Performance Errors of in Group/ Out Group Stimuli and Valence Association in the Implicit Association Test: Brain Bias of Ingroup Favoritism

Raphael Guex[1,5], Francisco Cerić[3,4], Esteban Hurtado[3,4], Ramiro González[3,4], Álvaro Navarro[3], Facundo Manes[1] and Agustín Ibáñez*[,1,2,3]

[1]*Laboratory of Experimental Psychology and Neurosciences, Institute of Cognitive and Behavioural Neurology, INECO, Argentina*

[2]*National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina*

[3]*Laboratory of Cognitive Neuroscience, Universidad Diego Portales, Santiago, Chile*

[4]*Pontificia Universidad Católica de Chile*

[5]*Laboratory for Behavioral Neurology and Imaging of Cognition, University of Geneva, Geneva, Switzerland*

Abstract: The goal of this study is to assess the role of membership and valence effects on errors performed in a racial implicit association test indexed by event-related potentials (ERPs). Non-indigenous participants performed an implicit association test (IAT) paradigm emphasizing the feedback of error due to misclassification of ingroup (non-indigenous) and outgroup (indigenous) faces as well as positive and negative words. As expected, participants responded to the compatible task with higher accuracy than to incompatible tasks. This is the first report demonstrating that IAT errors produce electrophysiological ERP modulation. Our results suggest that medial frontal negativity is modulated not only by IAT error of membership and valence classifications but also by IAT compatible and incompatible tasks. These results provide a basis for the future use of the misclassification error in the IAT recorded simultaneously with ERPs in other classic social psychology contexts.

INTRODUCTION

The commission of classification errors by subjects while watching ethnic markers can be influenced by racial bias, especially when those markers are associated with a positive or negative valence aspect. The goal of this study is to assess the role of membership and valence effects on errors performed in a racial implicit association test indexed by reaction times and neural correlates of error processing.

The electroencephalogram (EEG) technique provides an excellent method for temporal determination of event-related brain activity changes *via* the measurement of event-related potentials (ERP). When ERPs are related to error detection, an early negative deflection or Medial Frontal Negativity (MFN) with a specific topology and brain source is observed [1]. The MFN has been first described as a measure of the motor response [2, 3], as well as negative feedback to the response [4]. In the first case, self-generated information in performance monitoring produces an error related negativity (ERN, -10 to 50ms after the response; [3]). In the second case, after the feedback the response is observed as a Feedback Related Negativity (FRN) produced around 110-250 milliseconds after the stimulus and is related to negative feedback perception, which is externally provided as part of performance monitoring [4]. Both measures are useful for studying the neural correlates of error processing. A possible critical factor in MFN modulation is the subject's expectation towards the task, based on the relationship between an action and its expected consequence, as a general conflict evaluation mechanism [5] or as an action self-monitoring function [6]. For this reason, expectation towards the feedback or conflict and its result will directly modulate MFN amplitude.

With respect to the neural correlates of expectation, interesting evidence is obtained from social neuroscience, which suggests that expectations can be strongly regulated by bias factors from groups to minorities. For example, a study by Amodio and colleagues showed that the ERN component is modulated when a bias error detection task using stereotypes is applied [7]. This research is particularly interesting because the bias related to the outgroup (African-Americans) generated higher ERN amplitudes for white subjects' error responses with respect to the African-Americans wearing a tool in the task trial, in which subjects confused the tool with a gun. From a theoretical point of view, the error monitoring displayed by the white subjects required a stereotype inhibition when the African-Americans wearing tools were shown; during such trials, white subjects needed a higher control to respond correctly. The inhibition presented by these subjects recalls the idea of conflict detection and/or breaking of expectation, consistent with functional neuro-

*Address correspondence to this author at the Laboratory of Experimental, Psychology and Neuroscience; Institute of Cognitive Neurology (INECO) & CONICET, Castex 3293 (CP 1425) Buenos Aires, Argentina; Tel/Fax: +54 (11) 4807-4748; E-mail: aibanez@neurologiacognitiva.org

anatomical evidence for the participation of specific brain areas (anterior cingulate cortex, ACC) involved in error detection and for performance monitoring in this type of task [8]. According to the conflict detection theory, these brain areas comprise MFN's main neural generators [6].

Current research on prejudice has used prejudice implicit bias conceptualization to suggest that an individual's attitude toward a member of an outgroup is mainly based on automatic associative processes, which would explain biased behavior. This would suggest that implicit measurement techniques can reflect introspectively unidentified marks of past experiences [9]. One measure of implicit attitude, among many others used in social psychology, is the Implicit Association Test (IAT) [10]. The IAT is a simultaneous stimulus categorization task that works by comparing subjects' reaction times when classifying a word or image shown on the computer screen into one of two categories of response. There has been extensive research on implicit and explicit attitudes in the field of racial differences, and numerous findings have been reported on black and white subjects [11], Asians [12], Germans and Turks [13], and Chilean-indigenous people [14, 15]. The present study reports IAT errors of Chilean non-indigenous people, using stimuli of indigenous vs. non-indigenous faces and positive vs. negative valence words. Mapuches are Chile's largest indigenous group and are also one of the most deprived social groups due to their historical conflicts with Spaniards and with the Chilean Republic [16, 17].

Recent studies using ERPs and IAT simultaneously reported that an early and a late modulation occurred when responding to IAT stimuli of faces and words [15, 18, 19]. ERP studies of race-IAT found that there is an early blending (N170) of stimuli content and contextual information [15, 18] and a LPP component modulated by compatible and incompatible tasks [18]. To our knowledge, no electrophysiological correlation of error responses in IAT studies has been previously reported, and errors in the IAT tasks are not frequently considered in social psychology [20]. Then this study looks for the efficacy of errors in the IAT, providing useful and interesting behavioral and electrophysiological information about racial bias and IAT effects.

We expected to find here that the contextual associated valence related to the stimuli influence the cerebral processing. Further, since the MFN has an early and a late temporal window, we can establish the valence emotional effect in the cerebral process of the IAT errors, both in the early (automatic), and in the late window (more related to uncertainty). In this way, we propose that the cerebral activity associated with the expectation (MFN) is modulated by membership and valence effects. Finally, we look for those effects in both temporal windows of the motor response and after the feedback.

MATERIALS AND METHODS

Participants

Eighteen Chilean non-indigenous subjects, between the ages of 18 and 40, (mean age average 25.7 years, SD = 5.7), including 7 women, participated in this study. The participants were students or university professors and participated free of charge. Their non-indigenous origin was determined by a questionnaire wherein participants identified themselves as not coming from any Chilean ethnic group or from indigenous families, and not having any indigenous surname. Signed consent was obtained from all subjects accepting their participation in the study and ensuring protection of their identities in accordance with the Helsinki Declaration. All were right-handed, without any visual alterations or psychopathology. Three participants were excluded from the study due to a low number of total errors (less than five valid ERNs trials, the minimum required for an accurate ERP estimation of each category).

Procedure

We used a race-IAT previously tested in mapuche-chileans stimuli comparison [14, 15, 19]. Participants listened to the instructions and then sat in front of a computer with the electrodes placed on their heads and responded by pressing with each forefinger one of two keys on a response pad (see Fig. **1**). Initially, participants were informed that the study assessed recognition processes and opinions regarding people and words. Faces were displayed for 100 milliseconds and words for 300 milliseconds. The category names were 'Indigenous', 'Non-indigenous', 'Positive' and 'Negative'. Stimuli were centered horizontally and vertically on the screen. Incorrect responses were indicated with an 'X' in the central part of the screen that appeared immediately after the participants responded.

For each block, categories requiring a response were displayed in the top left- and right-hand corners. Those labels indicated the stimuli categorization based on membership (ingroup or outgroup stimuli) or valence (positive vs. negative stimuli). A block of IAT trials which implies a response to positive words and non-indigenous faces with one key by one side; and to negative words and indigenous faces with another key, implies a 'compatible with prejudice' block task. In this report we only considered the IAT errors in our analysis. Elicitation of errors allowed us to measure simultaneously the MFN associated to motor responses and feedback of errors. We computed the membership group errors (ingroup faces categorized as outgroup or vice versa) as well as valence errors (positive stimuli classified as negative and vice versa). In addition, we consider errors present in the task that is compatible with the participant's race perspective (ingroup/positive and outgroup/negative) and errors present in the task that is incompatible (ingroup/negative and outgroup/positive).

Following the model of Greenwald and colleagues [21], trial sequences were divided into practice block combinations (a photograph, a word and the combination of photograph and word) and register blocks (picture and word task), balancing to left or right the membership and valence associations. Once the experiment was finished, the participants were thanked, and the research goal was thoroughly explained to them.

In ERN paradigms the feedback is usually presented after the response with a delay of 200-1000ms. When feedback appears quickly after motor response, a greater temporal anticipation of FRN should be seen [22]; especially in rapid forced-choice tasks. In our case, in order to maintain the classical design of IAT, we included the immediate feedback (a red "X") in the moment that the participant pushes the

response button. We don't use then a time delay between the response and the feedback, so both negativities (motor-locked and feedback-locked) were expected in the same ERPs: early MFN associated to the response (0-50ms) and a late MFN associated to feedback (110-220ms). Then the motor response is triggered simultaneously (i.e., without delays) with the feedback, and those components (ERN and FRN) become overlapped in the same temporal window. Fig. (**1**) shows the apparition of the negative feedback simultaneous to the motor response. Because ERN (motor) and FRN (feedback) had never been elicited simultaneously in the same time window we were not clear about what differences are expected between both components. Nevertheless, since the motor response can occur without subjective awareness of the error [23] we expected full automatic effects in this time window. On the other hand, since previous studies suggests that FRN is more related to the predictability of an action's outcome [24], we expected that outgroup stimuli would be processed as a more uncertain outcome, which would modulate the FRN amplitude through a lack of familiarity or alternatively due to ingroup favoritism.

Electrophysiological Recording

Signals were recorded on-line using a GES300, 129-channel system with HydroCel Sensors (Electrical Geodesic, Inc.) with a DC coupling amplifier, 24-bit A/D converter, 200 MΩ input impedance, 0.7 µV RMS/1.4 µV pp noise and NetStation[TM] software. Analog filters excluded information outside the 0.1 - 100 Hz frequency band. A bandpass digital filter between 0.5 and 30 Hz was later applied to remove unwanted frequency components. Signals were sampled at 500 Hz.

During recording, captured signals were referenced to the vertex. Afterward, signals were re-referenced off-line to electrodes average. To monitor vertical and horizontal ocular movements (EOG), two bipolar derivations were used. Continuous EEG data were segmented from 200ms before the stimulus and 800ms after it. All segments with eye movement contamination were excluded from further analysis, using an automatic method for removing eye-blink artifacts and visual procedure [25]. Artifact-free segments were aver-

aged to obtain ERPs. ERP waveforms were averaged separately for each experimental condition. EEGLAB Matlab toolbox and T-BESP software (http://neuro.udp.cl/software) were used for EEG off-line processing and analysis.

Statistical Analyses

Behavioral Measures (IAT Errors)

An implicit racial bias association was calculated for each subject based on reaction times obtained from IAT errors (when a negative feedback -"X"- was triggered). In order to obtain accuracy for IAT errors, correct responses were divided by the total number of responses in each category (faces, words and compatible-incompatible tasks) yielding an accuracy score between 0 and 1. Accuracy was calculated based on (a) compatible and incompatible blocks; (b) on the separate responses to faces (ingroup vs. outgroup) and; (c) words (pleasant and unpleasant). A repeated-measure ANOVA was applied to determine the tendencies of the results. For significant effects (α = 0.05), averages and contrasts were calculated using Tukey's *post-hoc* tests.

ERPs

Two main sites associated with ERN and FRN (FCz and Fz: electrodes E11 from the front midline site and electrode E6 at the center-frontal position) were considered to represent the ERP components. Although the figures show the ERPs grand averages for each group, all statistical calculations were performed using individual participants' data. For the statistical analysis of each component, the average was taken from the 0-50ms (response) and 110-152ms (feedback) temporal window at the FCz location. In addition, the same early response (0-50ms) and the 126-220ms (feedback) temporal windows were taken at the Fz location. For each component, ANOVA of repeated measures with the following within-subject factors was performed: Membership (INGROUP vs. OUTGROUP), and Valence (POSITIVE vs. NEGATIVE). In addition, data were reorganized in another within-subject factor, Task. This factor involves two levels (COMPATIBLE vs. INCOMPATIBLE). Those effects are following detailed:
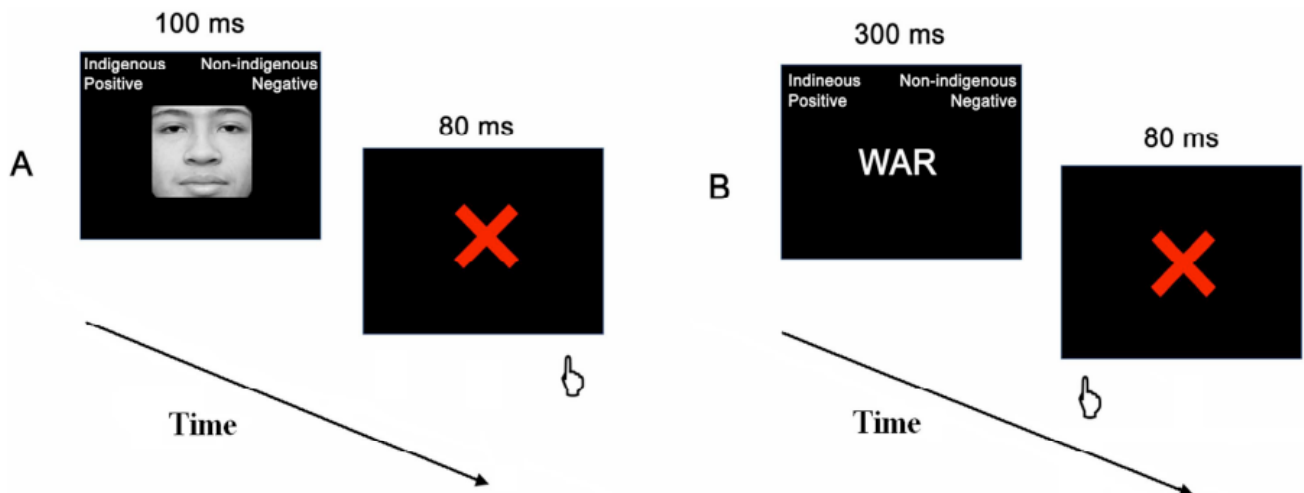

Fig. (1). IAT Paradigm. Examples of sequence errors (from faces or words) in the task. Presentation of the stimulus on the first slide (Face [in **A**] or Word [in **B**]); participant's motor response (second slide); finally, the error feedback (red X) after an incorrect response. Note that in the actual experiment the labels were presented in Spanish language (i.e., *Mapuche* for Indigenous, and *Chileno* for Non-indigenous)

1. Membership Effects

In this study, we analyzed this factor as the misclassification of ingroup or outgroup stimuli. Those errors implicate the incorrect classification of the ingroup stimuli (or non-indigenous) as outgroup (indigenous); and vice versa (an indigenous-outgroup stimulus classified as non-indigenous outgroup). In brief, we call those errors Ingroup vs. Outgroup errors.

2. Valence Effects

In this study, we analyzed this factor as the commission of errors based on emotional information. Those errors are defined as when a positive stimuli is classified as negative, or at reverse, when a negative is classified as a positive. Therefore, we compare in this factor two kinds of errors: Positive vs. Negative.

Possible interactions between Membership and Valence imply four categories of misclassification: errors performed on task which implies ingroup/positive association; outgroup/negative association; ingroup/negative association and outgroup/positive association.

3. Task Compatibility

In this study, we analyzed this factor as the errors made in the compatible task (blocks of the IAT which contains the association ingroup-negative and outgroup-positive) with respect to incompatible task (block of the IAT which contains the association ingroup-positive and outgroup-negative). Here we compare the results of the compatible vs. incompatible blocks.

Results were corrected using Greenhouse-Geisser and Bonferroni's methods to adjust the unvaried output of repeated measures ANOVA for violations of the compound symmetry assumption. The Tukey HSD test was used to calculate *post-hoc* contrasts.

RESULTS

Behavioral Results

The non-indigenous group displayed a relatively good general accuracy (M= 0.69; SD = 0.017). In response to pleasant words, an accuracy value of 0.72 (SD = 0.02) was obtained, while for unpleasant words, the accuracy value was 0.75 (SD = 0.02). With respect to faces, participants displayed an accuracy of 0.85 (SD = 0.02) in response to indigenous faces and of 0.81 (SD = 0.02) in response to non-indigenous faces. As expected, non-indigenous participants responded to the compatible task with higher accuracy (M = 0.84, SD = 0.01) than to incompatible tasks (M = 0.83; SD = 0.01; $F_{(1, 15)}$=9.828, p < 0.000).

ERPs

The two main sites where MFN exhibits its maxima were considered in the analysis.

Fz: Frontal Midline Electrode

ERN: Zero-50ms temporal window.

Membership and Valence Effects

The first temporal window that was selected (0-50 milliseconds, considering 0ms as the motor response moment)

displayed no membership significant differences between the ingroup (-0.23 uV, SD=0.784) and the outgroup (-0.74 uV, SD=0.876), with an $F_{(1,15)} = 1.272$ (p = 0.358). No main effect of Valence factor was shown. Nevertheless, an interaction between Membership and Valence was observed: $F_{(1, 15)} = 9.939$, p < 0.0001. Post-hoc comparison of this interaction revealed significant differences between the positive outgroup association (-0.067 uV, SD=0.115) and the positive ingroup association (-1.04 uV, SD=0.48, p<0.05) and also between the negative ingroup (0.01 uV, SD=0.226) and positive ingroup association (-1.04 uV, SD=0.48, p < 0.05); showing only the positive ingroup category association differences with all the other categories.

Task Compatibility

When compatible (-1.35 uV, SD=1.151) and incompatible (0.11 uV, SD= 0.362) tasks were considered, significant differences were found ($F_{(1,15)}$=6.409, p<0.05), with responses to the compatible category exhibiting a more negative amplitude (see Fig. **2B**).

FRN: 126-220ms temporal window

Membership and Valence

Statistically significant differences were found for Membership, between the outgroup (0.64 uV, SD=1.080) and the ingroup (1.95 uV, SD=0.934) conditions ($F_{(1,15)}$= 4.904; p < 0.05), with the ingroup presenting a larger positive amplitude (see Fig. **2C**). No effects of Valence were found, but a iteration between Membership and Valence was shown (F(1, 15) = 7.232 p < 0.001). Post hoc contrast performed over this interaction revealed differences between the positive outgroup (-0.85 uV, SD=1.26) and negative outgroup (1.32 uV, SD=1.061, p < 0.05) and also between the negative ingroup (2.29uV, SD=0.419) and positive outgroup (1.42uV, SD=0.854, p < 0.05). As a general pattern, each positive subcategory elicited smaller positive amplitude than the negative subcategories.

Task Compatibility

For the second temporal window (126-220ms, associated with the feedback), no significant differences were found when compatible (1.43 uV, SD=0.814) and incompatible (0.69 uV, SD=0.797; $F_{(1,15)}$= 0.453; p > 0.507) tasks were considered.

FCz: Fronto-Central Midline electrode

ERN: Zero-50ms temporal window.

In this electrode no differences were shown at this time window for any of the factors involved in this study.

FRN: 110-152ms temporal window

Membership and Valence

As presented in Fig. (**2A**), for the 110-152 millisecond window, comparison of the ingroup categories (2.77 uV, SD= 0.799) and outgroup categories (0.414 uV, SD=0.841) revealed significant main differences of membership effects ($F_{(1, 15)}$=7.342, p=0.010). No effects of Valence were shown. Once again, a Membership X Valence interaction effect was shown (F(1, 52) = 4.343476, p < 0.01). Post hoc comparisons of this interaction also showed significant differences. These were evidenced by comparison of the positive out-
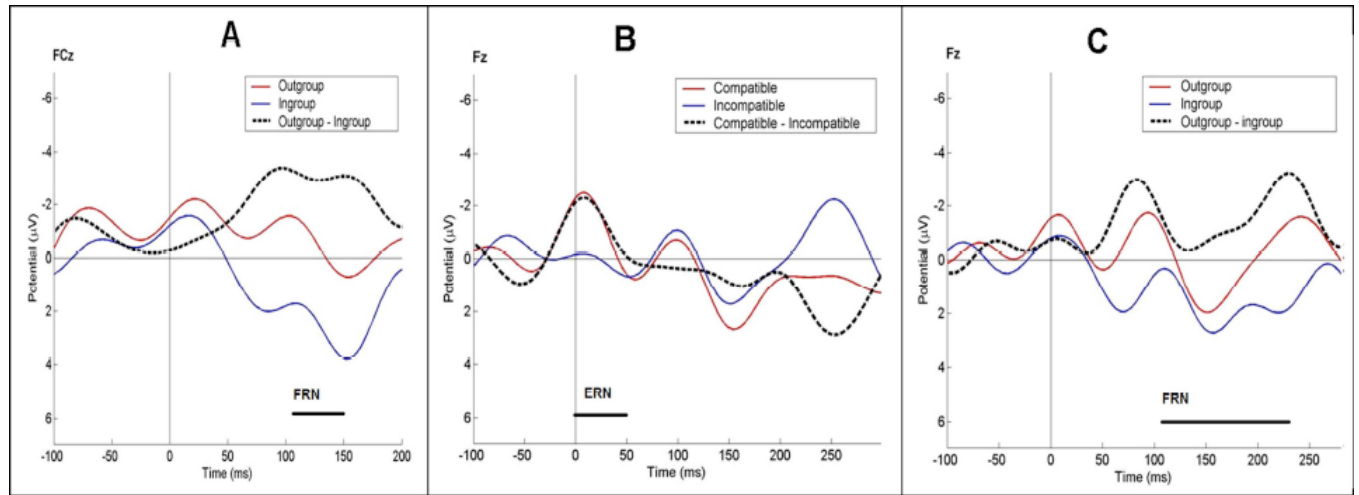
Fig. (2). ERP graphics for two electrodes with a Fronto-medial negativity. FCz: Frontal Central Midline; Fz: Frontal Midline. **A**) and **C**): ERP's for FCz and FZ electrodes, comparing outgroup conditions (red line) with ingroup conditions (blue line). **B**): FZ Comparison between Compatible (red line) and Incompatible (blue line). The dotted line corresponds to a subtraction between the respective categories. The time windows where the significant statistical analysis was performed are highlighted with a line.

group (-0.38 uV, SD= 0.831) with the positive ingroup (2.88uV, SD=0.583, $p < 0.05$) and of the positive outgroup (-0.38 uV, SD= 0.831) with the negative ingroup (2.26uV, SD=0.791, $p < 0.01$), demonstrating a larger negativity in the positive indigenous subcategory.

Task Compatibility

When compatible (1.25 uV, SD= 0.933) and incompatible (1.38 uV, SD=0.738) tasks were compared, no significant differences of membership factor were found ($F_{(1.15)}$=0.778, p=0.384).

DISCUSSION

Behavioral data

Participants displayed a higher level of accuracy for the compatible task (association favorable to the ingroup) than for the incompatible task (association favorable to outgroup). This result suggests that subjects' errors on these tasks may be produced by an implicit association between ethnic markers and valence, and that they may possibly result from a racial bias that leads to ingroup favoritism and outgroup devaluation.

ERPs

The MFN was associated in a general way with the conflict detection process and/or with expectation violations [8], which are, in this case, triggered by motor response and feedback [6]. In the present study, ERPs were modulated mainly by variables such as valence association and ethnic markers, both of which generated differences in the associated ERPs amplitudes.

In the current study, the topological location of the obtained ERN and FRN corresponds to the medial frontal area of the scalp, which is congruent with the general models that we find in the literature addressing conflict detection sources [6]. Brain areas such as the anterior cingulate cortex have been most often found to be related to the detection and interpretation of cognitive conflicts of this nature [26]. Our results specifically show that when making a decision about

another ethnic group, a subject's belonging to a specific ethnic group modulates the magnitude of electrophysiological activity that occurs, generating expectations and conflicts of different MFN magnitude, which behaves as a detection and processing marker of the error.

The electrophysiological correlates of IAT errors reported in this work are novel and relevant in the context of social psychology. Consistent with a previous report of medial-frontal negativity (MFN) and social bias [7], we found negativities associated with membership effects. Our results demonstrate that neural mechanisms of conflict detection are sensitive to racial stereotypes, opening a new field of research within social implicit associations and using IAT as a paradigm. On the other hand, the compatible task led to a larger early negativity (0-50 milliseconds) associated with the motor response compared with the incompatible task. While these data seem to be the opposite of what was expected (i.e., the incompatible task, being contrary to the participant's perspective, might be expected to generate a larger negativity), we think that this is due to a modulation based on expectations. In other words, the errors committed in this task are less expected by the participants, resulting in a larger gap between expectation and perception. This is evidenced in the interaction of membership and valence, where the presence of positive outgroup effects show the largest negativities. Although these results are somewhat controversial and will require future research to clarify their explanation, they clearly suggest that the detection process of IAT association may operate below the level of awareness and therefore that it does not necessarily rely on conscious reflection.

The IAT constitutes a very efficient method and is considered as a more robust measurement than the explicit evaluation [27, 28]. The MFN, particularly the ERN, doesn't depend on the explicit activity or awareness [29], and is not related to error detectability [30], then the MFN constitutes an implicit measurement complementary to the IAT score.

An issue to be considered is the relationship between MFN's elicitation and the activity when no error is present.

With respect to this parameter, it is important to ascertain that electrophysiological responses to correct answers were not included in the analysis because they are not strictly comparable to ERPs. This is because presenting an explicit feedback (an X mark) for an error, which is not present in correct answers, generates a time-locked answer that is specific to this stimulus and that would be expected to influence conflict monitoring. Perceptual differences seem to alter the FRN component. Since that FRN is modulated by the deviation of feedback stimuli from a perceptual template [31], it does not allow one to compare a response with negative feedback ("X") to a response without a perceptual stimulus (correct response without feedback).

As far as we are aware, although there are some reports registering both ERN and FRN negativities, no research has simultaneously measured error and feedback responses within a single temporal window. Our data suggest that although the early processes (ERN) and later processes (FRN) are modulated by similar stimuli, they may also show differential responses. Several studies have shown that the ERN and concomitant error-related activations of the medial frontal cortex in fMRI can occur without subjective awareness of the error [23]. The early ERN modulation in the Compatible/Incompatible task described in this study suggests that the task is highly automatic and independent of conscious control, as has previously been suggested by IAT behavioral versions [21]. This result is quite interesting, since the discrimination of the Compatible/Incompatible task implicitly suggests the complex conjugation of racial clues (i.e., ingroup/outgroup ethnic markers) and a semantic-associated valence (positive or negative words).

The observation that FRN is modulated exclusively by membership in a given group suggests that it reflects a process that is closer to conscious awareness because no modulation based on the implicit association between the combination of ingroup/outgroup stimuli and positive/negative valence was shown. Interestingly, Holroyd and colleagues [32] reported increased ERN, but nearly absent FRN amplitudes, in conditions with high predictive values. On the other hand, FRN amplitude variations are inversely related to the predictability of an action's outcome [24]. This suggests that FRN based on membership modulation is related to a higher uncertainty with respect to outgroup stimuli, which generate higher negativity. Future studies will be needed to determine whether the observed uncertainty is related to a lack of familiarity [33] or to ingroup favoritism [34]. It is also important to highlight that the patterns of ERN were modulated by implicit bias, according to IAT's paradigm, strengthening the conclusion that this marker is sensitive at an automatic level. For research of this type, which considers social bias, conflict detection would operate at a highly grounded or automatic level, which would not necessarily depend on the complex conscious monitoring of action, according to Amodio and colleagues [7]. In this context, the present work helps us significantly to understand social bias in a cultural context.

In agreement with previous IAT scores reported in race-IAT in Mapuches studies [14, 15, 19] the errors in the IAT were behaviorally and electrophysiologically consistent with the racial bias present in this racial-IAT score. Nevertheless, since MFN are not locked to the IAT stimuli but to the motor response which elicits an error, direct conclusion about IAT scores and MFN error modulation cannot be assessed. Future studies with a similar design can test if conflict detection on specific types of errors predicts lower IAT bias scores.

Like the results of Amodio and colleagues [7] our results show that ERN is a sensible component to errors associated with racial bias. More importantly, our results highlight important contributions:

1. Behavioral and Neural Markers of IAT Errors

Although some studies have suggested that feedback can be an important source of information (i.e., social desirability effects becomes less effective when feedback is suppressed: [35]), usually the IAT errors are not considered in the reports of classic and modified IATs [36, 37]. Our result shows clearly that behavioral measures of errors are very informative for the social bias. More important, we report the first electrophysiological correlate of IAT errors, which is modulated by social bias. The behavioral and electrophysiological measurements of the errors are simple and robust measurements which can be very informative of the implicit social and general bias. Further, the existence of an MFN in the IAT task indicates that one can unify two very distinct areas of investigation and would provide a new, interesting way to understand the manner in which the brain processes errors in IAT. For example, the reinforcement learning theory of the MFN [29], constitutes a possible background to interpret the errors in the IAT task since the MFN reflects a high-level error-processing system in humans [32].

2. Early Contextual Effects of Valence Present in the IAT Errors

Our results do not show a principal effect of valence. However, many effects of interactions between valence and membership factors were found. Those effects suggest that the valence is processed in a differential and contextual way. The same information (i.e. positive) is processed differently depending on if it is contextually linked with an outgroup or an ingroup association. This is consistent with research on contextual malleability of attitudes [38, 39]. More important, it is not only the valence but also the specific contextual association which seems to be highly automatic. The contextual modulation of valence is evidenced in an early step of processing (0-50ms). Growing information from several ERPs studies show early contextual effects of valence [39, 19, 35] supports the previous assertion.

3. Improvement of IAT/ERPs Recent Findings

Recent studies of IAT recording with ERPs have been [15, 18, 19, 40]. The problem with those previous studies lies in difficulty in adapting IAT to ERPs settings (i.e., eye movements due to the labels, the facts that faces vs. words elicits differences ERPs effects, effects of the blocks on hand responses, high number of loss trials, etc). The possibility of obtaining simple measurement based only on the commission of the errors through the ERN implies a direct shortcut to the cerebral correlates of the IAT, and avoids the methodological confounds that arises when IAT stimuli is the trigger to ERPs. Our study evidences that the ERN present in the IAT is a direct and simple method to investigate the implicit effects of IAT in racial research (social ingroup re-

search). Future investigation would test other social psychology domains in attitude research.

4. IAT Controversies

On the other hand, the IAT results have been recently criticized. The scores of IAT are relative in that it is impossible to interpret the absolute value and sign of a score (positive Vs. negative), and his score doesn't depend only of the category used, but as well of the particularity of the characteristics of the stimuli used (i.e., similarity, saliency, stimuli type, etc) and of other task-dependent aspects [41]. Because of that, the errors of the IAT (and then of the MFN) can serve as complementary measurements to establish several process that determines the IAT response. Further studies manipulating IAT attributes and measuring the MFN should be considered in the future. The convergence of measures assessing the ingroup favoritism at different levels of analysis (behavioral and electrophysiological) can help the co-construction of inter-level theories [42, 43].

CONCLUSION

Our results suggest that several processes interact in various ways at early stages of neural processing, and that different combinations of stimuli probably present dynamic patterns of association. The contribution of electrophysiological, multi-level studies would facilitate advances toward a dynamic [43] and more detailed racial process using the IAT. The possibility of assessing those differential effects using ERPs in IAT research calls for further inquiries. The combination of these instruments and types of measurements may open a new research area in social neuroscience, contributing to the investigation of electrophysiological correlates of implicit associations in social attitudes research.

ACKNOWLEDGMENTS

REFERENCES

[1] Yeung N, Botvinick MM, Cohen, JD. The neural basis of error detection: conflict monitoring and the error related negativity. Psychol Rev2004; 111: 931-59.

[2] Falkenstein M, Hohnsbein J, Hoormann J, Blanke L. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr Clin Neurophysiol 1991; 78: 447-55.

[3] Gehring WJ, Goss B, Coles MG, Meyer DE, Donchin E. A neural system for error detection and compensation. Psychol Sci1993; 4: 385-9.

[4] Miltner W, Braun CH, Coles MGH. Event-related brain potentials following incorrect feedback in a time estimation task: evidence for a 'generic' neural system for error detection. J Cognit Neurosci1997; 9: 788-98.

[5] Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD. Errors in reward prediction are reflected in the event related brain potential. Neuroreport 2003; 14: 2481-4.

[6] Luu P, Flaisch T, Tucker DM. Medial frontal cortex in action monitoring. J Neurosci 2000; 20: 464-9.

[7] Amodio DM, Harmon-Jones E, Devine PG, Curtin JJ, Hartley SL, Covert AE. Neural signals for the detection of unintentional race bias. Psychol Sci 2004; 15: 88-93.

[8] Carter CS, Braver TS, Barch DM, Botvinick MM, Jonathan DN, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 1998; 280: 747-9.

[9] Greenwald AG, Banaji MR. Implicit social cognition: Attitudes, self-esteem, and stereotypes. Psychol Rev 1995; 102: 4-27.

[10] Greenwald AG, McGhee DE, Schwartz JLK. Measuring individual differences in implicit cognition: The implicit association test. J Pers Soc Psychol 1998; 74: 1464-80.

[11] Dasgupta N, McGhee DE, Greenwald AG, Banaji MR. automatic preference for white americans: eliminating the familiarity explanation. J Exper Soc Psychol 2000; 36: 316-28.

[12] Dunham Y, Baron AS, Banaji MR. From American City to Japanese Village: a cross-cultural investigation of implicit race attitudes. Child Devel 2006; 77: 1268-81.

[13] Bohner G, Siebler F, González R, Haye A, Schmidt EA. situational flexibility of ingroup-related attitudes: a single category iat study of people with dual national identity. Group Processes Intergroup Relat 2008; 11: 301-17.

[14] Brown R, González R, Zagefka H, Manzi J, Cehajic S. Nuestra culpa: collective guilt and shame as predictors of reparation for historical wrongdoing. J Pers Soc Psychol 2008; 941: 75-90.

[15] Ibanez A, Gonzales R, Hurtado E, Haye A. Perceptual and semantic blending of ingroup/outgroup facial stimuli and word valence: early N170 and late LPP electrophysiological evidence. Pychophysiology 2008; 45: S101.

[16] Dillehay T. Monuments, empires, and resistance: the araucanian polity and ritual narratives. UK: Cambridge Studies 2007

[17] Saiz JL. In Morales JF, Páez D, Kornblit AL, Asún D, Ed. Psicología social. Atribución de estereotipos: Los indígenas mapuches que perciben los chilenos. Settlement of stereotypes: The indigenous Mapuches perceived Chileans. Buenos Aires: Prentice Hall-Pearson Education. 2002; pp. 145-51.

[18] Hurtado E, Haye A, Gonzalez R, Manes F, Ibanez. Contextual blending of ingroup/outgroup face stimuli and word valence: LPP modulation and convergence of measures. BMC Neurosci 2009; 101:69.

[19] Ibáñez A, Hurtado E, González R, Haye A, Manes F. Neural markers of early contextual blending: N170 modulation of ingroup/outgroup relative position and associated valence. Front Hum Neurosci 2010; 4: 188.

[20] De Houwer J, Teige-Mocigemba S, Spruyt A, Moors A. Theoretical claims necessitate basic research: reply to Gawronski, Lebel, Peters, and Banse 2009 and Nosek and Greenwald 2009. Psychol Bull 2009; 1353: 377-9.

[21] Greenwald AG, Nosek BA, Banaji MR. Understanding and using the implicit association test: I. An improved scoring algorithm. J Pers Soc Psychol 2003; 85: 197-216.

[22] Goyer JP, Woldorff MG, Huettel SA. Rapid electrophysiological brain responses are influenced by both valence and magnitude of monetary rewards. J Cogn Neurosci 2008; 2011: 2058-69.

[23] Endrass T, Reuter B, Kathmann N. ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. Euro J Neurosci 2007; 6: 1714-20.

[24] Heldmann M, Rüsseler J, Münte TF. Internal and external information in error processing. BMC Neurosci 2008; 9: 33.

[25] Gratton G, Coles MGH, Donchin E. A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 1983; 55: 468-84.

[26] Pardo JV, Pardo PJ, Janer KW, Raichle ME. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proceedings of the National Academy of Sciences of the United States of America 1990; vol. 87: pp. 256-259.

[27] Asendorpf JB, Banse R, Mücke D. Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. J Pers Soc Psychol 2002; 832: 380-93.

[28] Hofmann W, Rauch W, Gawronski B. And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. J Exper Soc Psychol 2007; 43: 497-504.

[29] Nieuwenhuis S, Holroyd CB, Mol N, Coles MG. Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. Neurosci Biobehav Rev 2004; 284: 441-8.

[30] Maier M, Steinhauser M, Hübner R. Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. J Cogn Neurosci 2008; 2012: 2263-73.

[31] Liu Y, Gehring WJ. Loss feedback negativity elicited by single-versus conjoined-feature stimuli. Neuroreport 2009; 206: 632-6.

[32] Holroyd CB, Coles MG. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 2002; 109: 679-709.

[33] Kinoshita S, Peek-O'Leary M. Two bases of the compatibility effect in the Implicit Association Test IAT. Quart J Exper Psychol 2006; 59: 2102-20.

[34] Efferson C, Lalive R, Fehr E. The coevolution of cultural groups and ingroup favoritism. Science 2008; 321:1844 -9.

[35] Olofsson JK, Nordin S, Sequeira H, Polich J. Affective picture processing: an integrative review of ERP findings. Biol Psychol 2008; 773: 247-65.

[36] Banse R, Seise J, Zerbes N. Implicit attitudes towards homosexuality: reliability, validity, and controllability of the IAT. J Exper Psychol 2001; 482: 145-60.

[37] Greenwald AG, Farnham SD. Using the implicit association test to measure self-esteem and self-concept. J Pers Soc Psychol 2000; 796:1022-38.

[38] Dasgupta N, Greenwald AG. On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. J Pers Soc Psychol 2001; 815: 800-14.

[39] Rudman LA, Ashmore RD, Gary ML. "Unlearning" automatic biases: the malleability of implicit prejudice and stereotypes. J Pers Soc Psychol 2001; 815: 856-68.

[39] Codispoti M, Ferrari V, Bradley MM. Repetition and event-related potentials: distinguishing early and late processes in affective picture perception. J Cogn Neurosci 2007; 194: 577-86.

[40] Gonzalez R, López V, Haye A, Hurtado E, Ibañez A. N170 and LPP discrimination of same race versus other race facial stimuli and positive and negative words in indigenous and non-indigenous participants. Clin Neurophysiol 2008; 119: 155.

[41] De Houwer J, Teige-Mocigemba S, Spruyt A, Moors A. Implicit measures: a normative analysis and review. Psychol Bull 2009; 1353:347-68.

[42] Cosmelli D, Ibáñez A. Human cognition in context: on the biologic, cognitive and social reconsideration of meaning as making sense of action. Integr Psychol Behav Sci 2008; 422: 233-44.

[43] Ibáñez A, Haye A, González R, Hurtado E, Henríquez R. Multilevel analysis of cultural phenomena: The role of ERP approach to prejudice. J Theor Soc Behav 2009; 39: 81-110.