ELECTROPHYSIOLOGICAL MARKERS OF PREJUDICE RELATED TO SEXUAL GENDER

ALICE MADO PROVERBIO, ANDREA ORLANDI AND EVELINA BIANCHI

Abstract—Previous studies have shown that Event-related potentials (ERPs) are sensitive to violations of gender-based stereotypes. In the present investigation, we used ERPs to measure the detection of a discrepancy between gender-based occupational stereotypes and written material presented to fifteen Italian viewers in a completely implicit task. No awareness or judgment about stereotypes was required, no decision had to be made on sentence acceptability or congruence, and no prime words related to gender were presented (which might reveal the matter of the investigation). EEG was recorded while participants engaged in a task that consisted of pressing a key in response to animal words, while ignoring the overall study’s purpose. Two hundred forty sentences that did or did not violate gender stereotypes were presented randomly with 32 other sentences ending with an animal word. Terminal words violating gender stereotypes (such as “The notary is BREASTFEEDING” or “Here is the commissioner with HER HUSBAND”) elicited a greater anterior N400 response and left anterior negativity (LAN) than words conforming to the gender stereotype (e.g., “The chemist put on a nice TIE”). LAN modulation suggests that gender stereotypes are processed automatically (as if they were morpho-syntactic errors) and hints at how they are deeply rooted in our linguistic brain. According to the inverse solution, the neural representation of gender-based stereotypes mostly involved the middle frontal gyrus (MFG). The temporoparietal junction (TPJ) supporting theory of mind (TOM) processes was also engaged, along with the superior and middle temporal gyri representing person information.

Key words: ERPs, prejudice, social cognition, TOM, N400, world knowledge.

INTRODUCTION

The aim of this study was to investigate the temporal and neuro-anatomical functional correlates of processes related to the representation of prejudices by means of EEG/ERPs and source reconstruction techniques. In particular, the object of investigation was gender prejudice, which is a series of stereotypes associated with typical male or female appropriate or common behaviors, attitudes, clothing, jobs, personality traits, abilities, and others. This information refers to the so-called world knowledge shared by a given social group, and can be investigated by means of ERP measures (Ibañez et al., 2009; Amodio et al., 2013; Hagoort and van Berkum, 2007; Metzner et al., 2015) as demonstrated for example by Hagoort et al. (2004) in their seminal study. They found that a violation of common knowledge (for example the color of trains in Holland) elicited a N400 component similar to the semantic N400 found for violations of semantic relatedness or congruence in the neurolinguistic literature (Kutas and Hillyard, 1980, 1984; Kutas and Federmeier, 2011; Hald et al., 2007; Lau et al., 2008). The N400 has also been found to be affected by personal semantics (Coronel and Federmeier, 2016), that is by violations relative to subjective knowledge (i.e., personal preferences (likes and dislikes) across a wide range of topics, including foods, sport teams, music, films, etc.) or by action plausibility (Proverbio and Riva, 2009). As for the Italian language, several ERP studies have provided evidence of a N400 modulation for semantic (or syntactic) incongruences (see Angrilli et al., 2002). For example Proverbio et al. (2009) presented pairs of common or proper names’ words that could be semantically related (“Woody Allen” or “social security”), or not (“Sigmund Parodi” or “judicial cream”) and they found much larger N400s to unrelated than related items. Again N400 responses were obtained both for Slovenian and Italian semantically incongruent terminal words (e.g., “La struttura della citta era troppo invidiosa” –The structure of the city was too envious– or “Polglasno petje jo je privolilo”) in the bilingual brain of Italian/Slovenian speakers (Proverbio et al., 2002), with an effect of language age of acquisition (that is L1 vs. L2) but no effect of language (Italian vs. Slovenian) per se.

A few studies have used the N400 response as an index to investigate the neural representation of stereotypes (Bartholow et al., 2001, 2003). Osterhout et al. (1997) showed participants sentences referring to stereotypically male or female occupations and pronouns that did or did not match the gender stereotypically implied by the job (for example, “The beautician put herself through school” vs. “The beautician put himself through school”). They found increased N400 responses in association with the prejudice violation.

Whereas gender prejudice has been rarely investigated from a neuroscientific point of view or with
ERPs, racial bias has been much more commonly studied (e.g., see the famous other-race effects found for face processing (Hart et al., 2000; Caldara et al., 2003; Ronquillo et al., 2007) or empathy for pain (Contreras-Huerta et al., 2013). For example, Gehman et al. (2014) recorded ERPs in a sequential priming task in which positive or negative stereotypes related to black (African-American) or white (Caucasian) people followed the presentation of either black or white faces acting as primes. The results showed that the N400 was larger in the trials in which the face prime was incongruent with the target trait compared to trials in which the face prime was congruent with the target trait. In Correll et al. (2006)’s study, subjects were required to make a quick shoot/don’t-shoot decision in response to armed or unarmed characters, half of whom were Caucasian as opposed to Black. Self-reported measures of cultural stereotypes predicted both the racial ERP differentiation and racial bias in the game. The degree of ethnic differentiation in the P200 and N200 amplitude predicted behavioral bias in the videogame and correlated with explicit measures of race-based cultural stereotypes. With a similar N400 paradigm, the prejudice of Chinese city residents against rural migrant workers was tested (Wang et al., 2010). For example, in this study, the positive adjective “clean” preceded by a prime related to rural migrants elicited a larger N400 than the negative adjective “dirty”, as it violated their social representation.

Several neuroimaging studies suggest that in addition to the amygdala, there is a crucial area of the brain that plays a key role in the formation of the prejudices: the medial prefrontal cortex (mPFC). It has been observed, for example, that judging a person based on his/her political views in the absence of other information, or considering a person solely on the basis of ethnicity or political views in the absence of other information, or considering a person solely on the basis of ethnicity or political views in the absence of other information, or considering a person solely on the basis of ethnicity or political views in the absence of other information, is accompanied by increased activity of the mPFC (Mitchell et al., 2006). Accordingly, lesions of the medial prefrontal cortex are associated with a decrease of a whole series of implicit beliefs that produce gender bias (Milne and Grafman, 2001). These findings suggest, therefore, that the mPFC plays an important role in the generation of prejudices and stereotypes. Quadflieg et al. (2009) investigated the neural bases of gender prejudice by means of fMRI. Bold signals were obtained while subjects made judgments about the likely person or location of a series of activities, some of which were associated with a specific gender stereotype. The data showed that stereotyping was associated with enhanced activity in the ventral mPFC and in the amygdala among other areas. Notably, brain activity associated with stereotypic judgments was correlated with the strength of participants’ explicit and implicit gender stereotypes.

Freeman et al. (2010) used functional magnetic resonance imaging to dissociate the neural mechanisms underlying social judgments and found that BOLD activity in the anterior medial prefrontal cortex was reduced when participants were thinking about a person that did not belong to their social group (out-group), such as drug addicts, homeless people or people of non-white racial origins, compared to when they thought about in-group individuals.

Another region that seems to be strongly involved in the representation of negative bias or prejudice is the precuneus. Bruneau and Saxe (2010) presented Arabic, Israeli and American individuals with sentences about the Middle East situation from an in-group vs. an out-group perspective. Participants were required to rate how reasonable they found the statements while fMRI data were obtained. They found an enhanced activation of the precuneus while reading pro-out-group compared to reading pro-ingroup sentences, and this increased activation strongly correlated with both explicit and implicit measures of a negative prejudice toward individuals belonging to other ethnic groups.

As for gender stereotypes, White et al. (2009) used an explicit paradigm in which words related to gender (e.g., ‘Women’ or ‘Men’) were followed by a word that was either consistent or inconsistent with gender stereotypes (e.g., “Women: Nurturing vs. Aggressive”). Participants were required to decide whether the words matched based on gender stereotypes. Stereotype-incongruent word pairs resulted in greater N400 responses and slower reaction times as compared to congruent word pairs.

One of the possible problems with these paradigms is that subjects are aware of the object of the study or the manipulation because they are explicitly primed with ‘Women’ or ‘Men’ types of words and asked to access their gender stereotypes to make decision. This issue has only been partially avoided in other studies on gender stereotypes (e.g., Osterhout et al., 1997). If ERPs are recorded to reflexive pronouns that referred to a stereotypically male or female antecedent noun and if there were no sex-based primes, participants are still asked to consciously make a decision about the sentence acceptability, which might involve bias inhibition or negation processes linked to social desirability.

This study is the first study, to the best of our knowledge, in which sexual prejudice neural representation was investigated without requiring the participants’ awareness or judgment. No decision had to be made whatsoever on the sentence acceptability or congruence, and no words related to gender were presented (which might reveal the object of the investigation). Brain activation was observed while participants engaged in an animal word detection task, which completely ignored the study’s purpose. An implicit task was used to elicit automatic activation of subjective and unrepressed information related to gender occupational biases.

Independent implicit measures of strong gender prejudice have been created for the Italian population through all age ranges, including elderly individuals and children as a young as 8 years old (Siyanova-Chanturia et al., 2015).

On the basis of previous evidence in the literature, it was expected that we would find an enhanced N400 response to prejudice violations. Indeed, other electrophysiological studies investigating the neural representation of gender or racial prejudice have clearly determined the effect of prejudice violation on the amplitude of the N400 component. One of the main differences between the other studies and the present
study, besides its implicit nature, is that it involved reading whole sentences, whereas most previous studies simply used two words differently coupled (a gender-related prime word and a target word) and not a meaningful sentence. This simplicity can limit the access to complex contextual or semantic representations, such as those provided by sentences vividly describing a multiplicity of characters engaged in different occupational situations.

**EXPERIMENTAL PROCEDURES**

**Participants**

Fifteen University students, on average 24.4 years of age (SD = 3.1, min = 18, max = 35), participated in the study. There were 7 women ranging in age between 22 and 29 years old (24.42, SD = 4.79) and 8 men from 23 to 27 years old (24.3, SD = 1.7). All of them were Italian citizens and had heterosexual preferences as ascertained by a written questionnaire. All participants had normal or corrected-to-normal vision. They were strictly right-handed as assessed by the Oldfield Inventory (mean score = 0.88; SD = 0.13) and reported no history of drug abuse or neurological or mental disorders. Experiments were conducted with the understanding and written consent of each participant according to the Declaration of Helsinki (BMJ 1991; 302: 1194), with approval from the Ethics Committee of University of Milano-Bicocca.

**Stimuli**

The stimuli were 252 sentences, half of which (126) expressed notions congruent with the common sense or prevalent sexual prejudice, thereafter named congruent phrases (e.g., “The financial controller soiled HIS PANTS”, in Italian: “Il controllore finanziario si macchiò I PANTALONI”), and the other half expressed notions incongruent with common sense prejudices, thereafter named incongruent phrases (e.g., “The engineer stained HER SKIRT”, in Italian “L’ingegnere si macchiò la GONNA”). Sentences were created in a way so that the gender of the agent (the person engaging in a given professional activity, sport, or behavior) was made explicit only at the very end (the terminal word). Note that the English translations of the stimuli used in this study do not respect this rule. The prejudice violation involved males in half of the cases and females in the other half. Phrases concerning men and women (both congruent and incongruent) were balanced for semantic domains (child care, jobs, hobbies, housework, clothes, physical strength, hairstyle), as shown in Table 1.

For ERP averaging, EEG epochs were synchronized with the onset of the terminal words. The latter were balanced across categories in length and frequency of use. The frequency of use of the terminal word was determined on the basis of the CoLFIS database by Bertinetto et al. (2006). CoLFIS is a lexical database of written Italian language. It comprises 3,798,275 lexical occurrences taken from newspapers, magazines and books, including textbooks. Statistical comparisons (ANOVAs) performed on the mean frequency of use of terminal words across the congruent vs. incongruent condition showed no differences \( [F(1, 125) = 0.037, p = 0.84] \) whatsoever (congruent = 115.91, SD = 25.68; incongruent = 110.17, 20.48). A further ANOVA performed on mean word length (\# of letters) showed balance across stimulus classes \( [F(1, 125) = 0.004, p = 0.94] \): the length

![Fig. 1.](image1.png) Distribution of scores obtained for congruent and incongruent sentences during stimulus validation, reflecting the different terminal word cloze probabilities (Kutas and Hillyard, 1984).

Table 1. Some examples of occupational gender-biased prejudices. A switch among columns represents a violation of the common-sense prejudice

<table>
<thead>
<tr>
<th></th>
<th>Man</th>
<th>Woman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job</td>
<td>Engineer</td>
<td>Dance teacher</td>
</tr>
<tr>
<td></td>
<td>Chemical engineer</td>
<td>Kindergarten teacher</td>
</tr>
<tr>
<td></td>
<td>Vascular surgeon</td>
<td>Baby sitter</td>
</tr>
<tr>
<td>Architect</td>
<td>Secretary</td>
<td></td>
</tr>
<tr>
<td>Computer scientist</td>
<td>Beautician</td>
<td></td>
</tr>
<tr>
<td>Housework</td>
<td>Fix roller shutter</td>
<td>Ironing</td>
</tr>
<tr>
<td></td>
<td>Changing tires</td>
<td>Cooking</td>
</tr>
<tr>
<td></td>
<td>Building fireplace</td>
<td>Mending</td>
</tr>
<tr>
<td></td>
<td>Roof maintenance</td>
<td>Child care</td>
</tr>
<tr>
<td></td>
<td>Oil change</td>
<td>Laundry</td>
</tr>
<tr>
<td>Sport</td>
<td>Body-building</td>
<td>Cheer-leader</td>
</tr>
<tr>
<td></td>
<td>Soccer</td>
<td>Classical dance</td>
</tr>
<tr>
<td></td>
<td>Rally</td>
<td>Synchronized swimming</td>
</tr>
<tr>
<td></td>
<td>Boxing</td>
<td>Pilates</td>
</tr>
<tr>
<td></td>
<td>Boxing</td>
<td>Ravelli</td>
</tr>
<tr>
<td></td>
<td>Rugby</td>
<td>Figure skating</td>
</tr>
</tbody>
</table>

![Table 1.](table1.png)
was on average 7.62 letters (SD = 0.19) for congruent terminal words and 7.64 letters (SD = 0.18) for incongruent words. Terminal words were also balanced for grammatical category (number of proper names, adjectives, verbs, nouns, professions, and adjectives used as nouns) and for concreteness and imagery value. All the balances were also checked across the female character and male character categories.

Stimulus validation

To determine whether the sentences actually represented (or violated) common sense prejudices for the specific populations (university students in the Milan metropolitan area), the stimuli underwent validation. Two hundred fifty-two phrases were randomly mixed and presented in a written questionnaire to a group of 20 University students (10 men and 10 women) ranging in ages between 18 and 35 years. All participants were Italian citizens and heterosexual. They were asked to rate, by means of a 3-point Likert scale, how they reacted to reading the terminal word of the phrase (the terminal word printed in bold capital characters). Scale units were as follows: 0 = Actually, I was a bit surprised, 1 = I do not know, and 2 = I kind of saw that coming. On the basis of the results, 7 incongruent phrases were modified, and 6 of them were eliminated. As a result, a corpus of 120 phrases violating common prejudices rated from 1.5 to 0 was obtained. Eleven congruent phrases were modified, and 6 congruent phrases were eliminated. As a result, a corpus of 120 phrases supporting common gender-biased prejudices rated from 1.5 to 2 was obtained (see the scores in Fig. 1 and some examples of phrases in Appendix 1).

Thirty-two other sentences (sharing the syntactic, semantic and lexical characteristics of the previous ones) were created for the task’s stated purpose and had a familiar animal as the terminal word (e.g., rabbit, woodpecker, horse, kangaroo).

All sentences were presented randomly, mixed in 8 different experimental sequences. Sentences were flashed for 1000 ms and arranged in two or three short rows centered circularly around the fixation point. The terminal words followed after an ISI of 700 ms and were typed in uppercase. Their size was 5.89 cm/\text{character}, 1.1 cm/\text{letter}, which was 10°18′ × 1°54′ of the visual angle (minimum length = 1.8 cm, maximum length = 12.2 cm). Their duration was 1000 ms, and they were followed by an ITI of 1200 ms. The text was printed in yellow on a black background. Terminal words were equiluminant across experimental categories as determined by means of a Minolta CS-100 luminance meter. The mean luminance values of terminal words in the congruent/incongruent classes were subjected to an ANOVA. The results were not statistically significant \[F(1, 112) = 0.24, p = 0.62\] (Congruent = FL = 3.78, SE = 0.04; Incongruent = FL = 3.81, SE = 0.04).

Procedure

Participants comfortably sat in a faradized and acoustically shielded cubicle in front of a PC monitor placed 100 cm from their eyes. The PC was located outside the cubicle and was visible through a mirror. Participants were asked to fixate on the center of the screen where a red dot served as a fixation point. They were instructed to sit relaxed but still while avoiding any head or body movements or ocular saccades. They were also trained to minimize eye blinks during EEG recordings. The task consisted of responding as quickly and accurately as possible to the terminal word when it represented an animal by pressing a joypad key with the index finger of the right or left hand. The response hand was alternated across trials and announced at the beginning of each trial. The sequence order and presentation and response hand order were randomized across subjects. A brief training using novel sentences preceded the beginning of actual EEG recording. Each experimental run lasted 2 min and 40 s and was followed by a 30-s pause. 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” on the screen. Both the warnings and thanks were typed in uppercase characters.

The task was implicit in that the potential congruence or incongruence with gender-biased prejudices was not made explicit or suggested to participants, who were focused on looking for an animal terminal word. Behavioral data of targets were recorded.

EEG recording and analysis

EEG signals were continuously recorded from 128 scalp sites located according to the 10–5 International System (Oostenveld and Praamstra, 2001) at a sampling rate of 512 Hz. Horizontal (hEOG) and vertical (vEOG) eye movements were also recorded. Linked mastoids served as the reference lead. The EEG and EOG were filtered with a half-amplitude band pass of 0.016–70 Hz. Electrode impedance was maintained below 5 KOhm. EEG epochs were synchronized with the onset of stimulus presentation. Computerized artifact rejection was performed prior to averaging. The artifact rejection criterion was a peak-to-peak amplitude exceeding 50 μV. This procedure resulted in a rejection rate of about 5%. Evoked response potentials (ERPs) from 100 ms before to 1000 ms after stimulus onset were averaged off-line. The mean amplitude area of the N170 response, sensitive to orthographic properties, was recorded from occipito/temporal POO9h and POO10h electrode sites in the 170- to 190-ms temporal window. The mean amplitude area of the N400 response sensitive to semantic or pragmatic incongruence was recorded from anterior and fronto-central sites (AFP3h, AFP4h, FFC5h, FFC6h, FCC5h, FCC6h) in the 350- to 450-ms temporal window. The mean amplitude area of the (late) Left anterior negativity (LAN) was recorded from fronto-temporal sites (FT7, FT8, FFT7h, FFT8h) in the 600- to 800-ms time window. For each ERP component, the mean area amplitude values were subjected to repeated-measures ANOVAs whose factors of variability were prejudice (congruent, incongruent), electrode (depending on the component of interest), and hemisphere (left, right). Tukey’s post hoc comparisons
were used. Behavioral data were not statistically analyzed, but response time and accuracy were collected as a record.

Two low-resolution electromagnetic tomographies (LORETA; Pasqual-Marqui et al., 1994) were performed on the ERP waveforms relative to the N170 and N400 potentials using ASA4 Software. LORETA is a discrete linear solution to the inverse problem, which corresponds to the 3D distribution of neural electrical activity that maximizes similarity (that is maximizes synchronization) in terms of orientation and strength between neighboring neural populations (represented by adjacent voxels). Here, an improved version called standardized weighted low-resolution brain electromagnetic tomography was used (i.e., swLORETA; Palmero-Soler et al., 2007). The data were automatically re-referenced to the average reference as part of the LORETA analysis.

RESULTS

Behavioral results

The mean RT to targets was 634.5 ms (637.7 ms for the left hand and 630.6 ms for the right hand). RT was on average 592 ms for women (597 ms for the left hand and 583.6 ms for the right hand) and 676.99 ms for men (678.4 ms for the left hand, and 677.59 ms for the right hand). Accuracy was very high, equal to 98% correct responses, indicating attentive reading by participants.

Electrophysiological results

Fig. 2 shows grand-average ERPs recorded for congruent vs. incongruent terminal words. An N400 negative deflection over the left frontal area is visible, followed by a later negativity (LAN) in response to terminal words violating common gander-biased prejudices.

N170 (170–190). The ANOVA carried out on N170 amplitude values showed no effect of prejudice \((p = 0.73)\) but a significant effect of hemisphere \((F(1, 14) = 12.34, p < 0.0035)\) with larger responses recorded over the left \((\mu V = -2.61, SE = 0.65)\) than right hemisphere \((\mu V = -0.76, SE = 0.75)\).

N400 (350–450 ms). The ANOVA carried out on N400 amplitude values showed a significant effect of prejudice \((F(1, 14) = 10.14, p < 0.01)\) with greater responses to incongruent \((\mu V = 1.77, SE = 0.83)\) than congruent.
words (μV = 2.2, SE = 0.82), as shown in Fig. 3 (left). The electrode factor trended in significance [F(2, 28) = 3.28, p = 0.052], indicating a larger N400 response over fronto-central sites (FFC5h-FFC6h: 1.65, SE = 0.87 μV, FCC5h-FCC6h: 1.83, SE = 0.84 μV) than anterior frontal sites (AFp3h-AFp4h: μV = 2.47, SE = 0.83), which was confirmed by post hoc comparisons (p = 0.02).

LAN (600–800 ms). The ANOVA performed on LAN mean amplitude area yielded a significant effect of prejudice [F(1, 14) = 7.34, p < 0.01], with larger responses to incongruent (μV = 2.15, SE = 0.56) than congruent (μV = 2.64, SE = 0.58) stimuli, as shown in Fig. 3 (right). The hemisphere factor was also found to be significant [F(1, 14) = 23.09, p < 0.01] with larger potentials over the left (μV = 1.55, SE = 0.46) than right hemisphere (μV = 3.24, SE = 0.69). The significant interaction of prejudice × hemisphere [F(1, 14) = 8.9, p < 0.01] and subsequent post hoc comparisons showed how the effect of prejudice was highly significant over the left hemisphere (p < 0.002; incongruent = 1.11, SE = 0.51 μV; congruent = 2.01, SE = 0.45 μV), but it did not affect LAN amplitude over the right hemisphere (p = 0.97; incongruent = 3.19, SE = 0.67 μV; congruent = 3.28, SE = 0.73 μV).

Whereas prejudice did not affect N170 response at all (devoted to orthographic and sub-lexical processing), an swLORETA was applied to grand-averaged ERPs of final words (regardless of congruency) in the latency range of 170–190 ms (corresponding to N170 peak) to identify its neural generators during word reading. The inverse solution localized its most powerful generator activity to the fusiform gyrus (BA37) of the left temporal lobe (see Fig. 4), which strongly agrees with neuroimaging literature that calls this region the Visual Word Form Area (Cohen and Dehaene, 2004; Proverbio et al., 2008).

To identify the neural bases of prejudice representation, a difference-wave was computed by subtracting ERPs for congruent minus incongruent words, and an inverse solution was applied to the difference voltage recorded in the 350–450 ms time-window, corresponding to the N400 peak, whose scalp distribution is displayed in Fig. 5. Table 2 shows a list of significant electro-magnetic dipoles, and swLORETA results are shown in Fig. 6. The strongest generator of surface activity was localized in the middle prefrontal cortex (Fig. 6, above). The temporo/parietal junction (TPJ) (Fig. 6, below) and the superior and middle temporal gyri were also engaged.

**DISCUSSION**

The first deflection that was shown to be sensitive to experimental manipulation emerged approximately 400 ms from the presentation of the stimuli. Specifically, it reached its peak between 350 and 450 ms over frontal areas. This negativity can be considered an N400, a cognitive component that occurs when incoming information is perceived as anomalous from the semantic point of view or as a result of the violation of the contextual constraints of the stimulus (Kutas and Hillyard, 1980). In our case, the component was strongly modulated by the experimental condition. In fact, we found a greater negativity in response to incongruent sentences (with violation of prevalent social prejudice about female and male attitudes, professional activities, etc.). This result is consistent with previous ERP investigations on sexual prejudices carried out with explicit paradigms (Osterhout et al., 1997; White et al., 2009).

At a later time window, between 600 and 800 ms, another negative deflection sensitive to prejudice violation reached a maximum amplitude over fronto-temporal regions: a late Left Anterior Negativity (LAN). The LAN is a linguistic ERP component sensitive to complex morpho-syntactic aspects, such as subject/verb, and verb/article agreement (Morris and Holcomb, 2005; Dowens et al., 2010; Tanner and Van Hell, 2014).

In this study, LAN was greater in response to incongruent sentences (with violation of common sense bias). We also found a strong left hemisphere asymmetry in LAN scalp distribution, as predicted by previous literature.

Interestingly, in Osterhout et al.’s (1997) study, the pronouns inconsistent with the definition of gender, or the gender bias of a name, elicited not a Late Anterior Negativity but a P600. The P600 or SPS (Syntactic Positive Shift) is a positive ERP component with a more posterior distribution and typically late latency. It is typically observed when some aspect of the structure of the sentence violates the rules of language, as in the case of the phrase “The doctor prepared herself for the operation”, in which we would expect a male doctor and not a female. This syntactic ERP component is supposed to reflect costs of repair and revision of structural mismatches and/or higher order integration processes (Friederici, 2002;...
The fact that (besides the N400) the prejudice violation elicited enhanced grammar-related ERP components, such as the P600 and the LAN, might suggest that gender prejudices are so deeply rooted in our memory that their violation is treated like a morpho-syntactic or linguistic error. But this is just a speculation and it deserves further research. On the other hand, the presence of a P600 (as in Osterhout et al.’s (1997) case) instead of a LAN (as in the present case) might be related to the degree of subjects’ awareness, which depends on the explicit or implicit nature of the task. Indeed, the P600 is thought to reflect more cognitive control, and LAN is thought to represent more automatic syntactic processing (Jiménez-Ortega et al., 2014).

An swLORETA inverse solution was applied to the difference-wave (incongruent minus congruent) to identify the neural generators of the anterior N400 to prejudice violations and it showed strong involvement the superior and especially the middle frontal cortex, which is particularly relevant to the theory of mind (TOM: Radua et al., 2014). According to the neuroimaging literature, the medial frontal cortex represents social information that refers to others, particularly outgroup stereotyping and prejudice (Mitchell et al., 2006). In particular, sub-regions of the medial prefrontal cortex (mPFC) differentiate between thinking about the attributes and/or mental states of similar versus dissimilar others (Mahy et al., 2014). In a recent study on the neural basis of prejudice, it was found that the left superior-frontal cortex (BA10) was particularly involved in representing negative prejudices related to others (Proverbio et al., 2016), which strongly fits with the current findings.

It should be mentioned that the middle frontal cortex (and especially the dorsolateral prefrontal cortex, BA9...
Table 2. Talairach coordinates (in mm) and localization of active electromagnetic dipoles explaining the surface voltage (incongruent – congruent to prejudice) recorded between 350- and 450-ms post-stimulus, according to swLORETA inverse solutions. Magn = magnitude in nA; T = Talairach; Hem = hemisphere; BA = Brodmann areas; L = left, R = right

<table>
<thead>
<tr>
<th>Magn</th>
<th>T-x</th>
<th>T-y</th>
<th>T-z</th>
<th>Hem.</th>
<th>Lobe</th>
<th>Gyrus</th>
<th>BA</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.00</td>
<td>40.9</td>
<td>55.3</td>
<td>7.0</td>
<td>R</td>
<td>Frontal</td>
<td>Middle Frontal</td>
<td>10</td>
<td>Prejudice: Proverbio et al. (2016), Mitchell et al. (2006), Mahy et al. (2014)</td>
</tr>
<tr>
<td>3.68</td>
<td>–58.5</td>
<td>–13.7</td>
<td>36.6</td>
<td>L</td>
<td>Parietal</td>
<td>Supramarginal</td>
<td>40</td>
<td>Temporo/parietal junction (Saxe, 2010) TOM</td>
</tr>
<tr>
<td>3.49</td>
<td>–18.5</td>
<td>41.4</td>
<td>41.8</td>
<td>L</td>
<td>Frontal</td>
<td>Superior Frontal</td>
<td>8</td>
<td>TOM</td>
</tr>
<tr>
<td>2.47</td>
<td>1.5</td>
<td>8.5</td>
<td>65.9</td>
<td>R</td>
<td>Frontal</td>
<td>Superior Frontal</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3.30</td>
<td>–38.5</td>
<td>10.4</td>
<td>48.1</td>
<td>L</td>
<td>Frontal</td>
<td>Middle Frontal</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2.82</td>
<td>60.6</td>
<td>13.3</td>
<td>21.4</td>
<td>R</td>
<td>Frontal</td>
<td>Inferior Frontal</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>2.92</td>
<td>–58.5</td>
<td>5.3</td>
<td>2.7</td>
<td>L</td>
<td>Temporal</td>
<td>Superior Temporal</td>
<td>22</td>
<td>Social affective information</td>
</tr>
<tr>
<td>2.21</td>
<td>40.9</td>
<td>9.1</td>
<td>–27.5</td>
<td>R</td>
<td>Temporal</td>
<td>Superior Temporal</td>
<td>38</td>
<td>Gender prejudice (Quadflieg et al., 2009) Person information, Gender prejudice (Quadflieg et al., 2009)</td>
</tr>
<tr>
<td>2.82</td>
<td>60.6</td>
<td>–57.9</td>
<td>5.6</td>
<td>R</td>
<td>Temporal</td>
<td>Middle Temporal</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>2.69</td>
<td>–58.5</td>
<td>–1.4</td>
<td>–20.8</td>
<td>L</td>
<td>Temporal</td>
<td>Middle temporal</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Axial and sagittal views of the active sources relative to the incongruent minus congruent ERP difference-waves recorded in the 350–450 time-window, which corresponds to the N400 peak, according to swLORETA analysis. The most active brain area engaged in representing sexual prejudice were the right middle prefrontal cortex (above) and the left temporo/parietal junction (supramarginal gyrus, below).
and BA46, DLPFC) is also involved in decision-making processes and attentional control (e.g., Rahnev et al., 2016). However, in our study, BA10 activation derived from the incongruent minus congruent-to-prejudice subtraction, not just from the activity recorded during the experimental session. Furthermore, participants’ attentional and decisional processes were engaged in performing the task, consisting in responding as accurately and quickly as possible to animal names, while sentences carrying social information about persons were task irrelevant. Therefore, it can be safely assumed that, in our study, the MFG activation (BA10) in response to sentences violating gender bias was specifically related to the processing social attributes, occupational stereotypes and prejudices.

The second most powerful source (in order of magnitude) was the supramarginal gyrus (BA40). This area is located in the parietal lobe and represents the dorsal part of the temporoparietal junction (TPJ). The TPJ has been associated with the ability to attribute intentions and meanings to the behavior of others, which is part of TOM (Saxe, 2010; Young et al., 2010).

Supposedly, during sentence reading, volunteers formed a mental representation of the scene depicted by imagining the person who was taking the action described, and by forming an impression about his/her character while also accessing available information about their own experiences and knowledge on the matter (world knowledge). This process resulted in the activation of specific brain regions, including the superior temporal gyrus (STG, BA38 and BA22), which was the third most highly activated area after TOM-related areas (prefrontal cortex and TPJ). The STG is believed to encode social information about the body and is involved in action understanding (Saxe et al., 2006; Jellema et al., 2000; Perrett et al., 1985). The middle temporal gyrus (BA 21) was also found to be active, and it is engaged in processing socio-affective information and judgments about other individuals. The activation of these areas fully agrees with other findings in the literature, most notably the fMRI study by Quadflieg et al. (2009) showing how the neural representation of gender bias was linked to the activity of the medial temporal gyrus (BA21), the superior frontal gyrus and the STG. Consistently, both neuropsychological (Semenza, 2009) and electrophysiological evidence have been provided in the literature of the role of the medial temporal gyrus in representing information concerning persons and proper names (Proverbio et al., 2001, 2009).

One of the best merits of this study is that we investigated the timing and neural circuits underpinning the representation of gender prejudice using a completely implicit paradigm, thus avoiding the activation of mental processes (and related brain areas) not related to prejudice representation but to prejudice suppression and inhibition caused by the well-documented social desirability pressure (Paulhus, 1984).

A possible study limitation is the lack of implicit measures of subjective gender stereotypes that might have been collected through the administration of the Implicit Association Test (IAT). However, the existence of strong gender stereotypes has been shown for the Italian population, including students and adults in the 18–25 year age range (Siyanova-Chanturia et al., 2015).

**AUTHOR CONTRIBUTION STATEMENT**

AMP conceived and designed the study, interpreted the data and wrote the paper. E.B. and AO developed stimuli, gathered and analyzed data.

**CONFLICT OF INTEREST STATEMENT**

The authors declare that the research was conducted in the absence of any real or perceived conflicts of interest.

**REFERENCES**


GLOSSARY

ANOVA: analysis of variance
BA: Brodmann area
BEM: boundary element model
BOLD: blood oxygenation level dependent
EEG: electroencephalogram
EOG: electro-oculogram
ERP: event-related potentials
fL: footlambert
fMRI: functional magnetic resonance imaging
fMRI: functional magnetic resonance imaging
IAT: implicit Association Test
isi: inter-stimulus-interval
LAN: left anterior negativity
LORETA: low-resolution electromagnetic tomography
mPFC: medial prefrontal cortex
ORE: other race effect
nA: nanoampere


APPENDIX

Some examples of phrases used as stimuli. The whole stimulus-set comprised 240 sentences balanced for grammatical category, concreteness, imagery value, length and frequency of written use of terminal words, semantic domain, sex of agent, and gender-biased prejudice. Please note that in the Italian language, personal pronouns are often omitted (e.g., in the sentence: “(he) went to Rome”): consequently, the agent’s sex can remain unknown till the end of the phrase. See, for example, sentence n°1 where the agent’s sex is never specified, but two actions are described that are typically associated with the two genders: shaving for a man and preparing tomato sauce (prejudicially) for a woman.

**Sentences incongruent with prejudices related to men**

Preparò il sugo e se fece la BARBA
L’assistente sociale si lisciò i BAFFI
Lasciò il pattinaggio artistico quando divenne PADRE
Esibendosi alla trave d’equilibrio si è INFORTUNATO
Stese i panni e raggiunse la MOGLIE
Faremmo ore ed ore di shopping, se fosse per GABRIELE
Uscendo dal corso di danza classica perse la PIPA
Lavorò all’asilo nido con la FIDANZATA
Portava una gonnellina colorata quel CAMERIERE
Quei gilet li ha fatti a maglia mio CUGINO

**Sentences incongruent with prejudices related to women**

Finito di imbiancare, era STRAVOLTA
Il notaio sta ALLATTANDO

L’ingegnere si macchiò la GONNA
Per cambiare l’olio al motore si è MACCHIATA
Il nome del sindaco è Niccolini GIUSI!
L’informatico musicale si fece la CERETTA
Il pugile ha appena PARTORITO
Cadendo dal tetto, l’antennista si è quasi AMMAZZATA

Nessuno in palestra è forte come mia MAMMA
Ecco l’assessore con suo MARITO

**Sentences congruent with prejudices related to men**

Il chimico si mise una bella CRAVATTA
Il marinaio ha una nuova AMMIRATRICE
Ho fatto il pieno dal solito BENZINAIO
Il gommista soffre di PROSTATITE
Ha tinteggiato tutto il locale, è ESAUSTO
Il financial controller si sporca i PANTALONI
Il tecnico di laboratorio si rovinò il CAMICE
Quel poliziotto era LUCA
Operò nell’intelligence finché non fu SCACCIATO
Finito di mettere le piastrelle, era STREMATO

**Sentences congruent with prejudices related to women**

Ha preparato una coreografia in piscina di cui va FIERA
Lavora come baby-sitter ed è molto MATERNA
Non lavora ma viene MANTENUTA
Alla gara di ballo ha partecipato pure ANNA
Con indosso il suo tutù raggiunse il podio ECCITATA

(continued on next page)
Fece mangiare la bambina e andò dalla PARRUCCHIERA
Lavorare al centralino piace molto a LAURA
Amava così tanto i bambini che diventò MAESTRA
Smise di sciare appena fu MADRE
Per usare ago e filo si è FERITA

Fed the little girl and went to the lady hairdresser
Laura enjoys working as a switchboard operator
Loved kids so much that became an Elementary school teacher
Dropped out of skiing as soon as she became mother
She hurt herself doing needle work

<table>
<thead>
<tr>
<th>Target phrases related to animals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luca ha l’abitudine di andare al parco con il CANE</td>
</tr>
<tr>
<td>E’ sicuramente daltonico perché colora di rosa le PECORE</td>
</tr>
<tr>
<td>Luana ha voluto comprare un PESCE</td>
</tr>
<tr>
<td>Mio nipote ha fotografato un FENICOTTERO</td>
</tr>
<tr>
<td>Alessandro è un vero appassionato di GUFI</td>
</tr>
</tbody>
</table>

(Received 23 January 2017, Accepted 19 June 2017)
(Available online 28 June 2017)