol.* 10, 232–237
39. Novembre, G. *et al.* (2014) Motor simulation and the coordination of self and other in real-time joint action. *Soc. Cogn. Affect. Neurosci.* 9, 1062–1068
40. Sowden, S. and Catmur, C. (2015) The role of the right temporoparietal junction in the control of imitation. *Cereb. Cortex* 25, 1107–1113
41. Bardi, L. *et al.* (2015) Eliminating mirror responses by instructions. *Cortex* 70, 128–136
42. Mattiassi, A.D.A. *et al.* (2014) Conscious and unconscious representations of observed actions in the human motor system. *J. Cogn. Neurosci.* 26, 2028–2041
43. Naish, K.R. *et al.* (2016) Stimulation over primary motor cortex during action observation impairs effector recognition. *Cognition* 149, 84–94
44. Cross, K.A. and Iacoboni, M. (2014) To imitate or not: avoiding imitation involves preparatory inhibition of motor resonance. *Neuroimage* 91, 228–236
45. Hogeveen, J. *et al.* (2014) Power changes how the brain responds to others. *J. Exp. Psychol. Gen.* 143, 755–762
46. Pisoni, A. *et al.* (2014) Fair play doesn't matter: MEP modulation as a neurophysiological signature of status quo bias in economic interactions. *Neuroimage* 101, 150–158
47. Wang, H. *et al.* (2016) Rhythm makes the world go round: a MEG–TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex* 75, 68–81
48. Kelly, Y.T. *et al.* (2014) Attributing awareness to oneself and to others. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5012–5017
49. Schuwerk, T. *et al.* (2014) Inhibiting the posterior medial prefrontal cortex by rTMS decreases the discrepancy between self and other in theory of mind reasoning. *Behav. Brain Res.* 274, 312–318
50. De Coster, L. *et al.* (2014) Effects of being imitated on motor responses evoked by pain observation: exerting control determines action tendencies when perceiving pain in others. *J. Neurosci.* 34, 6952–6957
51. Jacquet, P.O. *et al.* (2016) Changing ideas about others' intentions: updating prior expectations tunes activity in the human motor system. *Sci. Rep.* 6, 26995
52. Tidoni, E. *et al.* (2013) Action simulation plays a critical role in deceptive action recognition. *J. Neurosci.* 33, 611–623
53. Pobric, G. *et al.* (2016) Hemispheric specialization within the superior anterior temporal cortex for social and nonsocial concepts. *J. Cogn. Neurosci.* 28, 351–360
54. Avenanti, A. *et al.* (2013) Vicarious motor activation during action perception: beyond correlational evidence. *Front. Hum. Neurosci.* 7, 185
55. Silvanto, J. *et al.* (2008) State-dependency in brain stimulation studies of perception and cognition. *Trends Cogn. Sci.* 12, 447–454
56. Avenanti, A. and Urgesi, C. (2011) Understanding 'what' others do: mirror mechanisms play a crucial role in action perception. *Soc. Cogn. Affect. Neurosci.* 6, 257–259
57. Silvanto, J. and Pascual-Leone, A. (2012) Why the assessment of causality in brain-behavior relations requires brain stimulation. *J. Cogn. Neurosci.* 24, 775–777
58. Lang, N. *et al.* (2004) Preconditioning with transcranial direct current stimulation sensitizes the motor cortex to rapid-rate transcranial magnetic stimulation and controls the direction of after-effects. *Biol. Psychiatry* 56, 634–639
59. Siebner, H.R. *et al.* (2004) Preconditioning of low-frequency repetitive transcranial magnetic stimulation with transcranial direct current stimulation: evidence for homeostatic plasticity in the human motor cortex. *J. Neurosci.* 24, 3379–3385
60. Siebner, H.R. *et al.* (2009) How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex* 45, 1035–1042
61. Bestmann, S. *et al.* (2010) The role of contralesional dorsal premotor cortex after stroke as studied with concurrent TMS-fMRI. *J. Neurosci.* 30, 11926–11937
62. Avenanti, A. *et al.* (2013) Compensatory plasticity in the action observation network: virtual lesions of STS enhance anticipatory simulation of seen sctions. *Cereb. Cortex* 23, 570–580
63. Enticott, P.G. *et al.* (2012) Transcranial direct current stimulation (tDCS) of the inferior frontal gyrus disrupts interpersonal motor resonance. *Neuropsychologia* 50, 1628–1631
64. Knoch, D. *et al.* (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832
65. Knoch, D. *et al.* (2009) Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20895–20899
66. Soutschek, A. *et al.* (2015) The importance of the lateral prefrontal cortex for strategic decision making in the prisoner's dilemma. *Cogn. Affect Behav. Neurosci.* 15, 854–860
67. Balconi, M. and Canavesio, Y. (2014) High-frequency rTMS on DLPFC increases prosocial attitude in case of decision to support people. *Soc. Neurosci.* 9, 82–93
68. Izuma, K. *et al.* (2015) A causal role for posterior medial frontal cortex in choice-induced preference change. *J. Neurosci.* 35, 3598–3606
69. Ferrari, C. *et al.* (2016) Interfering with activity in the dorsomedial prefrontal cortex via TMS affects social impressions updating. *Cogn. Affect Behav. Neurosci.* 16, 626–634
70. Young, L. *et al.* (2010) Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc. Natl. Acad. Sci. U. S. A.* 107, 6753–6758
71. Baumgartner, T. *et al.* (2014) Diminishing parochialism in intergroup conflict by disrupting the right temporo-parietal junction. *Soc. Cogn. Affect. Neurosci.* 9, 653–660
72. Patterson, K. *et al.* (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987
73. Hamilton, D.L. and Sherman, J.W. (1994) Stereotypes. In *Handbook of Social Cognition (2nd edn), Vol. 2: Applications* (Wyer, R.S., Jr and Srull, T.K., eds), pp. 1–68, Psychology Press
74. Contreras, J.M. *et al.* (2012) Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Soc. Cogn. Affect. Neurosci.* 7, 764–770
75. Gilbert, S.J. *et al.* (2012) Evaluative vs. trait representation in intergroup social judgments: distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia* 50, 3600–3611

76. Gallate, J. *et al.* (2011) Noninvasive brain stimulation reduces prejudice scores on an implicit association test. *Neuropsychology* 25, 185–192
77. Wong, C.L. *et al.* (2012) Evidence for a social function of the anterior temporal lobes: low-frequency rTMS reduces implicit gender stereotypes. *Soc. Neurosci.* 7, 90–104
78. Abu-Akel, A. and Shamay-Tsoory, S. (2011) Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia* 49, 2971–2984
79. Saxe, R. and Kanwisher, N. (2003) People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *Neuroimage* 19, 1835–1842
80. Bardi, L. *et al.* (2017) Repetitive TMS of the temporo-parietal junction disrupts participant's expectations in a spontaneous theory of mind task. *Soc. Cogn. Affect. Neurosci.* 12, 1775–1782
81. Schurz, M. *et al.* (2017) Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: a review using probabilistic atlases from different imaging modalities. *Hum. Brain Mapp.* 38, 4788–4805
82. Koster-Hale, J. *et al.* (2017) Mentalizing regions represent distributed, continuous, and abstract dimensions of others' beliefs. *Neuroimage* 161, 9–18
83. Koster-Hale, J. *et al.* (2013) Decoding moral judgments from neural representations of intentions. *Proc. Natl. Acad. Sci. U. S. A.* 110, 5648–5653
84. Carter, R.M. and Huettel, S.A. (2013) A nexus model of the temporal–parietal junction. *Trends Cogn. Sci.* 17, 328–336
85. Zaitchik, D. *et al.* (2010) Mental state attribution and the temporoparietal junction: an fMRI study comparing belief, emotion, and perception. *Neuropsychologia* 48, 2528–2536
86. Crescentini, C. *et al.* (2014) Virtual lesions of the inferior parietal cortex induce fast changes of implicit religiousness/spirituality. *Cortex* 54, 1–15
87. Crescentini, C. *et al.* (2015) Excitatory stimulation of the right inferior parietal cortex lessens implicit religiousness/spirituality. *Neuropsychologia* 70, 71–79
88. Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
89. Kubota, J.T. *et al.* (2012) The neuroscience of race. *Nat. Neurosci.* 15, 940–948
90. Amodio, D.M. (2014) The neuroscience of prejudice and stereotyping. *Nat. Rev. Neurosci.* 15, 670–682
91. Botvinick, M.M. *et al.* (2001) Conflict monitoring and cognitive control. *Psychol. Rev.* 108, 624–652
92. Stanley, D. *et al.* (2008) The neural basis of implicit attitudes. *Curr. Dir. Psychol. Sci.* 17, 164
93. Richeson, J.A. *et al.* (2003) An fMRI investigation of the impact of interracial contact on executive function. *Nat. Neurosci.* 6, 1323–1328
94. Frith, C.D. and Frith, U. (1999) Interacting minds – a biological basis. *Science* 286, 1692–1695
95. Harris, L.T. and Fiske, S.T. (2006) Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychol. Sci.* 17, 847–853
96. Amodio, D.M. and Frith, C.D. (2006) Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277
97. Amodio, D.M. *et al.* (2006) Alternative mechanisms for regulating racial responses according to internal vs external cues. *Soc. Cogn. Affect. Neurosci.* 1, 26–36
98. Cattaneo, Z. *et al.* (2011) The role of the prefrontal cortex in controlling gender-stereotypical associations: a TMS investigation. *Neuroimage* 56, 1839–1846
99. Mattavelli, G. *et al.* (2015) Transcranial magnetic stimulation of medial prefrontal cortex modulates implicit attitudes towards food. *Appetite* 89, 70–76
100. Conrey, F.R. *et al.* (2005) Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *J. Pers. Soc. Psychol.* 89, 469–487
101. Ranganath, K.A. *et al.* (2008) Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *J. Exp. Soc. Psychol.* 44, 386–396
102. Sellaro, R. *et al.* (2015) Reducing prejudice through brain stimulation. *Brain Stimul.* 8, 891–897
103. Luo, Q. *et al.* (2006) The neural basis of implicit moral attitude – an IAT study using event-related fMRI. *Neuroimage* 30, 1449–1457
104. Gladwin, T.E. *et al.* (2012) Anodal tDCS of dorsolateral prefontal cortex during an Implicit Association Test. *Neurosci. Lett.* 517, 82–86
105. den Uyl, T.E. *et al.* (2015) Transcranial direct current stimulation, implicit alcohol associations and craving. *Biol. Psychol.* 105, 37–42
106. Downing, P.E. and Peelen, M.V. (2016) Body selectivity in occipitotemporal cortex: causal evidence. *Neuropsychologia* 83, 138–148
107. Quadflieg, S. *et al.* (2015) The neural basis of perceiving person interactions. *Cortex* 70, 5–20
108. Quadflieg, S. *et al.* (2011) Stereotype-based modulation of person perception. *Neuroimage* 57, 549–557
109. Greven, I.M. *et al.* (2016) Linking person perception and person knowledge in the human brain. *Soc. Cogn. Affect. Neurosci.* 11, 641–651
110. Greven, I.M. and Ramsey, R. (2017) Person perception involves functional integration between the extrastriate body area and temporal pole. *Neuropsychologia* 96, 52–60
111. Cazzato, V. *et al.* (2017) Cathodal transcranial direct current stimulation of the extrastriate visual cortex modulates implicit anti-fat bias in male, but not female, participants. *Neuroscience* 359, 92–104
112. Eldaief, M.C. *et al.* (2011) Transcranial magnetic stimulation modulates the brain's intrinsic activity in a frequency-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* 108, 21229–21234
113. Wagner, T. *et al.* (2007) Noninvasive human brain stimulation. *Annu. Rev. Biomed. Eng.* 9, 527–565
114. Maister, L. *et al.* (2013) Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition* 128, 170–178
115. Peck, T.C. *et al.* (2013) Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious. Cogn.* 22, 779–787
116. Banakou, D. *et al.* (2013) Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12846–12851
117. Fini, C. *et al.* (2013) Embodying an outgroup: the role of racial bias and the effect of multisensory processing in somatosensory remapping. *Front. Behav. Neurosci.* 7, 165
118. Maister, L. *et al.* (2015) Changing bodies changes minds: owning another body affects social cognition. *Trends Cogn. Sci.* 19, 6–12
119. Sherman, J.W. *et al.* (2008) The self-regulation of automatic associations and behavioral impulses. *Psychol. Rev.* 115, 314–335
120. Nosek, B.A. *et al.* (2002) Harvesting implicit group attitudes and beliefs from a demonstration web site. . 6, 101–115
121. Di Lazzaro, V. *et al.* (2004) The physiological basis of transcranial motor cortex stimulation in conscious humans. *Clin. Neurophysiol.* 115, 255–266
122. Eldaief, M.C. *et al.* (2013) Transcranial magnetic stimulation in neurology: a review of established and prospective applications. *Neurol. Clin. Pract.* 3, 519–526
123. McClintock, S.M. *et al.* (2011) Transcranial magnetic stimulation: a neuroscientific probe of cortical function in schizophrenia. *Biol. Psychiatry* 70, 19–27

124. Hallett, M. (2007) Transcranial magnetic stimulation: a primer. *Neuron* 55, 187–199

125. Suppa, A. *et al.* (2016) Ten years of theta burst stimulation in humans: established knowledge, unknowns and prospects. *Brain Stimul.* 9, 323–335

126. Wischnewski, M. and Schutter, D.J.L.G. (2015) Efficacy and time course of theta burst stimulation in healthy humans. *Brain Stimul.* 8, 685–692

127. Chung, S.W. *et al.* (2016) Use of theta-burst stimulation in changing excitability of motor cortex: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 63, 43–64

128. Nitsche, M.A. and Paulus, W. (2001) Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* 57, 1899–1901

129. Fregni, F. *et al.* (2015) Regulatory considerations for the clinical and research use of transcranial direct current stimulation (tDCS): review and recommendations from an expert panel. *Clin. Res. Regul. Aff.* 32, 22–35

130. Liebetanz, D. *et al.* (2002) Pharmacological approach to the mechanisms of transcranial DC-stimulation-induced after-effects of human motor cortex excitability. *Brain* 125, 2238–2247

131. Greenwald, A.G. *et al.* (2003) Understanding and using the implicit association test. I. An improved scoring algorithm. *J. Pers. Soc. Psychol.* 85, 197–216