

RESEARCH ARTICLE

Learning positive social information reduces racial bias as indexed by N400 response

Alessandra Brusa ^{*}, Antonia Pesič, Alice Mado Proverbio

Department of Psychology, Neuro-Mi, Milan Center for Neuroscience, University of Milano-Bicocca, Milan, Italy

^{*} a.brusa@campus.unimib.it

Abstract

The present study used EEG/ERPs to detect the activation of implicit stereotypical representations associated to other-race (OR) people and the modulation of such activation through the previous presentation of positive vs. neutral social information. Electrophysiological signals were recorded in 40 Italian Caucasian participants, unaware of the overall study's purpose. They were presented with 285 sentences that could either violate, non-violate (e.g., "the Roma girl was involved in a robbery) or be neutral with regard to stereotypical concepts concerning other-race people (e.g. Asians, Africans, Arabic). ERPs were time-locked to the terminal words. Prior to the sentence reading task, participants were exposed to a 10 minutes colourful video documentary. While the experimental group was presented a video containing images picturing other-race characters involved in "prestigious" activities that violated stereotypical negative assumptions (e.g. a black neurosurgeon leading a surgery team), the control group viewed a neutral documentary about flora and fauna. EEG signals were then recorded during the sentence reading task to explore whether the previous exposure to the experimental video could modulate the detection of incongruence in the sentences violating stereotypes, as marked by the N400 response. A fictitious task was adopted, consisted in detecting rare animal names. Indeed, only the control group showed a greater N400 response (350–550 ms) to words incongruent with ethnic stereotypes compared to congruent and neutral ones, thus suggesting the presence of a racial bias. No N400 response was found for the experimental group, suggesting a lack of negative expectation for OR individuals. The swLORETA inverse solution, performed on the prejudice-related N400 showed that the Inferior Temporal and the Superior and Middle Frontal Gyri were the strongest N400 intra-cortical sources. Regardless of the experimental manipulation, Congruent terminal words evoked a greater P300 response (500–600 ms) compared to incongruent and neutral ones and a late frontal positivity (650–800 ms) was found to be larger to sentences involving OR than own-race characters (either congruent or incongruent with the prejudice) thus possibly indicating bias-free perceptual in-group/out-group categorization processes. The data showed how it is possible to modulate a pre-existing racial prejudice (as reflected by N400 effect) through exposure to positive media-driven information about OR people. Further follow-up studies should determine the duration in time, and across contexts, of this modulatory effect.

 OPEN ACCESS

Citation: Brusa A, Pesič A, Proverbio AM (2021) Learning positive social information reduces racial bias as indexed by N400 response. PLoS ONE 16(11): e0260540. <https://doi.org/10.1371/journal.pone.0260540>

Editor: Maria Serena Panasiti, Sapienza University of Rome, ITALY

Received: May 10, 2021

Accepted: October 26, 2021

Published: November 24, 2021

Copyright: © 2021 Brusa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available online [10.6084/m9.figshare.14937609](https://doi.org/10.6084/m9.figshare.14937609).

Funding: AB was supported by a Doctoral fellowship from the University of Milano-Bicocca. <https://www.unimib.it> The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Stereotyping is a cognitive process by which a specific set of attributes and beliefs is associated to a social group (e.g., [1]). Stereotypes are embedded within cultural systems and deeply impact how people interact with each other. They are learned and reinforced by numerous processes. They develop early in life, partly through associative learning processes [2], so that even young children easily learn social stereotypes based on features that are salient for categorization such as gender, age, race or physical attractiveness [3]. If on one hand a key role is played by the influence and interactions with parents, friends and peer groups [4, 5], the impact of media in the formation and maintenance of cultural stereotypes and prejudicial feelings toward out-groups seems to be preponderant [6]. Several studies demonstrated that mainstream media has historically marginalized and underrepresented minority groups [7–10]. For instance, researchers have suggested that the negative representation of African Americans and Latinos as criminal, aggressive and unintelligent in the media help reinforce hostile prejudice against these social groups [8]. Moreover, media representations have a strong impact on identity management strategies. Saleem et al. [11] revealed that Muslim American students who viewed negative media representations of their religious group were less likely to desire acceptance by other Americans and more likely to avoid interactions with majority members. Overall, stereotypes are often embedded in the cultural system people are immersed in and are shared within a given culture [12].

Once established, stereotypes become hard to deconstruct and they can influence people's lives in a variety of ways, shaping attitudes (prejudice) and behaviours (discrimination), influencing for instance academic performance, career choices and salary gap.

Researchers have developed several techniques to improve attitudes toward outgroups that have been found to be effective. Kawakami and colleagues [13] found that students who practiced responding in nonstereotypical ways to members of other groups became better able to avoid activating their negative stereotypes on future occasions. Moreover, when people are exposed to or are invited to think about counterstereotypic persons, for instance strong women [14] or Black role models [15], they became less prejudicial towards that social group.

Allport [1] provided one of the first contributions in understanding what are the conditions that allow stereotypes to be modulated and prejudices to be reduced in the publication "The nature of prejudice". He formulated the "Contact Hypothesis", according to which to reduce prejudices and hostilities it is best to encourage contact between members of different groups. Allport additionally identified a series of conditions to be met so that contact can produce positive effects (e.g., Equality of status, Social and institutional support, Pleasant contact and Intergroup cooperation in order to achieve a common goal).

Following this line of thoughts, if on the one hand the media can reinforce certain prejudices and stereotypes, on the other they can be used as educational tool that can raise awareness and influence attitudes towards members of different social groups to reduce intergroup prejudices [9, 10, 16].

For example, Ramasubramanian and colleagues [10] showed that activation of implicit racial stereotypes decreased when people were exposed to news stories disconfirming stereotypes and that there are media-based strategies to reduce accessibility to stereotypes. Also Soble and colleagues [17] showed that a short video could modify the racial attitudes of white college students towards peers of colour. In the experimental design students were assigned to two experimental groups: one group viewed a short documentary depicting the life of a Caucasian and an Afro American subject, followed by hidden cameras over the course of a day, the other group did not view any video. The authors found that the experimental group presented reduced prejudicial attitudes and reduced fear of minorities compared to the control group as

assessed by the *Color-Blind Racial Attitudes Scale* (CoBRAS; [18]), the *Psychosocial Costs of Racism to Whites* scale (PCRW; [19]), and the *Quick Discrimination Index* (QDI; [20]).

Because stereotypes often operate out of awareness, social psychologists and neuroscientists have developed methods for assessing them indirectly. Research using implicit measures suggest that stereotypes and prejudices are easily activated when we encounter members of different social groups [21].

To overcome issues related to social desirability, researchers have effectively studied the implicit activation of stereotypes at the brain level by means of EEG/ERPs. Several studies have shown that ERPs are sensitive to violations of race- and gender-based stereotypes [22–32].

Specific attention has been paid to the N400 response, a negative deflection of the ERPs peaking around 400 ms, that indexes a difficulty in accessing and integrating incoming semantically incongruent information with previous knowledge [33]. Indeed, researchers discovered that lexical-semantic knowledge and general knowledge about the world were both integrated in mental representations in the same time interval during sentence comprehension (about 300 ms from stimulus presentation; [27]). Furthermore, the N400 response was found effective to explore the activation of conceptual/semantic associations related to world knowledge such as stereotypes using both visual or written stimuli [22–24, 27, 29–31, 34–36].

For instance, Hehman and colleagues [22] argued that the N400 component could be considered an index of stereotype accessibility. The authors used a prime-target paradigm and measured a greater N400 reactivity in response to trials in which the prime (face: Black/White) was incongruent with the stereotypical target (trait: positive/negative) than when primes and targets matched. Similarly, Wang et al. [31] found that the N400 amplitude was modulated by the congruent or incongruent association of stereotypical traits to pictures of Chinese rural workers or city dwellers. In the context of social stereotypes the N400 response might therefore index the detection of information not corresponding to available world knowledge, including social attributes stereotypically associate to a social or ethnic group.

In the realm of gender stereotypes, Proverbio and collaborators [23, 24] also used ERPs to measure, by means of a N400 response, the detection of a discrepancy between gender-based occupational stereotypes and non-stereotypical written material. According to the authors [23, 24] using complete and meaningful sentences allowed participants to access and shape more easily mental representations and to form impressions according to their own experiences. Adapting the same paradigm of Proverbio et al. [23] to the study of race-based stereotypes, Brusa and colleagues [34] also found a greater N400 negativity in response to words violating ethnic stereotypes, compared to words that confirmed them or were neutral in this regard.

Moreover, the amplitude of the N400 response was found to be sensitive to several experimental manipulations. For instance, smaller amplitudes were recorded when the incoming information was more expected and therefore easier to process [37].

Since sentence comprehension occurs accounting for one's knowledge, context and expectations deriving from prior knowledge can influence the processing of incoming information from the earliest stages [38]. Indeed, sentence comprehension can be facilitated when the right context is provided. For instance, Hald and colleagues [39] investigated with ERPs the modulation of world knowledge comprehension by providing context. The authors found that when world knowledge incongruities (e.g. *The city Venice has many roundabouts and beautiful buildings*) were preceded by local context that made violations more acceptable (e.g., *The large and increasing amount of cyclists in the inner city of Venice had to be regulated. The city council decided 10 years ago to replace traffic lights with other road layouts that ease traffic flow*) the N400 was reduced compared to when the current discourse context accentuated the violation. The results fostered the claim that local discourse can modulate the integration of world knowledge information retrieved from memory.

Also Duffy and Keir [40] investigated the effect of discourse context on processing of gender stereotypes during reading. After providing adequate disambiguating context, the effect of longer reading times in response to stereotype anomalies disappeared.

The modulation of the N400 effect by providing media-based context information was studied also by Jin and colleagues [41]. The authors explored the role of the media in emphasizing racial stereotypes, as reflected by the N400 effect. In particular, the authors studied the attitude of Han Chinese towards the Uygur minority using ERPs as an index of the cognitive and neural processes associated with the activation of stereotypes. The participants, all of Han ethnicity, were subdivided into two experimental groups: a priming group viewed a video that portrayed negative aspects of the Uygur group; while the control group didn't watch any video.

The paradigm consisted in the vision of unknown faces belonging to either the Han or Uygur ethnicity followed by a lexical decision task of positive or negative words, while recording EEG signal. The results showed that the priming group, who observed the video, unlike the control group, showed a greater N400 amplitude in response to positive adjectives associated with the presentation of Uygur faces compared to when the same faces were associated with negative adjectives.

The results demonstrated that negative media information might influence people judgments toward other groups and this effect is reflected in the deflection of the N400 amplitude.

As seen so far, previous literature has demonstrated that providing local context that is able to modulate expectation on the upcoming information and therefore facilitate the integration of semantic and social incongruities can modulate the N400 effect [38–41]. When considering social stereotypes, only few studies have investigated the role of local context on the N400 (e.g. [41]). Indeed, the study by Jin and colleagues [41] demonstrated that local context provided in the form of media-driven information is able to play an influence on attitudes towards social groups modulating the N400 response. While their study showed that negative media driven information towards a social group can increase the N400 effect when associating the same social group to positive attributes, there aren't studies to our knowledge that have explored whether associating visual positive information to a social group could in turn reduce the N400 effect. Unlike the previously used paradigms, the emphasis here was on the possibility to measure a change in the representation of social attributes, due to the incoming of newly acquired knowledge. In other words, through the N400 paradigm we aimed at investigating whether it was possible to change mental ethnic stereotypes, by measuring their mutable effects, rather than just assessing their presence.

The paradigm used in the present study was similar to the one used by Brusa et al. [34] with sentences that could either violate (incongruent), non-violate (congruent) or be neutral with regard to stereotypical concepts concerning non-Caucasian ethnic groups. While our previous study allowed to implicitly access the neural representation of ethnic stereotypes avoiding the activation of control processes related to social desirability pressure, here the investigation aimed to confirm previous results and further explore, by means of EEG/ERPs technique, whether the previous exposure to positive and progressive social information about other-race members (e.g. Arabs, Africans, Asians, South Americans) in the form of a video documentary could modulate, in a group of Italian Caucasian participants, the access of such stereotypical representation, thus the detection of stereotypes incongruities in the subsequent sentence reading task, as marked by the N400 response.

It was hypothesized that, unlike participants in the control group who were exposed to a neutral video on flora and fauna, the participants exposed to the experimental video representing other-race individuals associated to positive attributes (e.g. a Muslim activist marching for human rights or a Black Boeing's pilot), would present a smaller N400 effect in response to

sentences that violated stereotypical assumptions compared to congruent and neutral sentences.

Based on studies on the influence of context on the N400 effect and the power of media to modulate stereotypical mental representation [9, 16, 17, 39, 41], it was hypothesized that watching a documentary could update, even if only temporarily, one's knowledge about the social world modulating their expectations, as revealed by the N400 effect.

As previously seen in several studies (e.g. [23, 24, 26, 34]), the neural generators of the N400 response elicited by the violation of social stereotypes were hypothesized to reside in cortical structures that are involved in semantic knowledge and in brain regions connected to social cognition processes, such as the representation of social groups (e.g. ethnic groups), impression formation and processing and retrieval of information about other people (e.g. [23, 34, 42–46]). Indeed, the representation of stereotypes seems to be accompanied by the activation of some brain areas involved in mnemonic processes, in particular the temporal lobe (lateral and anterior) and the inferior frontal gyrus, and of areas belonging to the social cognition network, as the middle prefrontal cortex, the temporo/parietal junction, the superior temporal sulcus and the anterior temporal cortex [23, 24, 26, 34, 47–49]. Based on the previous literature we expected that the neural sources best explaining stereotype-dependent N400 potentials would involve regions of the social brain. The literature has especially highlighted the role of the right superior/medial frontal gyrus (SFG; BA 8) in representing social attributes and stereotypes (e.g., [42, 43]). While the typical semantic N400 response (which appears focused at centro/parietal scalp sites) seems to reflect mostly the activity of the anterior medial temporal lobe (AMTL) (e.g., [50, 51]), we hypothesized that the present anterior N400 would tap at social brain areas. Indeed N400 violations were not semantic in nature but purely based on ingroup/outgroup social categorizations. The present findings will get further new lights on the cortical dynamics of social knowledge.

Moreover, as evidenced in our previous study [34], the present study aimed to investigate post-N400 positivities related to congruence and language prediction [52, 53]. Regardless of the group assigned, a P300 effect was expected to be found in response to sentences congruent with ethnic bias compared to incongruent and neutral ones, as already found in the previous study by Brusa and colleagues [34]. The P300 response, an index of contextual updating, has been shown to reflect the integration into mental representations of semantic information that was anticipated during sentence reading [54–58] and would fit the hypothesis that biased mental representations were more easily expected in the group of participants.

Furthermore, also replicating the results of our previous study [34], a late frontal positivity was expected to be found in both groups in response to sentences involving other-race individuals compared to own-race, neutral, characters. Indeed, the late positivity is a component sensitive to the emotional value of a stimulus and is commonly found in response to emotionally relevant and arousing stimuli compared to neutral ones [59–64]. For instance, Cuthbert and colleague [59] revealed that emotionally arousing pictures evoked a larger positive amplitude compared to neutral ones, while Schacht and Sommer [64], investigating the emotional influence on visual word processing, found a late ERP effect related to emotional (positive and negative) words processing compared to neutral ones. It was hypothesized here that sentences referring to OR individuals with either a stereotypical or counter-stereotypical content would be more emotionally relevant (thus evoking a larger frontal positivity) than own-race/neutral sentences.

Materials and methods

Participants

Forty (22 females, 18 males) volunteers aged between 18 and 35 years ($M = 22.75$, $SD = 2.68$) and ranging in education from 12 to 18 years of study ($M = 16.15$, $SD = 1.46$), participated in

the experiment. Participants were randomly assigned to but demographically balanced across two experimental groups. The data from 4 participants were discarded for excessive recording artifacts. The final EEG data analyses therefore included thirty-six ($n = 36$) participants, of whom 18 (age = 22.59, SD = 2) belonging to the experimental group “People” and 18 (age = 22.50, SD = 3.29) to the control group “Nature”. To avoid cultural differences, all volunteers were Italian by birth, Caucasian and with both Italian biological parents. They had normal or corrected-to-normal vision, were right handed as assessed by the Oldfield Inventory ([65]; $M = 0.89$, SD = 0.14) and reported no history of drug abuse or neurological or mental disorders.

Subjects were recruited through the Sona-System University portal and were granted university credits in exchange for their participation.

At the end of the experiment, participants were administered an adapted version of the Pettigrew and Meertens’ Subtle and Blatant Prejudice Scale (see [66] for the Italian translation), to assess the presence of racial bias towards immigrants. Response to the questionnaire were provided on a 5-point Likert scale ranging from ‘Not at all’ to ‘Very much’ according to the agreement with the statements presented.

All the participants expressed their understanding and written consent for the study according to the Declaration of Helsinki (BMJ 1991; 302: 1194), with approval from the University Ethics Committee (prot. n° RM-2019-178). Participants were informed of the general characteristics of the study (which was defined as an ERP study on reading). However, information about the study’s hypotheses were only provided at the end of the experiment, minimizing the social desirability bias.

Stimuli for the video manipulation

Prior to the EEG recording involving the language task, each subject, according to the experimental group assigned, was presented with one of two video documentaries, created *ad hoc* for the study using PowerPoint.

The two documentaries were comparable in the format, differing from each other only for the image content. They were featured the presentation of 95 colourful and copyright free-images, collected from the web. Each video lasted about 10 minutes (9’42”) and was musically accompanied by a romantic piano piece (Debussy -Arabesque N° 1 -Arabesque N° 2 -Clair De Lune).

The documentary presented to the experimental group (named “People of the world”) showed images depicting the lives of multiple subjects belonging to various ethnic groups (e.g. Arabs, Afro-Americans, Roma) portrayed in situations and contexts violating discriminatory stereotypes (for instance, a female Afro-American astronaut at flight controls, a Muslim human rights activist during a protest, two Afro-American Airline pilots, an Arab woman wearing the hijab doing medical research, etc.). This type of “inverted bias” conformed to what was described in the incongruent sentences of the subsequent sentence reading task during EEG recording.

The control group viewed a documentary (named “Nature of the world”) showing images depicting beautiful natural landscapes and different animal species, excluding human beings and their modified landscapes (for instance, a fjord landscape, two rabbits in a meadow, a sea-side landscape, a goat climbing on alpine rocks, desert dunes, etc.).

Procedure for the video manipulation

This manipulation occurred 20 minutes prior EEG recording. The video-documentaries were presented in free vision on a TV monitor placed at a distance of 100 cm, while people were

comfortably seated and in the absence of any EEG recording at this stage. To monitor participants' attention while watching the documentary, they were instructed to watch the documentary attentively, because the content of the documentary would have been object of specific questions in a subsequent memory task, administered at the end of the video presentation. At this purpose, a paper and pencil questionnaire involving the mnemonic recall of some of the viewed content, based on 4 YES/NO questions (Table 1), was employed.

Stimuli for the language task

The linguistic task performed during EEG recording involved the reading of 285 sentences belonging to three experimental conditions based on their congruency in relation to typical ethnic stereotypes present in the Italian culture. The ethnic groups considered were various (Eastern Europeans, Arabs, Africans, Asians, South Americans, Roma) and were balanced between conditions.

The Congruent condition included 95 sentences that confirmed ethnic stereotypes, such as “Naadir is the leader of a group of REBELS” (in Italian “Naadir è a capo di un gruppo di RIBELLI”); The Incongruent condition included 95 sentences that violated stereotypes expressing notions deviating from common prejudice; e.g., “Mustafa is part of a group of ALPINISTS” (in Italian: “Mustafa fa parte di un gruppo di ALPINISTI”); The Neutral condition included 95 sentences that expressed neutral concepts relative to Italian characters; e.g., “Mirco attends a group of ACTORS” (in Italian: “Mirco frequenta un gruppo di ATTORI”).

Sentences were created in a way so that the congruency with the prejudice was made explicit only at the very end by the terminal word.

Moreover, 24 extra sentences sharing the syntactic, semantic and lexical characteristics with the previous sentences acted as target stimuli. They all ended with an animal name as the terminal word (e.g. “That man was assaulted by a TIGER”, in Italian: “Quel signore è stato assalito dalla TIGRE”) for the fictitious experimental task.

The stimuli have been previously employed in another study by the present research group [34] and contextually balanced syntactically, semantically, lexically across conditions and validated in respect to common sense prejudices for the specific population (University students in the Milan metropolitan area).

For ERP averaging of all stimulus types, EEG epochs were time-locked to the terminal words.

Procedure for the language task

Participants were blindly and randomly assigned to either the experimental (“People”) or the control (“Nature”) group by the researchers, preserving demographical balance across groups.

Table 1. Example of questions related to the documentary content administered to participants right after the end of the video (memory test).

#	People Documentary	Nature Documentary
1	Is there an image of a woman throwing a basketball in the documentary? (Yes)	Is there an image of two rabbits in the documentary with their noses close together? (Yes)
2	Is there an image of a policeman in the documentary stopping a car on the street and looking at the driver? (Yes)	Is there an image of a dune in the desert with a palm tree in the foreground in the documentary? (Yes)
3	Is there an image of a woman on the beach cleaning a surfboard in the documentary? (No)	Is there an image of a rattlesnake in attack position in the documentary? (No)
4	Is there an image of a musician in the documentary while playing the saxophone in a club? (No)	Is there an image of a storm that hits the beach in the documentary? (No)

<https://doi.org/10.1371/journal.pone.0260540.t001>

Regardless of the group assigned, participants underwent the identical procedure that differed only for the documentary previously presented. After positioning the EEG headset, participants were invited to seat in an electrically and acoustically shielded cubicle at 100 cm from a PC screen and were asked to fixate its centre where a dot served as fixation point.

The experimental procedure involved two distinct subsequent sessions: the viewing of the documentary followed by the mnemonic recall questionnaire (without EEG recording) and the sentence reading task (with EEG recording).

In the first session, participants were provided written instructions that invited them to attentively watch the video documentary and later to answer a questionnaire. The documentary was presented on a PC screen using *Quicktime Player* and background music was reproduced through a pair of Sennheiser (202HD model) headphones. At the end of the presentation, the brief mnemonic questionnaire was provided (Table 1) and answers were collected to assess the attentive state of participants.

In the second session, where EEG signal was recorded during the sentence reading task, participants were instructed to sit relaxed but still, avoiding any head or body movements as well as ocular saccades and blinks.

Sentences were presented with *Evoke v2.2* (ANT Neuro, Hengelo, The Netherlands) in a randomized order across 8 sequences of 2 min and 40 s each interspersed with a 30-s pause. Each sentence was presented for 1000 ms at the centre of the screen arranged in a maximum of three short rows and followed, after an ISI of 700 ms, by the terminal word which was presented for 1000 ms in uppercase. The ITI was 1200 ms. The average size of terminal words was 4.47 cm x 0.8cm, implying a visual angle of $2^{\circ}33'40'' \times 2^{\circ}33'40''$ (minimum length = 1.5cm, maximum length = 7.9 cm). The text was printed in yellow on a grey background. An example is shown in Fig 1.

The fictitious task consisted of responding as quickly and accurately as possible to the terminal word of sentences when it contained an animal name, by pressing a joystick key with the index finger of the right or the left hand. The response hand was alternated across trials. Participants were unaware of the overall study purposes and no judgement about stereotypes was required. Response time to target terminal words was recorded only to control the attentive state and accuracy of participants.

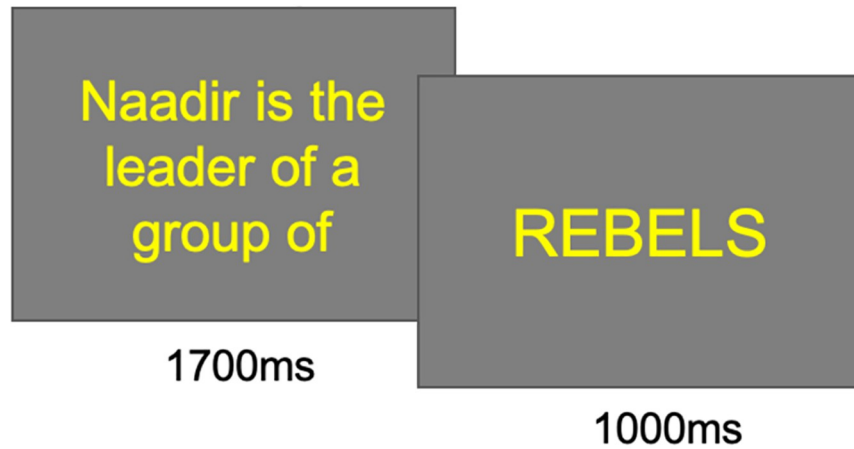
Two training runs were administered to participants to allow them to familiarize with the procedure. A longer pause was allowed at about half the recording time. Each experimental sequence started with the presentation of 3 warning signals of 700 ms of duration (“Ready, Set, Go!”) and ended with “thank you” word on the screen. Both the warnings and thanks were typed in uppercase characters.

At the end of the second session, participants were administered the Subtle and Blatant Prejudice Scale.

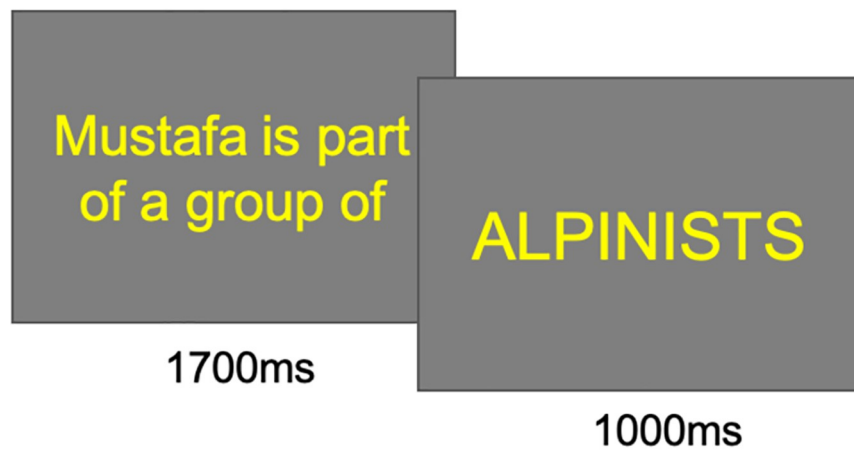
EEG recording and analysis

EEG data were recorded using a standard EEG cap with 128 electrodes located according to the 10–5 International System [67] using *EEProbe v2.2* software (ANT Neuro, Hengelo, The Netherlands) at a sampling rate of 512 Hz (band-pass 0.016–70 Hz). Horizontal (hEOG) and vertical (vEOG) eye movements were also recorded and linked mastoids served as the reference lead. Electrode impedance was maintained below 5 KOhm. Computerized artefact rejection and manual eye-inspection was performed to remove EEG segments contaminated by ocular artefacts (saccades and blinks), muscle-related potentials, or amplifier blockages. The computerized artefact rejection criterion was a peak-to-peak amplitude exceeding 50 μ V. The rejection rate was ~18% in both groups and the average number of trials per condition that went into ERP averaging was 77.58 for congruent, 78.47 for incongruent and 77.86 for neutral.

Congruent



Incongruent



Neutral

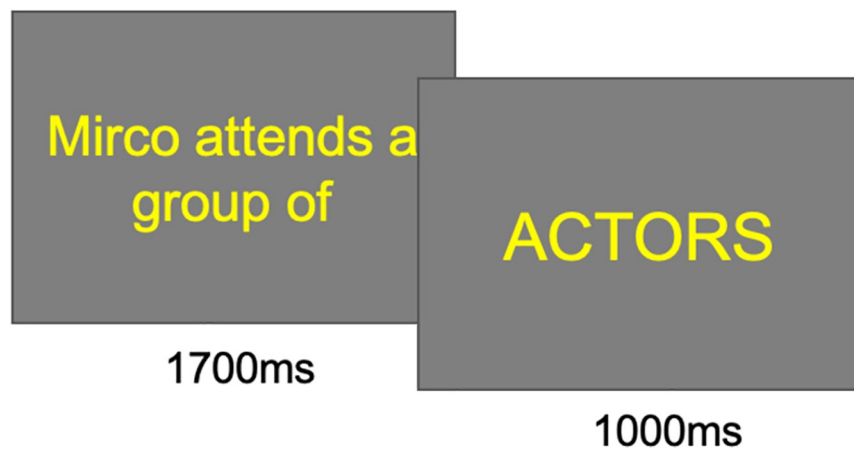


Fig 1. Experimental paradigm. Example of Congruent, Incongruent and Neutral sentences used in the language task. ERPs were time-locked to terminal words which were presented in uppercase for 1000 ms.

<https://doi.org/10.1371/journal.pone.0260540.g001>

EEG epochs were synchronized with the onset of the terminal word. Evoked response potentials (ERPs) were averaged off-line from 100 ms before to 1000 ms after stimulus onset and an off-line filter (band-pass 0.16–30 Hz) was applied to the ERPs.

The mean amplitude area of the N400 response was recorded from anterior, frontal and fronto-central sites (AFz, Fz, FCz) in the 350- to 550-ms temporal window. The mean amplitude area of the P300 was recorded from anterior and fronto-central sites (AFF1, AFF2, AFz, Fz, FCz) in the 500–600-ms time window. The mean amplitude area of the frontal positivity was recorded from anterior-frontal sites (AF3, AF4, AFF1, AFF2, AFz) in the 650–800-ms time window. ERP analyses were performed using *EEprobe v2.2* software (ANT Neuro, Hengelo, The Netherlands). The electrode clusters and time windows were selected on the basis of previous literature [22–24, 29–31, 34, 68] and on where they matched the strongest scalp recorded activity of N400, P300 and frontal positivity potentials.

For each ERP component, the mean area amplitude values were subjected to repeated-measures ANOVAs whose factors of variability were group (“People”, “Nature”), condition (congruent, incongruent, neutral) and electrode (depending on the component of interest). Fisher post-hoc comparisons were used. The effect size for the statistically significant factors was estimated using partial eta squared (η_p^2) and the Greenhouse-Geisser correction was applied to account for non-sphericity of the data. All the ANOVAs were performed using Statistica software (version 10) by StatSoft. Behavioural data from the fictitious task and the mnemonic questionnaire were not statistically analysed, but response time and accuracy were collected as a record.

A standardized weighted low-resolution electromagnetic tomography (swLORETA; [69, 70]) was performed on the ERP Difference Wave relative to N400 potentials, computed by subtracting ERP response to Congruent words from ERP response to Incongruent words in the N400 time window, using ASA4 Software.

Results

Memory task

Behavioural data showed that the mnemonic recall of documentary contents was comparable in the two groups. On a total of four questions, the average number of correct responses was 3.61 (SD = 0.70) for the “People” group and 3.55 (SD = 0.62) for the “Nature” group, indicating a good attentive state at the documentary.

Language task

Participants had an average Response Time (RT) to target terminal words of 625.48 ms (right hand = 627.73 ms; left hand = 625.52 ms). Moreover, the accuracy rate of responses was quite high (hits = 98%), suggesting that participants maintained an elevated attentive state throughout the experiment.

Subtle and Blatant Prejudice Scale

The prejudice scores towards other race individuals could range from 1 (minimum prejudice) to 5 (maximum prejudice). A single overall prejudice score was computed for the two groups separately averaging all the items of the Subtle and Blatant Prejudice Scale. The mean scores were comparable in the two groups as assessed by the independent-sample t-test ($t(34) = 0.40$,

Table 2. Summary of the results of the T-test comparisons on the scores obtained at the Subtle and Blatant Prejudice Scale across the two experimental groups.

Prejudice	Group	Descriptives	T Statistic
Overall score	People	M = 1.99, SD = 0.35	t (34) = 0.40, p = .694
	Nature	M = 2.04, SD = 0.44	
Blatant Scale	People	M = 1.58, SD = 0.33	t (34) = 0.60, p = .552
	Nature	M = 1.58, SD = 0.36	
Subtle Scale	People	M = 2.39, SD = 0.45	t (34) = 0.00, p = 1.000
	Nature	M = 2.49, SD = 0.59	
Group	Prejudice	Descriptives	T Statistic
People	Blatant	M = 1.58, SD = 0.33	t (17) = -9.44, p < .001
	Subtle	M = 2.39, SD = 0.45	
Nature	Blatant	M = 1.58, SD = 0.36	t (17) = -9.33, p < .001
	Subtle	M = 2.49, SD = 0.59	

M = mean; SD = standard deviation; t = T-test.

<https://doi.org/10.1371/journal.pone.0260540.t002>

p = 0.33) and it was 1.99 (SD = 0.35) in the “People” group and 2.04 (SD = 0.44) in the “Nature” group.

Moreover, in both groups the average score was significantly lower than the scale midpoint [3] (People: $t(17) = -12.31$, $p < .0001$; Nature: $t(17) = -9.17$, $p < .0001$), suggesting an overall low racial prejudice towards immigrants in the experimental population, regardless of the group assigned.

Considering the explicit and implicit (Blatant and Subtle) scales separately, participants in the “People” group obtained an explicit prejudice score of 1.58 (SD = 0.33) and implicit prejudice score of 2.39 (SD = 0.45), while the “Nature” group scored 1.58 (SD = 0.36) at the explicit prejudice score and 2.49 (SD = 0.59) at the implicit prejudice one. In both groups the explicit prejudice was significantly lower than the implicit prejudice (People: $t(17) = -9.44$, $p < .001$; Nature: $t(17) = -9.33$, $p < .001$). A summary of results is presented in Table 2 and Fig 2.

N400 response (350–550 ms)

The ANOVA performed on the N400 amplitude values recorded over the anterior, frontal and fronto-central midline sites (AFz, Fz and FCz) in the 350–550 ms time window, revealed a significant effect of the interaction between group (“People” and “Nature”) and condition (Congruent, Incongruent, Neutral) ($F(2, 68) = 3.22$, $p < 0.05$, $\eta_p^2 = 0.09$). Pairwise comparisons revealed that in the “Nature” group a greater negativity ($p = 0.037$) was recorded in response to Incongruent (Nature: 2.11 μV , SE = 0.68; People: 1.12 μV , SE = 0.68) than Congruent (Nature: 2.85 μV , SE = 0.76; People: 0.93 μV , SE = 0.76) and also ($p = 0.01$) compared to Neutral (Nature: 3.02 μV , SE = 0.78; People: 0.85 μV , SE = 0.78) terminal words, as visible in Figs 3 and 4. A similar effect has not been recorded in the “People” group (respectively $p = 0.588$ and $p = 0.445$).

P300 response (500–600 ms)

The ANOVA performed on the P300 amplitude values, measured between 500 and 600 ms over the anterior-frontal sites (AFF1, AFF2, AFZ, FZ and FCZ) showed a borderline significant main effect of condition ($F(2, 68) = 3.14$, $p = 0.05$; $\eta_p^2 = 0.08$), indicating that the P300 amplitude recorded in response to Congruent terminal words (4.18 μV , SE = 0.51) was overall greater ($p = 0.015$) than the one measured in response to Neutral terminal words (3.41 μV , SE = 0.53). Moreover, the analysis showed a significant interaction between condition and

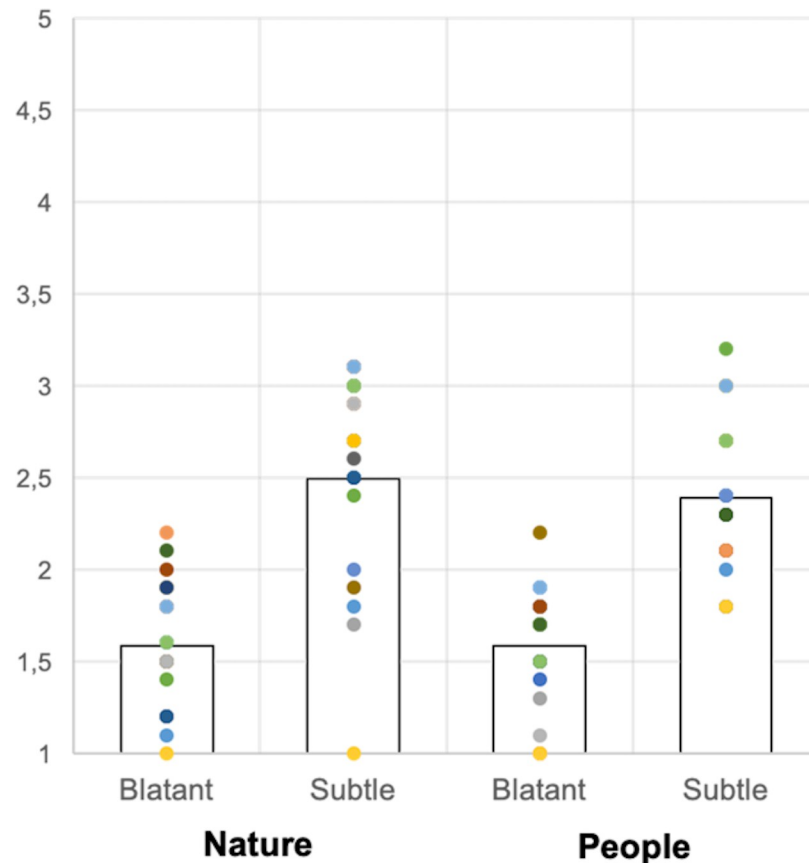


Fig 2. Subtle and Blatant Prejudice Scale. Scores obtained on the Subtle and Blatant Prejudice Scale by the two experimental groups. Scores could range from 1 (minimum prejudice) to 5 (maximum prejudice); for both groups, explicit prejudice (blatant) was significantly lower than the implicit one (subtle).

<https://doi.org/10.1371/journal.pone.0260540.g002>

electrode ($F(8, 272) = 4.03, p < 0.005; \eta_p^2 = 0.11$), regardless of the experimental manipulation. Overall, the P300 amplitude was greater ($p < 0.001$) in response to Congruent than Incongruent and Neutral terminal words at all electrode sites; in turn the P300 response was larger ($p < 0.001$) to Incongruent than neutral terminal words at more anterior but not FCZ site ($p = 1$). P300 mean amplitude values are reported in Table 3 and showed in Figs 4 and 5.

Finally, the analysis revealed a significant effect of the electrode ($F(4, 136) = 7.17, p < 0.001, \eta_p^2 = 0.17$), indicating that the P300 response was larger over the electrode FCZ ($4.16 \mu\text{V}, \text{SE} = 0.50$) compared to the electrodes AFF1 ($3.59 \mu\text{V}, \text{SE} = 0.47; p = 0.001$), AFF2 ($3.74 \mu\text{V}, \text{SE} = 0.46; p = 0.040$) and AFZ ($3.44 \mu\text{V}, \text{SE} = 0.44; p < 0.001$). The amplitude recorded over the electrode FZ ($3.94 \mu\text{V}, \text{SE} = 0.51$) was greater than the one recorded over the electrode AFZ ($p = 0.008$).

Frontal positivity (650–800 ms)

The ANOVA computed on the mean area values of frontal positivity, measured between 650 and 800 ms over frontal-anterior regions (AF3, AF4, AFF1, AFF2, AFZ) didn't reveal any difference across the two experimental groups. It showed a significant effect of condition ($F(2, 68) = 4.57, p < 0.05, \eta_p^2 = 0.17$) such that the positivity measured in response to stimuli regarding OR individuals, either Congruent ($2.61 \mu\text{V}, \text{SE} = 0.37; p = 0.026$) or Incongruent ($2.77 \mu\text{V}, \text{SE} = 0.36; p = 0.006$), was greater than the frontal positivity recorded in response to stimuli regarding own-race individuals, that is neutral words ($1.97 \mu\text{V}, \text{SE} = 0.40$).

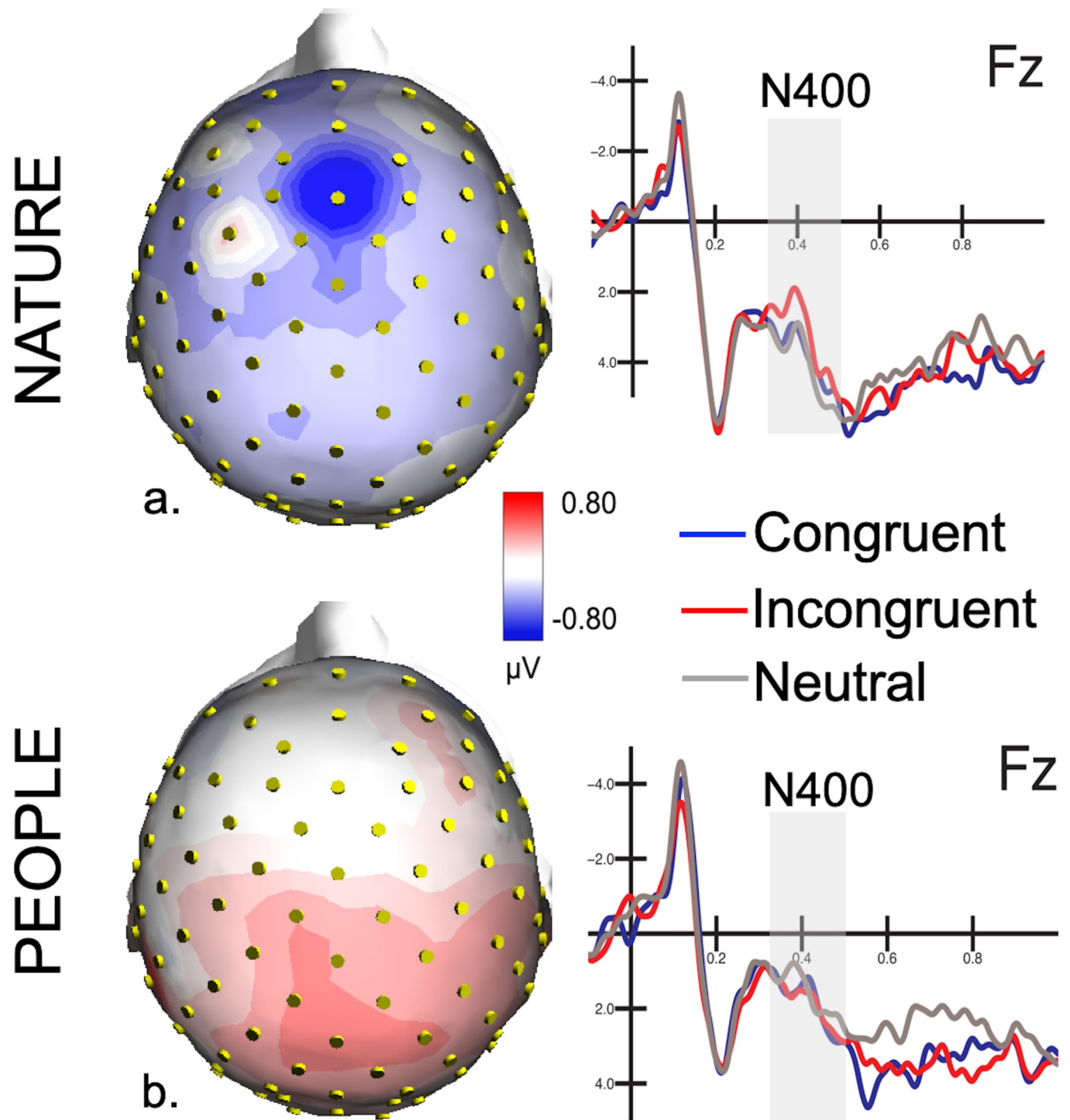


Fig 3. N400 topographical map. Topographical distribution of the differential voltage (Incongruent–Congruent) recorded between 350 and 550 ms post-stimulus and ERP waveforms recorded at frontal sites in response to Congruent, Incongruent and Neutral sentences; The top figure (a) shows results in the “Nature” group; the bottom figure (b) in the “People” group.

<https://doi.org/10.1371/journal.pone.0260540.g003>

The analysis also revealed a significant effect of electrode ($F(4, 136) = 7.30, p < 0.01, \eta_p^2 = 0.18$), indicating greater positivity recorded over the electrodes AFF1 (2.75 μV , SE = 0.39), AFF2 (2.91 μV , SE = 0.38) and AFZ (2.60 μV , SE = 0.36) compared to the electrodes AF3

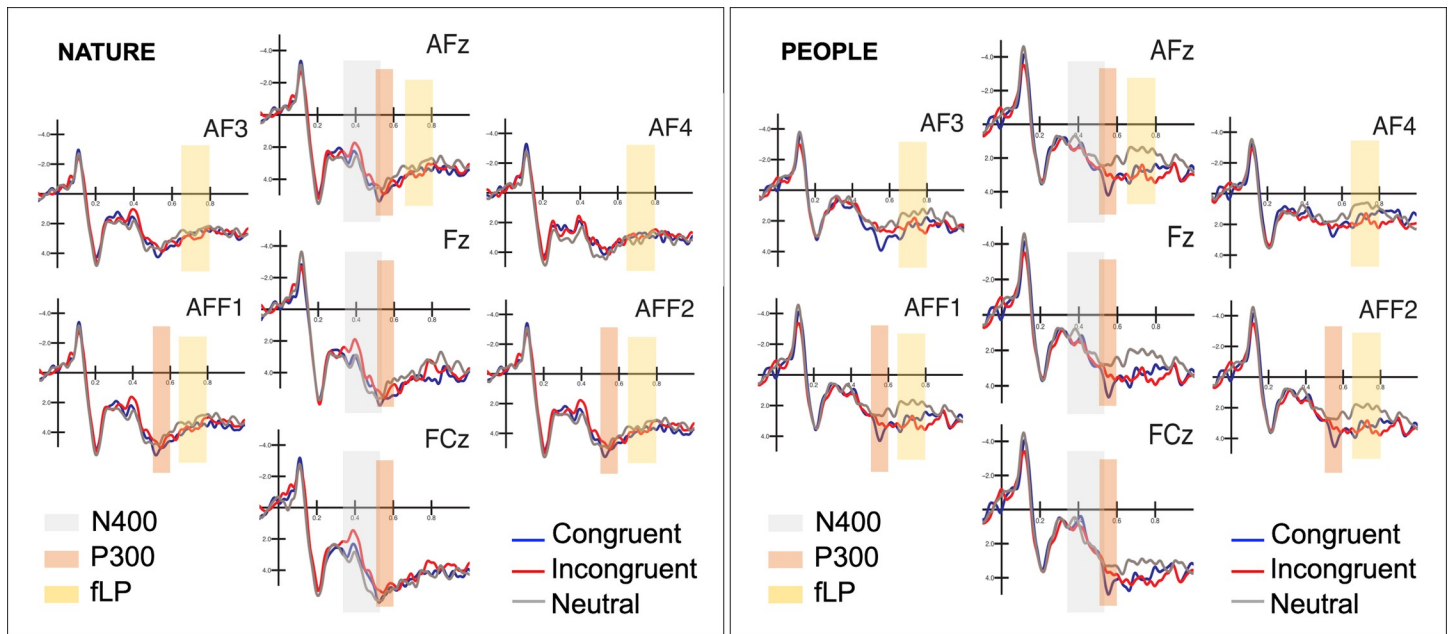


Fig 4. Grand average ERP waveforms recorded at anterior frontal sites in response to congruent, incongruent and neutral sentences. The time windows corresponding to the N400 (350–550 ms), P300 (500–600 ms) and frontal positivity (fLP; 650–800 ms) are highlighted. On the left recordings from the “Nature” group; on the right from the “People” group.

<https://doi.org/10.1371/journal.pone.0260540.g004>

(2.09 μ V, SE = 0.37; respectively p = 0.004; p <0.001; p = 0.027) and AF4 (1.90 μ V, SE = 0.35; respectively p < 0.001; p <0.001; p = 0.003). Results are showed in Figs 4 and 5.

swLORETA source reconstruction

The swLORETA inverse solution, applied to the Difference Waves obtained by subtracting ERP waveforms elicited by Congruent from that elicited by Incongruent stimuli in the

Table 3. P300 mean amplitude values recorded in the 500–600 ms time window in response to congruent, incongruent, and neutral terminal words, at anterior and front-central sites (AFF1, AFF2, AFZ, FCZ, FZ).

		P300 (500–600 ms)					
CONDITION	ELECTRODE	M	S.E.	C.I. -95%	C.I. 95%	N	
Congruent	AFF1	4.02	0.53	2.95	5.09	36	
Congruent	AFF2	4.18	0.52	3.13	5.23	36	
Congruent	AFZ	3.89	0.50	2.88	4.90	36	
Congruent	FCZ	4.49	0.51	3.45	5.52	36	
Congruent	FZ	3.34	0.56	3.21	5.47	36	
Incongruent	AFF1	3.57	0.45	2.64	4.49	36	
Incongruent	AFF2	3.71	0.45	2.79	4.63	36	
Incongruent	AFZ	3.44	0.43	2.56	4.32	36	
Incongruent	FCZ	3.99	0.51	2.96	5.03	36	
Incongruent	FZ	3.95	0.51	2.90	4.99	36	
Neutral	AFF1	3.19	0.54	2.09	4.30	36	
Neutral	AFF2	3.32	0.52	2.27	4.38	36	
Neutral	AFZ	2.99	0.50	1.97	4.01	36	
Neutral	FCZ	3.99	0.58	2.82	5.17	36	
Neutral	FZ	3.53	0.57	2.37	4.69	36	

M = mean; S.E. = Standard Error; C.I. = Confidence Interval; N = sample size.

<https://doi.org/10.1371/journal.pone.0260540.t003>

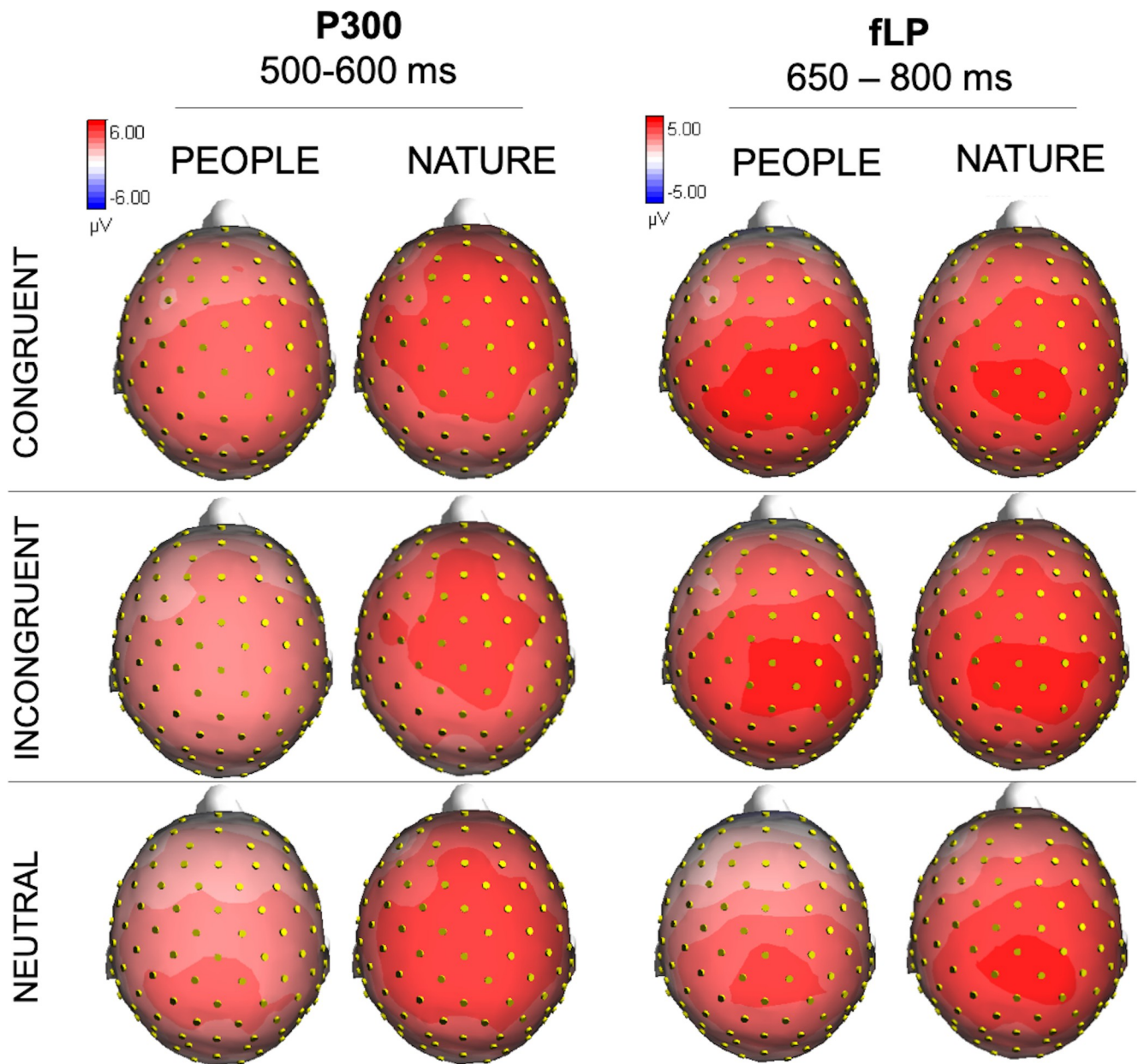


Fig 5. P300 and fLP topographical maps. Isocolour topographical maps of the voltage distribution (top view) of brain potentials in the time windows corresponding to the P300 (500–600 ms) and frontal late positivity (fLP; 650–800 ms).

<https://doi.org/10.1371/journal.pone.0260540.g005>

350–550 ms time window, corresponding to N400 component, showed (Fig 6) that the most active inner sources were the right superior frontal gyrus (SFG; BA 8) and the inferior temporal gyrus (ITG; BA 20). A complete list of all significant electromagnetic dipoles explaining the difference voltages are listed in Table 4.

Discussion

The present investigation aimed to explore, by means of EEG/ERPs technique, whether the exposure to a video documentary representing positive and counter-stereotypical social information about other-race members could modulate, in a group of Italian Caucasian

participants, the activation of implicit stereotypical representations associated to other-race (OR) people in a language task, as indexed by the N400 response.

The pool of participants of the study, all with Caucasian background, overtly presented a low prejudice towards OR people as assessed by the Prejudice Scale. Such overall low racial bias presented differences according to the explicitness of the scale used; indeed, the implicit scale revealed slightly higher prejudice scores compared to the explicit one. This result is in line with research that posits that implicit and explicit attitudes can be dissociated [71, 72]. While explicit prejudice shapes deliberative responses which people are able to monitor, implicit attitudes are automatically activated and usually influence responses without a person's full awareness or control [73, 74]. Therefore, these results highlight the presence, although little and below the scale midpoint, in our participants' group of an implicit racial bias that probably bypasses explicit learnt equalitarian values.

Participants were randomly assigned to two experimental groups and accordingly presented one of two video documentaries for the experimental manipulation.

As previously hypothesized, the previous observation of the experimental video documentary was associated to a smaller N400 response, as compared to the control video. Indeed, the results showed that the "People" experimental group, which observed several representations of OR individuals associated to positive social information, presented a smaller N400 effect in response to sentences that violated common racial stereotypes compared to the "Nature" control group.

On the other hand, similarly to our previous study, showing how the N400 response signalled the existence of implicit ethnic prejudices [34], here the N400 response was larger to terminal words that were incongruent with racial stereotypes (compared to congruent and neutral ones) in the control group only witnessing a nature documentary. Therefore, the "Nature" control group, which viewed the neutral video-documentary representing landscapes and animals, showed the typical N400 effect representing the pre-existing representation of race-biased social attributes (prejudices), as it was observed in the previous study [34]. Conversely, the "People" group who observed unbiased social information (progressive documentary) presented a lower N400 amplitude.

Although a possible limit of the present study might rely in the absence of a pre-intervention measure of the N400 effect to assert that the reduced N400 amplitude in the experimental group it is truly to ascribe to the experimental manipulation, based on previous literature [23, 24, 34, 41], it is reasonable to believe that being exposed to the un-biased and counter-stereotypical video documentary played a crucial role in the reduced N400 effect.

As well-known the N400 response is an ERP component sensible to the detection of semantic and conceptual incongruities; it reflects the integration of incoming information with pre-existing knowledge and it appears larger when such integration requires an effort due to unexpected content [33]. Several studies showed that when people are faced with information (in the form of images or sentences) that are incongruent with previous conceptual/semantic associations related to world knowledge, such as counter-stereotypical social knowledge, this incongruence is reflected in a greater N400 response [22–24, 27, 29–31, 34–36]. This allowed researchers to assess the presence of implicit stereotypical mental representations in the population, revealing the sensitivity of ERPs as an implicit measure.

The present study further investigated whether these deeply grounded mental associations in the form of stereotypes could be modulated and its results are in line with the literature that highlights the role of contextual information in modulating the N400 response. Providing local context can indeed modulate expectations and facilitate the integration of incongruent content into mental representations, thus reducing the N400 effect [39–41]. Duffy and Keir [40] have shown that offering adequate context to introduce information that are incongruent

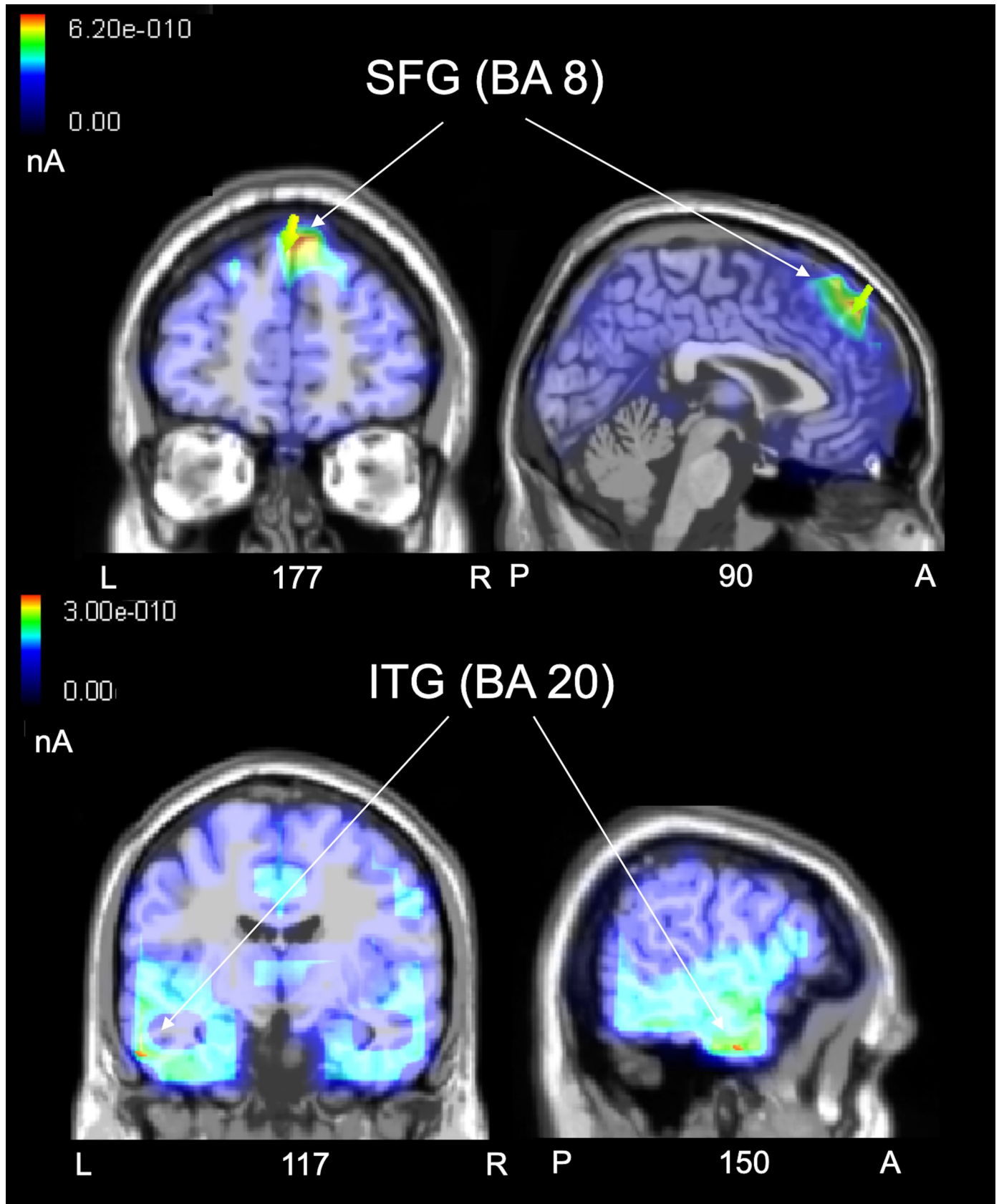


Fig 6. swLORETA. The most active sources relative to the ERP difference-waves (incongruent minus congruent) recorded in the 350–550 time-window (N400) in the “Nature” group were the Superior Frontal Gyrus (SFG) and the Inferior Temporal Gyrus (ITG).

<https://doi.org/10.1371/journal.pone.0260540.g006>

with gender stereotypes modulated the N400 in response to stereotype violations, possibly facilitating their integration and reducing biased expectations. On the other hand, Jin and colleagues [41], highlighting the role of media in shaping social attitudes, proved that exposing people to videos that show negative behaviours associated to a minority group (e.g. Uyghurs) can increase biased attitudes towards the same group as showed by a greater N400 effect in response to positive trait (compared to negative trait) associations to that group. The present study explored whether media exposure could also positively shape attitudes towards ethnic minorities, attenuating social bias, using the N400 response as a neural marker of this modulation. Abundant behavioural literature confirmed on one hand the impact of media in the formation and maintenance of stereotypes and prejudicial feelings toward out-groups [6–9] and on the other hand its ability to disconfirm stereotypes and modify racial attitudes [10, 17].

According to our findings, relying on the influence of context on the N400 effect and the power of media to modulate stereotypical mental representation [9, 16, 17, 39, 41, 75], we hypothesize that watching the experimental documentary possibly updated participants’ mental representations modulating their expectations, as revealed at a neural level by the reduced N400 effect. While the effect observed was only a between-person effect and it was assessed after a short time (about 20 minutes) occurred between the video manipulation and the language task, future studies should approach the matter applying a within-person effect and eventually enquire the actual duration of such effect in the long-term memory.

Moreover, similarly to our previous ERP studies on ethnic and sexual prejudices [23, 24, 34], the swLORETA computed to reconstruct the neural generators of the N400 response elicited by the violation of ethnic stereotypes in the control group “Nature” identified a networks of cortical structures involved in semantic knowledge and social cognition processes such as the inferior temporal gyrus (ITG; BA 20) and right superior frontal gyrus (SFG; BA 8).

This pattern of results fits well with available social neuroscientific literature that posits, indeed, that stereotypical associations reside in regions that underpin semantic memory, such as the anterior and lateral temporal lobes [44, 76–78]. Moreover, in line with our results, Olson and colleagues [49] proposed that the anterior portion of the inferior temporal gyrus (the anterior temporal lobe, ATL) plays a critical role in representing and retrieving social knowledge, including biographic memory about people and more generally memory for traits and social concepts. This, according to several authors, includes stereotype representations, considered to be social semantic associations [23, 24, 44, 46, 79].

Table 4. Source reconstruction of N400 response to stereotype violation.

Magn.	T-x	T-y	T-z	Hem.	Lobe	Gyrus	BA
6.42	1.5	40.5	50.7	R	F	Superior Frontal	8
2.62	60.6	-45.8	-9.5	R	T	Middle Temporal	37
2.62	-28.5	-90.3	20.8	L	O	Middle Occipital	19
2.50	-58.5	-8.7	-21.5	L	T	Inferior Temporal	20
2.16	40.9	-86.4	-12.4	R	O	Inferior Occipital	18

The table provides Talairach coordinates and localization of the active electromagnetic dipoles explaining the surface difference-voltage (incongruent–congruent) recorded in the 350–550 ms time window, according to swLORETA inverse solution.

Magn. = Magnitude; Hem = Hemisphere; BA = Brodmann areas; T = Talairach; R = right; L = left.

<https://doi.org/10.1371/journal.pone.0260540.t004>

Another powerful source in terms of magnitude was the superior frontal gyrus (BA 8), a region often implicated in the representation of social attributes and theory of mind processes (TOM; [43, 80–82]). The medial prefrontal cortex (mPFC) has been found together with other regions belonging to the social cognition network to be a particularly important structure for the processing of social information (see [83, 84]). Its activity has been primarily associated with impression formation about other people and mentalizing processes [85]. Interestingly, in a recent study on the neural basis of prejudice, it was also found that the left superior frontal cortex was particularly involved in representing negative prejudices related to others [82].

Another interesting piece of data from the present study revealed, regardless of the group assigned, a greater P300 effect in response to sentences congruent with ethnic bias compared to incongruent and neutral ones, as previously seen also in the study by Brusa and colleagues [34]. The P300 response is an index of contextual updating and has been shown to reflect the integration into mental representations of semantic information that was pre-activated during sentence reading [54–58]. In line with previous literature, it appears that in the present study Congruent sentences matched participants' previous knowledge and mental stereotypical representations, thus evoking a greater P300 response.

Similarly, a study by Correll and co-authors [29] investigated the properties of the P300 component in a racial bias experiment and demonstrated a greater P300 response in the presence of black armed subjects (compared to White armed ones) due to the automatic association of Black persons and the use of weapons. Here, both the experimental and the control group, showed a greater P300 in response to sentences that matched racial stereotypes compared to the other conditions. It is possible that, while learning through video documentary was able to facilitate the integration of upcoming stereotype-inconsistent content, hence reducing the N400 effect in response to incongruent stimuli, the P300 effect might possibly reflect that in both groups, regardless of the experimental video-manipulation, a terminal word confirmative of the stereotype was pre-activated, thus more expected, during sentence reading.

Furthermore, also in line with the results of our previous study on the measuring of implicit stereotypical mental representation [34], a late frontal positivity was found in both groups to be sensitive to the ethnicity of the characters involved in the sentences. Indeed, a greater frontal positivity was recorded in response to sentences involving other-race individuals, both in a stereotypical and counter-stereotypical way, compared to own-race neutral ones. It is possible to assume that stereotype relevant content involving OR persons might represent for the reader a more emotionally relevant and arousing material compared to neutral content involving individual of the same ethnicity. This finding fits previous literature which suggests that the late positivity is a component sensitive to the emotional and motivational value of a stimulus. It was found in response to emotional and arousing pictures, both positive and negative, [59, 60] or words [61–64] compared to neutral ones. The arousal hypothesis fits with neuroimaging literature reporting how perception of other- vs. own-race people increases the activity of amygdala, a subcortical structure that reflects arousal triggered by fast unconscious assessment of potential threat [86–88]. Moreover, in a study by Ito and colleagues [89] on social perception processes a late positive potential marked the evaluative differentiation of racial ingroup and outgroup members. Indeed, several studies have investigated the neural processes underlying ingroup-outgroup discrimination, mainly using face stimuli. For instance, Hart and colleagues [90] showed in an fMRI study that amygdala responses to human face stimuli were affected by the relationship between the perceived ethnicity of the facial stimulus and that of the observer. Moreover, greater fusiform activity has been found in response to faces of one's own racial group [91]. ERP research examining the N170 component, an index of face encoding, revealed differential processing of ingroup versus outgroup faces, even when groups

were defined arbitrarily [92]. Also, the orbitofrontal cortex has been shown to play a broader role in group-based evaluations [93].

In the present study, no significant difference was found between the two experimental groups (lack or presence of bias-free social documentary administration) in the amplitude of frontal positivity. Nonetheless, as visible in Fig 4, the effect appears stronger in the “People” experimental group. It is possible that viewing the experimental documentary depicting hundreds of individuals of OR ethnic groups might influence and increase the ingroup-outgroup discrimination. This claim isn’t supported by precise statistical data but deserves further analyses for reaching a specific conclusion. Overall, it is reasonable that adopting a larger sample size might overcome the possible limitations and strengthen the power of the present research inherent to the relatively small sample size.

Conclusions

The present study revealed the possibility that N400 effect reflecting pre-existing racial stereotypes, might be modulated and reduced through exposure to positive media-driven information about other race (OR) people. Here, the modulatory effect on the N400 response was tested about 20–40 minutes after the experimental manipulation, so that further follow-up studies should determine its duration in time by testing it after longer periods of time, such as one week or after one month. Moreover, the present results seem to point to a sovramodal interaction in social knowledge, such that a pictorial/filmic experience could modify event related potentials in a stereotype-based language task. Hence, these data possibly suggest that the N400 response may be a sovramodal marker for conceptual incongruence (for example, [36, 94–96]) and that social knowledge is also sovramodal in nature.

Acknowledgments

The authors wish to thank Giorgia Bordone and Francesco De Benedetto for their kind support.

Author Contributions

Conceptualization: Alessandra Brusa, Alice Mado Proverbio.

Formal analysis: Alessandra Brusa, Antonia Pesič.

Investigation: Alessandra Brusa, Antonia Pesič.

Methodology: Alice Mado Proverbio.

Supervision: Alice Mado Proverbio.

Writing – original draft: Alessandra Brusa.

Writing – review & editing: Alice Mado Proverbio.

References

1. Allport G. W., Clark K., Pettigrew T. The nature of prejudice. 1954. <https://doi.org/10.1037/h0062678> PMID: [13204492](https://pubmed.ncbi.nlm.nih.gov/13204492/)
2. Rudman L. A., Phelan J. E., Heppen J. B. Developmental sources of implicit attitudes. *Pers Soc Psychol Bull.* 2007; 33(12):1700–1713. <https://doi.org/10.1177/0146167207307487> PMID: [18000104](https://pubmed.ncbi.nlm.nih.gov/18000104/)
3. Bigler R. S., Liben L. S A developmental intergroup theory of social stereotypes and prejudice. *Adv Child Dev Behav.* 2006; 34:39–89. [https://doi.org/10.1016/s0065-2407\(06\)80004-2](https://doi.org/10.1016/s0065-2407(06)80004-2) PMID: [17120802](https://pubmed.ncbi.nlm.nih.gov/17120802/)
4. Aboud F. E., Doyle A. B. Parental and peer influences on children’s racial attitudes. *Int J Intercult Relat.* 1996; 20(3–4):371–383.

5. Schaller M., Conway III L. G. Influence of impression-management goals on the emerging contents of group stereotypes: Support for a social-evolutionary process. *Pers Soc Psychol Bull.* 1999; 25(7):819–833.
6. Brown R. *Prejudice: Its social psychology.* Cambridge, MA: Blackwell. 1995.
7. Dill K. E., Thill K. P. Video game characters and the socialization of gender roles: Young people's perceptions mirror sexist media depictions. *Sex roles.* 2007; 57(11):851–864.
8. Dixon T. L., Linz D. Race and the misrepresentation of victimization on local television news. *Commun Res.* 2000; 27(5):547–573.
9. Ramasubramanian S., Oliver M. B. Activating and suppressing hostile and benevolent racism: Evidence for comparative media stereotyping. *Media psychology.* 2007; 9(3):623–646.
10. Ramasubramanian S., Murphy C. J. Experimental studies of media stereotyping effects. In: Webster Murray, Sell Jane, editor. *Laboratory Experiments in the Social Sciences.* Second Edition. Academic Press. 2014. p. 385–402.
11. Saleem M., Ramasubramanian S. Muslim Americans' responses to social identity threats: Effects of media representations and experiences of discrimination. *Media Psychology.* 2019; 22(3):373–393.
12. Madon S., Guyll M., Aboufadel K., Montiel E., Smith A., Palumbo P., et al. Ethnic and national stereotypes: The Princeton trilogy revisited and revised. *Pers Soc Psychol Bull.* 2001; 27(8):996–1010.
13. Kawakami K., Dovidio J. F., Moll J., Hermsen S., Russin A. Just say no (to stereotyping): effects of training in the negation of stereotypic associations on stereotype activation. *J Pers Soc Psychol.* 2000; 78(5):871. <https://doi.org/10.1037//0022-3514.78.5.871> PMID: 10821195
14. Blair I. V., Ma J. E., Lenton A. P. Imagining stereotypes away: the moderation of implicit stereotypes through mental imagery. *J Pers Soc Psychol.* 2001; 81(5):828. <https://doi.org/10.1037//0022-3514.81.5.828> PMID: 11708560
15. Bodenhausen G. V., Schwarz N., Bless H., Wänke M. Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals. *J Exp Soc Psychol.* 1995; 31(1):48–63.
16. Paluck E. L., Green D. P. Prejudice reduction: What works? A review and assessment of research and practice. *Annu Rev Psychol.* 2009; 60:339–367. <https://doi.org/10.1146/annurev.psych.60.110707.163607> PMID: 18851685
17. Soble J. R., Spanierman L. B., Liao H. Y. Effects of a brief video intervention on White university students' racial attitudes. *J Coun Psychol.* 2011; 58(1):151. <https://doi.org/10.1037/a0021158> PMID: 21142353
18. Neville H. A., Lilly R. L., Duran G., Lee R. M., Browne L. Construction and initial validation of the color-blind racial attitudes scale (CoBRAS). *J Couns Psychol.* 2000; 47(1):59.
19. Spanierman L. B., Heppner M. J. Psychosocial costs of racism to whites scale (PCRW): construction and initial validation. *J Couns Psychol.* 2004; 51(2):249.
20. Ponterotto J. G., Burkard A., Rieger B. P., Grieger I., D'Onofrio A., Dubuisson A., et al. Development and initial validation of the Quick Discrimination Index (QDI). *Educ Psychol Meas.* 1995; 55(6):1016–1031.
21. Barden J., Maddux W. W., Petty R. E., Brewer M. B. Contextual moderation of racial bias: the impact of social roles on controlled and automatically activated attitudes. *J Pers Soc Psychol.* 2004; 87(1):5. <https://doi.org/10.1037/0022-3514.87.1.5> PMID: 15250789
22. Hehman E., Volpert H. I., Simons R. F. The N400 as an index of racial stereotype accessibility. *Soc Cogn Affect Neurosci.* 2014; 9(4):544–552. <https://doi.org/10.1093/scan/nst018> PMID: 23386742
23. Proverbio A.M., Orlandi A., Bianchi E. Electrophysiological markers of prejudice related to sexual gender. *Neuroscience.* 2017; 358:1–12 <https://doi.org/10.1016/j.neuroscience.2017.06.028> PMID: 28666645
24. Proverbio A. M., Alberio A., De Benedetto F. Neural correlates of automatic beliefs about gender stereotypes: males are more prejudicial. *Brain Lang.* 2018; 186:8–16. <https://doi.org/10.1016/j.bandl.2018.08.006> PMID: 30179752
25. Ibáñez A., Haye A., Gonzalez R., Hurtado E., Henriquez R. Multi-level analysis of cultural phenomena: The role of ERPs approach to prejudice. *J Theory Soc Behav.* 2009; 39(1):81–110.
26. Amodio D.M. The neuroscience of prejudice and stereotyping. *Nat Rev Neurosci.* 2014; 15:670–682. <https://doi.org/10.1038/nrn3800> PMID: 25186236
27. Hagoort P., Hald L., Bastiaansen M., Petersson K. M. Integration of word meaning and world knowledge in language comprehension. *Science.* 2004; 304(5669):438–441. <https://doi.org/10.1126/science.1095455> PMID: 15031438

28. Metzner P., von der Malsburg T., Vasishth S., Rösler F. Brain responses to world knowledge violations: A comparison of stimulus-and fixation-triggered event-related potentials and neural oscillations. *J Cogn Neurosci*. 2015; 27(5):1017–1028. https://doi.org/10.1162/jocn_a_00731 PMID: 25269112
29. Correll J., Urland G. R., Ito T. A. Event-related potentials and the decision to shoot: The role of threat perception and cognitive control. *J Exp Soc Psychol*. 2006; 42(1):120–128.
30. White K. R., Crites S. L. Jr., Taylor J. H., Corral G. Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Soc Cogn Affect Neurosci*. 2009; 4(2):191–198. <https://doi.org/10.1093/scan/nsp004> PMID: 19270040
31. Wang L., Ma Q., Song Z., Shi Y., Wang Y., Pfothenauer L. N400 and the activation of prejudice against rural migrant workers in China. *Brain Res*. 2011; 1375:103–110. <https://doi.org/10.1016/j.brainres.2010.12.021> PMID: 21167137
32. Amodio D. M., Bartholow B. D., Ito T. A. Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Soc Cogn Affect Neurosci*. 2014; 9(3):385–393. <https://doi.org/10.1093/scan/nst177> PMID: 24319116
33. Kutas M., Hillyard S. A. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 1980; 207(4427):203–205. <https://doi.org/10.1126/science.7350657> PMID: 7350657
34. Brusa A., Bordone G., Proverbio A. M. Measuring implicit mental representations related to ethnic stereotypes with ERPs: An exploratory study. *Neuropsychologia*. 2021; 155, 107808. <https://doi.org/10.1016/j.neuropsychologia.2021.107808> PMID: 33636156
35. Kutas M., & Federmeier K. D. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*. 2011; 62:621–647. <https://doi.org/10.1146/annurev.psych.093008.131123> PMID: 20809790
36. Proverbio A.M., Riva F. RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neurosci Lett*. 2009; 459:142–146. <https://doi.org/10.1016/j.neulet.2009.05.012> PMID: 19427368
37. Kutas M., Federmeier K. D. N400. *Scholarpedia*. 2009; 4(10):7790. <https://doi.org/10.1093/scan/nsp038> PMID: 19776220
38. Federmeier K. D., Segal J. B., Lombrozo T., Kutas M. Brain responses to nouns, verbs and class-ambiguous words in context. *Brain*. 2000; 123(12):2552–2566. <https://doi.org/10.1093/brain/123.12.2552> PMID: 11099456
39. Hald L. A., Steenbeek-Planting E. G., Hagoort P. The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Res*. 2007; 1146:210–218. <https://doi.org/10.1016/j.brainres.2007.02.054> PMID: 17433893
40. Duffy S. A., Keir J. A. Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Mem Cognit*. 2004; 32(4):551–559. <https://doi.org/10.3758/bf03195846> PMID: 15478749
41. Jin J., Pei G., Ma Q. They are what you hear in media reports: the racial stereotypes toward Uyghurs activated by media. *Front Neurosci*. 2017; 11:675. <https://doi.org/10.3389/fnins.2017.00675> PMID: 29270104
42. Amodio D.M., Frith C.D. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*. 2006; 7:268–277. <https://doi.org/10.1038/nrn1884> PMID: 16552413
43. Mitchell J.P., Macrae C.N., Banaji M.R. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*. 2006; 50:655–663. <https://doi.org/10.1016/j.neuron.2006.03.040> PMID: 16701214
44. Quadflieg S., Turk D.J., Waiter G.D., Mitchell J.P., Jenkins A.C., Macrae C.N. Exploring the neural correlates of social stereotyping. *J Cogn Neurosci*. 2009; 21:1560–1570. <https://doi.org/10.1162/jocn.2009.21091> PMID: 18752409
45. Freeman J.B., Schiller D., Rule N.O., Ambady N. The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Hum Brain Mapp*. 2010; 31:150–159. <https://doi.org/10.1002/hbm.20852> PMID: 19618409
46. Contreras J.M., Banaji M.R., Mitchell J.P. Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Soc Cogn Affect Neurosci*. 2012; 7:764–770. <https://doi.org/10.1093/scan/nsr053> PMID: 21908447
47. Amodio D.M. The social neuroscience of intergroup relations. *Eur Rev Soc Psychol*. 2008; 19:1–54.
48. Saxe R. The right temporo-parietal junction: a specific brain region for thinking about thoughts. In: *Handbook of theory of mind*. Taylor & Francis Group, editor. Psychology Press. 2010. p.1–35.
49. Olson I.R., McCoy D., Klobusicky E. Ross L.A. Social cognition and the anterior temporal lobes: a review and theoretical framework. *Soc Cogn Affect Neurosci*. 2013; 8:123–133. <https://doi.org/10.1093/scan/nss119> PMID: 23051902

50. McCarthy G., Nobre A.C., Bentin S., Spencer D.D. Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *J Neurosci.* 1995; 15(2):1080–9. <https://doi.org/10.1523/JNEUROSCI.15-02-01080.1995> PMID: 7869084
51. Halgren E., Dhond R.P., Christensen N., Van Petten C., Marinkovic K., Lewine J.D., et al. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage.* 2002; 17(3):1101–16 <https://doi.org/10.1006/nimg.2002.1268> PMID: 12414253
52. Van Petten C., Luka B. J. Prediction during language comprehension: Benefits, costs, and ERP components. *Int J Psychophysiol.* 2012; 83(2):176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015> PMID: 22019481
53. DeLong K. A., Kutas M. Comprehending surprising sentences: sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. *Lang Cogn Neurosci.* 2020; 35(8):1044–1063.
54. Donchin E. Is the P300 component a measure of context updating? *Behavioral and brain Science.* 1988; 11:357–374.
55. Roehm D., Bornkessel-Schlesewsky I., Rösler F., Schlesewsky M. To predict or not to predict: Influences of task and strategy on the processing of semantic relations. *J Cogn Neurosci.* 2007; 19(8):1259–1274. <https://doi.org/10.1162/jocn.2007.19.8.1259> PMID: 17651001
56. Freunberger D., Roehm D. Semantic prediction in language comprehension: evidence from brain potentials. *Lang Cogn Neurosci.* 2016; 31(9):1193–1205. <https://doi.org/10.1080/23273798.2016.1205202> PMID: 27868079
57. Molinaro N., Carreiras M. Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biol Psychol.* 2010; 83(3):176–190. <https://doi.org/10.1016/j.biopsycho.2009.12.006> PMID: 20026375
58. Vespignani F., Canal P., Molinaro N., Fonda S., Cacciari C. Predictive mechanisms in idiom comprehension. *J Cogn Neurosci.* 2009; 22:1682–1700.
59. Cuthbert B.N., Schupp H.T., Bradley M.M., Birbaumer N., Lang P.J. Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol Psychol.* 2000; 52:95–111. [https://doi.org/10.1016/s0301-0511\(99\)00044-7](https://doi.org/10.1016/s0301-0511(99)00044-7) PMID: 10699350
60. Schupp H.T., Cuthbert B.N., Bradley M.M., Cacioppo J.T., Ito T., Lang P.J. Affective picture processing: the frontal positivity is modulated by motivational relevance. *Psychophysiology.* 2000; 37:257–261 PMID: 10731776
61. Rostami H. N., Ouyang G., Bayer M., Schacht A., Zhou C., Sommer W. Dissociating the influence of affective word content and cognitive processing demands on the late positive potential. *Brain topography.* 2016; 29(1):82–93. <https://doi.org/10.1007/s10548-015-0438-2> PMID: 26012382
62. Fischler I., Bradley M. Event-related potential studies of language and emotion: words, phrases, and task effects. *Prog Brain Res.* 2006; 156:185–203. [https://doi.org/10.1016/S0079-6123\(06\)56009-1](https://doi.org/10.1016/S0079-6123(06)56009-1) PMID: 17015080
63. Herbert C., Junghofer M., Kissler J. Event related potentials to emotional adjectives during reading. *Psychophysiology.* 2008; 45(3):487–498. <https://doi.org/10.1111/j.1469-8986.2007.00638.x> PMID: 18221445
64. Schacht A., Sommer W. Time course and task dependence of emotion effects in word processing. *Cogn Affect Behav Neurosci.* 2009; 9(1):28–43. <https://doi.org/10.3758/CABN.9.1.28> PMID: 19246325
65. Oldfield R.C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia.* 1971; 9:97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4) PMID: 5146491
66. Arcuri L., Boca S. Pregiudizio e affiliazione politica: destra e sinistra di fronte all'immigrazione dal terzo mondo. *Psicologia e politica.* 1996;241–273.
67. Oostenveld R., Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol.* 2001; 112:713–719. [https://doi.org/10.1016/s1388-2457\(00\)00527-7](https://doi.org/10.1016/s1388-2457(00)00527-7) PMID: 11275545
68. Zhang X., Li Q., Sun S., Zuo B. The time course from gender categorization to gender-stereotype activation. *Soc Neurosci.* 2018; 13(1):52–60. <https://doi.org/10.1080/17470919.2016.1251965> PMID: 27778758
69. Pascual-Marqui R. D., Michel C. M., Lehmann D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int J Psychophysiol.* 1994; 18(1):49–65. [https://doi.org/10.1016/0167-8760\(84\)90014-x](https://doi.org/10.1016/0167-8760(84)90014-x) PMID: 7876038
70. Palmero-Soler E., Dolan K., Hadamschek V., Tass P.A. swLORETA: a novel approach to robust source localization and synchronization tomography. *Phys Med Biol.* 2007; 52:1783. <https://doi.org/10.1088/0031-9155/52/7/002> PMID: 17374911

71. Bargh J. A. The cognitive monster: The case against the controllability of automatic stereotype effects. In: Chaiken S. & Trope Y., editor. *Dual-process theories in social psychology*. The Guilford Press; 1999. p.361–382. <https://doi.org/10.1037//0022-3514.76.6.893> PMID: 10402679
72. Dovidio J. F., Fazio R. H. New technologies for the direct and indirect assessment of attitudes. In Tanur J. M., editor. *Questions about questions: Inquiries into the cognitive bases of surveys*. Russell Sage Foundation. 1992. p. 204–237.
73. Greenwald A. G., Banaji M. R. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol Rev*. 1995; 102(1):4. <https://doi.org/10.1037/0033-295x.102.1.4> PMID: 7878162
74. Chen M., Bargh J. A. Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *J Exp Soc Psychol*. 1997; 33(5):541–560.
75. Menenti L., Petersson K. M., Scheeringa R., Hagoort P. When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *J Cogn Neurosci*. 2009; 21(12):2358–2368. <https://doi.org/10.1162/jocn.2008.21163> PMID: 19016600
76. Zahn R., Moll J., Krueger F., Huey E.D., Garrido G. Grafman J. Social concepts are represented in the superior anterior temporal cortex. *Proc Natl Acad Sci*. 2007; 104:6430–6435. <https://doi.org/10.1073/pnas.0607061104> PMID: 17404215
77. Patterson K., Nestor P.J., Rogers T.T. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci*. 2007; 8:976–987. <https://doi.org/10.1038/nrn2277> PMID: 18026167
78. Quadflieg S., Macrae C.N. Stereotypes and stereotyping: What's the brain got to do with it? *Eur Rev Soc Psycho*. 2011; 22:215–273.
79. 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. <https://doi.org/10.1037/a0021102> PMID: 21381825
80. Fletcher P. C., Happe F., Frith U., Baker S. C., Dolan R. J., Frackowiak R. S., et al. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition*. 1995; 57(2):109–128. [https://doi.org/10.1016/0010-0277\(95\)00692-r](https://doi.org/10.1016/0010-0277(95)00692-r) PMID: 8556839
81. Mahy C., Moses L., Pfeifer J. How and where: theory-of-mind in the brain. *Dev Cogn Neurosci*. 2014; 9:68–81. <https://doi.org/10.1016/j.dcn.2014.01.002> PMID: 24552989
82. Proverbio A.M., La Mastra F. Zani A How negative social bias affects memory for faces: An electrical neuroimaging study. *Plos one*. 2016; 11(9):e0162671. <https://doi.org/10.1371/journal.pone.0162671> PMID: 27655327
83. Van Overwalle F. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp*. 2009; 30(3):829–858. <https://doi.org/10.1002/hbm.20547> PMID: 18381770
84. Mitchell J. P., Heatherton T. F., Macrae C. N. Distinct neural systems subserve person and object knowledge. *Proc Natl Acad Sci*. 2002; 99(23):15238–15243. <https://doi.org/10.1073/pnas.232395699> PMID: 12417766
85. Frith C. D., Frith U. The neural basis of mentalizing. *Neuron*. 2006; 50(4):531–534. <https://doi.org/10.1016/j.neuron.2006.05.001> PMID: 16701204
86. Ronquillo J., Denson T. F., Lickel B., Lu Z. L., Nandy A., & Maddox K. B. The effects of skin tone on race-related amygdala activity: an fMRI investigation. *Soc Cogn Affect Neurosci*. 2007; 2(1):39–44. <https://doi.org/10.1093/scan/nsl043> PMID: 18985117
87. Eberhardt J.L. Imaging race. *Am Psychol*. 2005; 60(2):181–190. <https://doi.org/10.1037/0003-066X.60.2.181> PMID: 15740450
88. Lieberman M.D., Harii A., Jarcho J.M., Eisenberger N.I., Bookheimer S.Y. An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nat Neurosci*. 2005; 8(6):720–2. <https://doi.org/10.1038/nn1465> PMID: 15880106
89. Ito T. A., Thompson E., Cacioppo J. T. Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Pers Soc Psychol Bull*. 2004; 30(10):1267–1280. <https://doi.org/10.1177/0146167204264335> PMID: 15466600
90. Hart A. J., Whalen P. J., Shin L. M., McInerney S. C., Fischer H., Rauch S. L. Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*. 2000; 11(11):2351–2354. <https://doi.org/10.1097/00001756-200008030-00004> PMID: 10943684
91. Golby A. J., Gabrieli J. D. E., Chiao J. Y., Eberhardt J. L. Differential fusiform responses to same- and other-race faces. *Nat Neurosci*. 2001; 4:845–850. <https://doi.org/10.1038/90565> PMID: 11477432
92. Ratner K. G., Amodio D. M. Seeing “us versus them”: minimal group effects on the neural encoding of faces. *J Exp Soc Psychol*. 2013; 49:298–301.

93. Van Bavel J. J., Packer D. J., Cunningham W. A. The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychol Sci.* 2008; 11:1131–1139 <https://doi.org/10.1111/j.1467-9280.2008.02214.x> PMID: 19076485
94. Proverbio A.M., Camporeale E., Brusa A. Multimodal Recognition of Emotions in Music and Facial Expressions. *Front Hum Neurosci.* 2020; 14:32. <https://doi.org/10.3389/fnhum.2020.00032> PMID: 32116613
95. Dwivedi V. D., Selvanayagam J. Effects of Dispositional Affect on the N400: Language Processing and Socially Situated Context. *Front Psychol.* 2021; 12:1045. <https://doi.org/10.3389/fpsyg.2021.566894> PMID: 33868066
96. Schnuerch R., Richter J., Koppehele Gossel J., Gibbons H. Multiple neural signatures of social proof and deviance during the observation of other people's preferences. *Psychophysiology.* 2016; 53(6):823–836. <https://doi.org/10.1111/psyp.12636> PMID: 26928085