Electrophysiological Characterization of Prejudice in Actors of the Colombian Armed Conflict

Jhon Jair Quiza Montealegre

Tesis para optar al título de Doctor en Ingeniería Electrónica y de Computación

Director
José David López, PhD

Codirector
Andrés Quintero Zea, PhD

Universidad de Antioquia
Facultad de Ingeniería
Doctorado en Ingeniería Electrónica y de Computación
Medellín
2024

Cita	(Quiza-Montealegre, 2024)
Referencia **Estilo APA 7 (2020)**	Quiza-Montealegre (2024). *Electrophysiological Characterization of Prejudice in Actors of the Colombian Armed Conflict* [Tesis doctoral]. Universidad de Antioquia, Medellín, Colombia.

Doctorado en Ingeniería Electrónica y de Computación, Cohorte XXII.
Grupo de Investigación SISTEMIC

Centro de Documentación de Ingeniería

Repositorio Institucional: http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: Jhon Jairo Arboleda Céspedes.
Decano Julio César Saldarriaga Molina.
Coordinadora de Posgrados: Natalia Gaviria Gómez.

Dedicatory Note

To my wife Lina, my son Miguel Ángel, and my daughter María José.

Acknowledgements

The author would like to express his gratitude to the following people and institutions that, in one way or another, helped me to successfully complete this challenge that I set myself four years ago.

To the University of Medellin for supporting me financially and giving me the necessary time to complete this doctorate.

To my directors, Professors José David López and Andrés Quintero, for all their teachings and emotional support. They were my primary support to do everything I could do.

To Professor Natalia Trujillo for her advice that helped me better understand the psychological aspects of my research work. In this sense, I would also like to thank research psychologist Stella Valencia.

To the director of the Research Program "Restaurar tejidos sociales en comunidades durante el postconflicto a través de un enfoque de intervención biopsicosocial comprehensivo: estrategias hacia a la construcción de paz en Colombia", Professor Juan Ugarriza, and to the principal investigators of the projects that make up the Program, for allowing me to use their data.

To my internship tutor, Professor Juan David Martínez, for everything he taught me about iML.

To my coworkers at the University of Medellin for their encouragement.

Finally, and very specially, to my wife Lina, for giving me strength when I had none left.

Índice

Lista de tablas

Lista de figuras

Resumen

Colombia ha experimentado el conflicto armado más prolongado en América Latina, resultando en numerosas víctimas. Los esfuerzos para desescalar este conflicto, incluyendo los acuerdos de paz con los paramilitares en 2002 y con las guerrillas de las FARC en 2016, han tenido resultados mixtos debido a los prejuicios persistentes entre los antiguos actores. Esta tesis se centra en caracterizar estos prejuicios para diseñar mejores intervenciones psicosociales para la reconciliación. El estudio caracteriza los patrones electrofisiológicos asociados a una prueba psicológica que evalúa el prejuicio entre los antiguos actores del conflicto. Involucra tres etapas: análisis de señales electrofisiológicas en el dominio del tiempo, análisis en el dominio de la frecuencia utilizando teoría de grafos, e integración de características electrofisiológicas con otros datos para entrenar un modelo de aprendizaje automático interpretable. En la primera etapa, se desarrolló una nueva metodología de análisis EEG-ERP utilizando inferencia bayesiana para evaluar el prejuicio a través de una tarea IAT. Los hallazgos indicaron mayor actividad cerebral en los participantes con prejuicio contra las víctimas en comparación con aquellos sin prejuicio o con prejuicio contra excombatientes. La segunda etapa introdujo un nuevo método de análisis de conectividad funcional basado en EEG, revelando diferencias en las configuraciones de redes cerebrales entre los antiguos actores del conflicto. Los resultados mostraron que las víctimas y los exparamilitares tenían más prejuicio contra las víctimas, mientras que los civiles y los exguerrilleros tenían más prejuicio contra los combatientes. Además, las víctimas y los exguerrilleros regulaban más eficazmente el prejuicio contra las víctimas, y los exparamilitares hacían lo mismo con los combatientes. En la etapa final, un modelo de aprendizaje automático interpretable identificó características clave para caracterizar y reclasificar sujetos, utilizando datos demográficos, conductuales y electrofisiológicos. Cinco características clave fueron suficientes para discriminar entre grupos y determinar las necesidades de reclasificación. Esta investigación proporciona una comprensión integral de los prejuicios persistentes entre los antiguos actores del conflicto en Colombia, ayudando a los psicólogos a diseñar estrategias de intervención específicas para la reconciliación.

Abstract

Colombia has had the most prolonged armed conflict, with the highest number of victims in Latin America. Since the beginning of the century, multiple efforts have been made to de-escalate it, such as the peace agreements with the paramilitaries in 2002 and the

FARC guerrillas in 2016. However, efforts to de-escalate the conflict have had mixed results due to persistent prejudices among former actors. Therefore, it is necessary to characterize this prejudice to design more effective psychosocial intervention strategies to promote reconciliation. In this thesis, we present the characterization of electrophysiological patterns associated with a psychological test that assesses prejudice among former actors in the Colombian armed conflict. This characterization was done in three stages: Analyzing electrophysiological signals in the time domain, analyzing them in the frequency domain using graph theory, and merging electrophysiological features with other domain features to train an interpretable machine learning model. In the first stage, we developed a novel methodology for EEG-ERP analysis based on massive univariate statistical methods and Bayesian inference hypothesis testing. This methodology was used to analyze ERP related to an IAT task designed to assess prejudice, and we found that participants with prejudice toward one (victims) group exhibited higher activity than participants without prejudice or with prejudice toward the other group (ex-combatants). In the second stage, we developed a novel methodology for EEG-based functional connectivity analyses that adopted the techniques currently at the forefront of engineering and incorporated Bayesian inference hypothesis testing. This methodology was used to detect and measure differences among the configuration of the brain networks of former conflict actors; as a result, we found that victims and ex-paramilitaries generate more prejudice against victims, and civilians and ex-guerrillas generate more prejudice against combatants. However, victims and ex-guerrillas regulate more the prejudice against victims, and ex-paramilitaries regulate more the prejudice against combatants. All of these results are consistent with the results of the IAT task. In the third stage, we developed a novel methodology based on global and local analysis of interpretable machine learning models to identify the most important features in the characterization of groups and to evaluate the convenience of reclassifying subjects individually. We used demographic, behavioral, and electrophysiological features to characterize the groups of the former actors. As a result, we found that five characteristics of the 128 evaluated are sufficient to discriminate between groups of actors in armed conflict and to determine whether or not a participant should be reclassified. This research has allowed us to comprehensively characterize the phenomenon of prejudice that persists among former actors of the Colombian armed conflict, which will allow psychologists to design more specific intervention strategies to promote reconciliation among them.

Keywords. ERP, Functional connectivity, iML, Bayesian inference, Prejudice, Colombian armed conflict, IAT, Social cognition

1 Introduction

Colombia has been confronted with the most prolonged armed conflict in Latin America, leading to the highest victim count (Comisión de la Verdad, 2022). Numerous efforts have been made since the beginning of the century to mitigate the intensity of the conflict, including the establishment of peace agreements with paramilitary groups in 2002 (de Colombia., 2005) and with FARC guerrillas in 2016 (para la Paz, 2016). Within the framework of these agreements, several organizations have worked to promote reconciliation among the parties involved in the conflict and undertake the task of reconstructing social structures.

The results of these institutionsésfforts have been mixed for several reasons. One is that prejudices, which are negative evaluations of a particular social group and tend to generalize to its members (Amodio and Cikara, 2021), are relatively stable over a person's lifetime, particularly for people who have experienced traumatic situations, such as those involved in armed conflicts, as highlighted by (Bar-Siman-Tov, 2004).

From an evolutionary perspective, the origins of prejudice can be understood as adaptive mechanisms that our ancestors developed to navigate their complex social environments (Dunbar, 1992). Until recently in human history, social groups lived in small units that were genetically and culturally homogeneous. These living conditions supported specific adaptations still a part of our nature today. We show strong tendencies to cooperate with people whom we perceive to be "like us," and we are suspicious of strangers judging and discriminating against those who are "not like us" (Liberman et al., 2017). Evolutionarily, quickly categorizing other individuals into groups could be crucial for survival, facilitating swift decisions about who might be an ally or a threat (Dovidio et al., 2010). Several authors have extensively researched these mechanisms from a neuroscientific standpoint, exploring how brain structures such as the amygdala and the prefrontal cortex are involved in the formation of prejudices and in regulating responses to social stimuli (Rösler and Amodio, 2022; Saarinen et al., 2021).

Accordingly, it is necessary to characterize this prejudice to design more effective psychosocial intervention strategies to promote reconciliation (Ugarriza et al., 2019). Measuring and characterizing prejudice is challenging because people often try to hide it due to self-image and socially desirable responses (Teige-Mocigemba et al., 2010). Thus, using tools based on direct measures and explicit questions is not a successful approach to assess it (Teige-Mocigemba et al., 2010). Instead, many authors prefer a methodology proposed

by Greenwald et al. (1998) to assess implicit psychological processes, known as the Implicit Association Test (IAT), in which these cognitive operations are less prone to distortions.

In the framework of this thesis, IAT has been complemented with electroencephalography (EEG) to try to find phenotypes of prejudice. Phenotypes are physiological patterns consistent with the behavior of a subject within a homogeneous group that can be used to characterize a population in experimental conditions (Johnstone et al., 2005).

In recent studies, behavioral measures and electroencephalography (EEG) have provided valuable insights into how these brain processes manifest. Although different techniques have been used to find electrophysiological phenotypes of prejudice, such as frequency or time-frequency analysis (Kato et al., 2018), and reconstruction of cortical sources (Healy et al., 2015; Schindler et al., 2015), the event-related potentials (ERP) (Barnes-Holmes et al., 2004) technique is by far the most widely used and has provided the most information about the cognitive processing of prejudice. The use of ERP still predominates due to its high temporal resolution, noninvasiveness, low cost, and simplicity (Luck, 2014). Furthermore, ERP studies have pinpointed the exact moments when the brain reacts to prejudice-related stimuli, highlighting components such as N2 (Chen et al., 2018; Healy et al., 2015), P3a (Healy et al., 2015; Portengen et al., 2022), P3b (Chen et al., 2018), N400 (Healy et al., 2015; Williams and Themanson, 2011), and LPP (Forbes et al., 2012), which are associated with processes of social categorization detection, goal-directed attention to category, category conflict and response selection, categorization decision, and evaluation of categorization decision (Amodio and Cikara, 2021). These investigations not only demonstrate the complexity of the cognitive processes behind prejudice and social categorization but also open the door to more informed interventions to combat prejudice in modern society.

ERP analysis involves acquiring, preprocessing, and averaging EEG signals associated with a particular stimulus. The amplitude and latency of specific segments of ERP, namely ERP components, are quantified, and the data is further analyzed using null hypothesis significance testing (NHST) to determine if there are statistically significant differences between the experimental conditions (Luck, 2014). This method exhibits two main issues. First, the use of NHST has been criticized in psychology and neuroscience because it produces reproducibility problems (Open Science Colaboration, 2015), and in ERP analyses because it could make it challenging to detect spurious effects (Luck and Gaspelin, 2017). Secondly, determining the precise location, latency, and duration of ERP components is a complex task in practical applications due to the influence of various factors,

such as the characteristics of the stimulus and the demographic makeup of the sample being studied (Fields and Kuperberg, 2020).

An alternative approach involves massive univariate statistical methods (Maris and Oostenveld, 2007) to perform data-driven analyses. This method does not require preselecting which regions of interest (ROI) or time windows are to be analyzed, as this is done automatically (Groppe et al., 2011). Although these methods can be more effective in reducing Type I errors (Fields and Kuperberg, 2020), they are not reliable in detecting and quantifying effects (Sassenhagen and Draschkow, 2019), nor in eliminating the possibility of Type II errors (Luck and Gaspelin, 2017; Groppe et al., 2011).

In light of the limitations associated with traditional methodologies and the challenges posed by massive univariate statistical methods, our first objective in this thesis was to develop an alternative methodology that addresses these issues while characterizing an IAT task to measure prejudice among former actors in the Colombian armed conflict.

In the context of understanding cognitive processes related to prejudice, in addition to ERPs, it is essential to consider the synchronized participation of multiple brain regions. Previous research has established that the brain has distinct regions dedicated to carrying out certain cognitive functions (Fornito et al., 2016; Yuste, 2015). However, cognitive processing requires a concurrent interaction between several regions (Friston, 1994). In particular, according to Amodio and Cikara (2021), prejudice involves several cognitive processes in which many regions of the brain participate, such as the medial temporal lobe, anterior temporal lobe, medial prefrontal cortex, dorsal anterior cingulate cortex, and rostral anterior cingulate cortex.

Given this complex interplay of brain regions in cognitive processes and their relevance in understanding prejudice, we chose to use a different technique to ERP to analyze the brain networks involved in prejudice, which is EEG-based Functional Connectivity. The functional connectivity technique finds the predominant relationships among brain areas (Friston, 2011) and expresses these relationships in a connectivity matrix (Sporns, 2011). The resultant connectivity matrix can be characterized using theoretical graph analysis (Stam et al., 2007). Previous studies have used EEG-based functional connectivity, suggesting that training in social cognition can improve emotional recognition and modify brain connectivity (Quintero-Zea et al., 2019; Valencia et al., 2020; Trujillo et al., 2017). However, these investigations left the question of whether tasks focusing on more complex social processes, such as those associated with prejudice, could be affected. The second

objective of this thesis was to apply this technique to EEG data from former Colombian armed conflict actors during the IAT test, shedding light on the reorganization of brain networks in prejudice events and providing information for targeted psychosocial intervention strategies.

In the third stage of our investigation, our objective was to characterize a group of former actors for whom we had collected behavioral test results, EEG records, and demographic data. In such a characterization, we did not perform typical analyses such as ANOVA, chi-square tests, MANOVA, and MANCOVA because they are only adequate when the relationships between variables are linear or when there are few variables to analyze Fokkema et al. (2022). Instead, we employed machine learning algorithms to uncover complex patterns and relationships within the collected data. These methods allowed us to identify subtle nuances and interactions between variables that traditional analyses may have overlooked, providing a more comprehensive understanding of the former actors under investigation (Dehghan et al., 2022).

ML models are known for their high predictive capacity. However, they are also opaque, making it difficult to infer how they make their predictions or what features are more critical to their predictions (Yarkoni and Westfall, 2017). The complexity of models often makes it challenging to balance their performance and interpretability. Fortunately, data analysis techniques have emerged that make opaque machine learning models interpretable and allow models with high predictive capacity and interpretability to be obtained simultaneously (Molnar, 2022). These techniques, known as interpretable machine learning methods, aim to provide insights into the decision-making process of complex models. They often involve feature importance analysis, rule extraction, or model-agnostic explanations. Using these techniques, researchers and practitioners can better understand how the models arrive at their predictions and identify the key factors driving them. These techniques enhance interpretability and enable trust and accountability when using ML models in various domains.

Therefore, the third objective of the thesis was to create an interpretable ML model that would use demographic, behavioral, and electrophysiological data to describe the people involved in the armed conflict in Colombia. These analyses are essential for us because they can determine whether these subjects should be reclassified, considering that according to their group, they will receive personalized social-cognitive training designed to mitigate prejudice among Colombian armed conflict actors.

The fulfillment of these objectives has allowed us to obtain a comprehensive characterization of the prejudice phenomenon that persists among former actors of the Colombian armed conflict, which will allow specialists to design more specific intervention strategies to promote reconciliation between them.

1.1 Objectives

Formulate a methodology to identify electrophysiological phenotypes of prejudice among individuals exposed to the Colombian armed conflict.

1. Characterize an existing prejudice task synchronized with EEG data applied to a sample of people who have been exposed to the Colombian armed conflict and perform exploratory statistical analyses using EEG-ERP signal analysis techniques to obtain a baseline for comparison.
2. Propose and validate a functional connectivity model to determine levels of prejudice among individuals exposed to the Colombian armed conflict, based on EEG signal analysis techniques and statistical analyses.
3. Formulate and validate a computational-intelligence-based methodology to correlate the obtained EEG-based phenotypes of prejudice with the results of behavioral tests applied to the target population.

1.2 Outline

The chapters of this thesis are organized as follows.

Chapter 2 expounds on the materials and methods shared throughout several subsequent chapters. The present study characterizes a cohort of individuals previously involved in armed conflicts. These individuals were subjected to a series of behavioral assessments and electroencephalographic recordings. Subsequently, an overview is provided on the behavioral tasks undertaken by the participants, elaborating on the characteristics of the Implicit Association Test (IAT), a test specifically devised to assess prejudice exhibited toward distinct groups of individuals. Following this, the methodology used to preprocess EEG data and extract ERPs is outlined, which serves as the basis for all subsequent studies.

Subsequently, we elucidate the methodology employed to acquire the network metrics required to perform the EEG-based functional connectivity analyses expounded in Chapter 4, which constitute integral components of the attributes employed in the machine learning models expounded in Chapter 5. The following section explains the Bayesian inference hypothesis testing approach employed in Chapters 3 and 4. Finally, we briefly described the machine learning models tested to achieve the third specific objective and the feature selection and interpretability strategies employed in chapter 5.

Chapter 3 presents the methodology and results of ERP analyses to characterize the sample of participants classified by levels of prejudice according to the result obtained in the IAT test. For this analysis, we developed a new methodology based on massive univariate statistical methods and hypothesis testing using Bayesian inference that solves most of the problems of sensitivity, reliability, and reproducibility of traditional methods. The results obtained with this methodology were compared with those obtained using the traditional method of prior selection of ERP components and the method based on massive univariate analyses.

Chapter 4 presents the methodology and results of the EEG-based functional connectivity analysis to characterize participants separated by conflict actor groups. To our knowledge, this is the first time this technique has been used in studying social biases. From the results obtained, we discuss the relationships found between some measures of global connectivity and the IAT score, which could be indicators of electrophysiological correlates of prejudice.

Chapter 5 presents the interpretable machine learning model developed to characterize the groups of actors in the Colombian armed conflict based on their demographic, behavioral, and electrophysiological characteristics. Global interpretability analyses resulted in only five features sufficient to accurately classify former armed conflict actors. Also, for the first time in social psychology studies, local interpretability analyses were performed to identify features that may cause a subject to be reclassified and, consequently, to receive sociocognitive training to mitigate prejudice different from the one he/she would receive if he/she were not reclassified.

Finally, general conclusions, main contributions derived from this study, and future work are presented in Chapter 6.

2 Materials and Methods

This chapter describes the materials and methods used in this thesis. In the Materials section 2.1, we describe the participants and the behavioral tasks they performed. In the methods section, we explain how acquired and preprocessed the EEG signals, the traditional methodologies to analyze ERP, the hypothesis testing Bayesian inference method – a cornerstone of the methodologies proposed –, the methods necessary to obtain EEG measures with EEG-base functional connectivity, and described briefly the ML models used, the techniques to do feature selection and interpretability de ML models.

2.1 Materials

2.1.1 Participants

In this study, we recruited 92 healthy Colombian volunteers, 56 of whom were men, between 18 and 70 years old ($M = 36.90$, $SD = 10.72$) and with an average education of 10.12 years ($SD = 3.50$), classified into four groups: **exguerrillas** (22 participants), **exparamilitaries** (31 participants), **victims** (23 participants), and **civilians nonexcombatants and nonvictims** (16 participants). The sample was not based on an estimate but on convenience, by direct invitation to ex-combatants of illegal armed groups (e.g., guerrillas, ex-paramilitaries) through the Colombian Agency for Reintegration, to people officially declared as victims by the government through key actors in municipalities with a high number of armed conflict events, and to civilians (not directly involved in the conflict) with similar demographic characteristics to the ex-combatants and victims. Before registration, participants were asked about their medical and psychiatric histories. The exclusion criteria included a history of severe mental disorders such as schizophrenia, epilepsy, or severe head trauma. All psychological evaluations and EEG recordings were performed in classrooms of educational institutions in Antioquia, Colombia, and not in a laboratory under controlled conditions.

The participants were informed about the purpose of the study, the confidentiality of the information collected, and the procedures for psychological tests and electroencephalographic recordings. Evaluations began once the participants signed the consent form. The research procedures were approved by the Research Ethics Committee of the Universidad del Rosario (Minute DVO005-063-CS048, 8 February 2018). The informed consent docu-

ment signed by the participants contains the following points: a general description of the research project for which they are asked to participate, the objective of the study, the reason why they were selected, the risks and benefits of participating in the study; the description of the procedures they will perform; and, the guarantee of confidentiality of the information and anonymity of participation.

2.1.2 Behavioral Tasks

All participants underwent a series of behavioral tests thoroughly described below.

Implicit Association Test – IAT

The IAT is an experimental behavioral task developed by Greenwald et al. (1998) to measure implicit bias based on the principle that if a congruent association between two concepts (e.g., target and stereotypical attribute) is readily accepted as accurate by a decision maker (e.g., victim → negative), then the reaction time (RT) to categorize such associations as equivalent is shorter. On the contrary, if an incongruent association between two concepts (e.g., target and counter-stereotypical attribute) is not readily accepted as accurate (e.g., victim → positive), then the RT is comparatively longer due to inhibitory processes required to override an automatic tendency to associate congruent concepts (Healy et al., 2015). The IAT effect, which can also be considered a measure of implicit bias, is calculated using the standardized difference (D) between the mean reaction time and the congruent and incongruent pairings. If an individual has a positive D score, it implies that the person is slower to respond to incongruent pairings, faster to respond to congruent pairings, or both (Forbes et al., 2012).

Given the low level of study of the participants, we used a modified version of the original IAT design introducing auditory stimuli. The task comprises seven blocks, each with a fixed number of trials, as shown in Table 1. In each block, trial images are the same and appear in the same positions, and the stimuli are randomly selected from a set of pre-established audios. The task design - including the decision to use auditory stimuli and the selection of the screens and words - was made for the team of psychologists of the research and could be seen in detail in Baez et al. (2020).

The test screens presented to the participants are shown in Figure 1. In our experiment, the image on the left of Figure 1 (A) represents the concept of "ex-combatant." The image on the right represents the concept of "victim." The happy and sad face icons

Tabla 1

IAT task structure.

Block	Trials	Function
1	6	Concept practice
2	20	Valence practice
3	20	Congruent trials - Practice
4	40	Congruent trials - Test
5	6	Concept practice
6	20	Incongruent trials - Practice
7	40	Incongruent trials - Test

in Figure 1 (B) represent positive and negative valence, respectively. The audio "Ganar" ("Win" in Spanish) is the stimulus, which must be associated with the happy face (good valence), for which the participant has up to three seconds to press the shift key on the left side of the keyboard. The screen shown in Figure 1 (C) corresponds to the congruent trials, and the screen shown in Figure 1 (E) corresponds to the incongruent trials. The IAT effect occurs when the average time to associate the stimulus with the corresponding valence in one of the configurations (congruent or incongruent trials) is shorter than in the other.

Participants are expected to associate the words with the corresponding valence image (pleasant words with smiley faces and unpleasant words with sad faces) within a 300 ms to 3000 ms time window following the auditory stimulus presentation. The test provides three types of responses: correct, incorrect, and non-response. Only trials with correct answers are recorded as valid. The test continues until it reaches the number of trials specified in the protocol for each block.

The Interpersonal Reactivity Index – IRI

The IRI is a self-report measure widely used to assess empathy. It has been applied in several studies to assess gender differences and prosocial behavior in adolescents and assess the inhibitory role of empathy in aggressive behavior (Davis, 1980). The IRI includes cognitive and emotional factors and comprises 28 items. All items have five response categories (1 = describes me not well, 2 = describes me a bit, 3 = describes me quite well, 4 = describes me well, and 5 = describes me very well) distributed in four subscales measuring four dimensions of the global concept of empathy, namely Perspective Taking (PT),

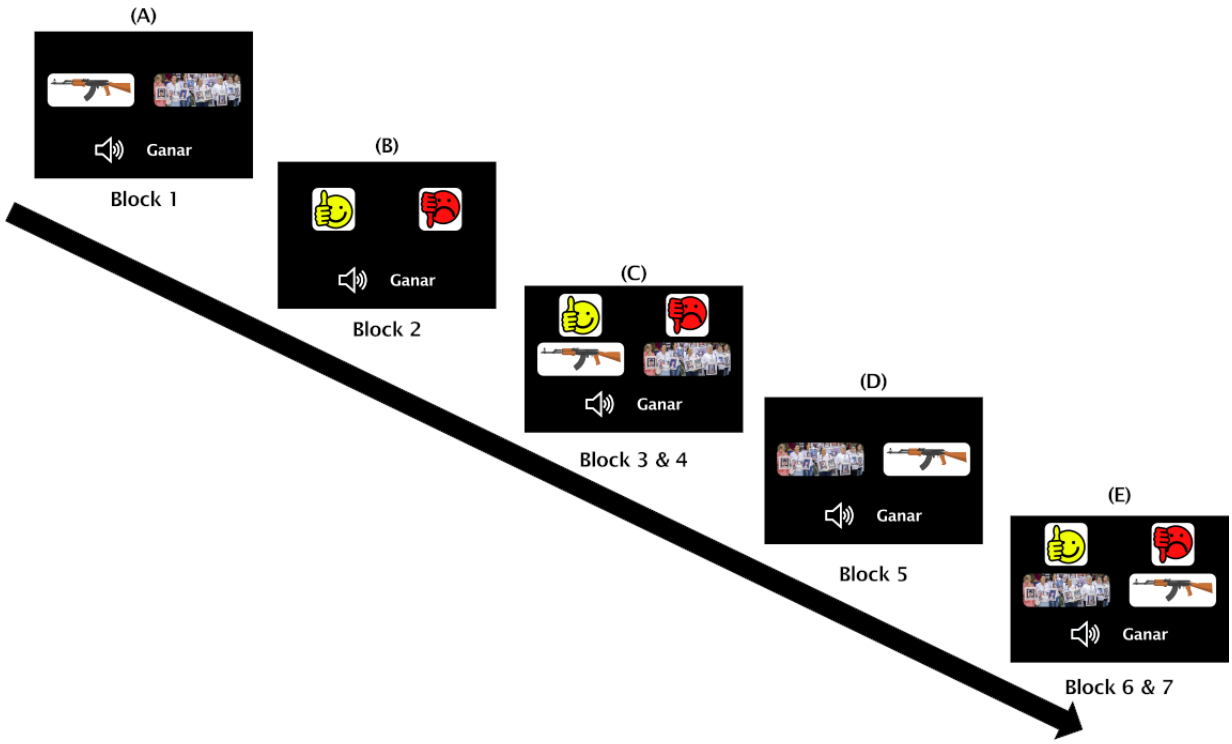


Figura 1

Implicit Association Test (IAT) sample screens and stimuli: (A) block 1, (B) block 2, (C) blocks 3 and 4, (D) block 5, and (6) blocks 6 and 7. A speaker and a word represent stimuli, but they are sounds. Ten words were spoken for each kind of stimulus.

Fantasy (FS), Empathic Concern (EC) and Personal Upset (PD); each of them comprising seven items. Validation was carried out for the Colombian context by Garcia-Barrera et al. (2017), in which 18 items were selected, and the four subscales were maintained.

The Motives for Aggression Inventory – IMA

The IMA is a tool to assess motivations for aggressive behavior. It is a self-report scale comprising 26 items rated on a 3-point Likert scale (1 = never/rarely, 2 = sometimes, and 3 = frequently), indicating the frequency of each motive leading to aggressive behaviors. Empirical solid support underpins the definitions and hypotheses on which this questionnaire is based, which attribute the fundamental motivation for aggressive behavior to particular causes, such as the reinforcement that the individual receives, the need to defend or promote his or her social identity, the presence of various events can cause discomfort, among others (Juárez Acosta and Montejo Hernández, 2008).

The Reactive-Proactive Aggression Questionnaire – RPQ

The RPQ is a self-report instrument created by Raine et al. (2006) and validated in a population exposed to armed conflict by Gómez et al. (2022). It comprises 23 elements based on the proactive motivational dimension (instrumental) vs. reactive (hostile). The subject scores the items on a frequency scale (0 = never, 1 = sometimes, and 2 = often). Proactive aggression has been characterized as instrumental, organized, and cold-blooded, with little evidence of autonomous arousal. Reactive aggression refers to acts that seek compensation for the damage caused to us. For this reason, proactive but not reactive aggression has been associated with higher levels of psychopathic personality, blunted affect, and stimulation-seeking tendencies.

The Scale for Mood States Assessment – EVEA

The EVEA assesses four clinically significant mood states: depression, anxiety, hostility, and joy. Its objective is to assess the current state of mind at the time of the assessment or at a specific time. It has been applied and validated in clinical and general populations (Sanz Fernández, 2001). The scale consists of sixteen items, all formulated in the same direction. It is quick to apply, and its response form is a Likert type of 11 points (from 0 to 10). The 16 sentences that make up the items have the same structure, starting with the words *I feel*, followed by an adjective that names a mood state, for example, I feel sad or happy. Each mood is represented by four items with different adjectives that define a subscale. The four subscales are joy (AL), sadness-depression (TD), anger-hostility (IH), and anxiety (AN).

The Extreme Experience Scale for Armed Conflict Contexts – EX2

The EX2 was adapted and validated for the Colombian setup by Giraldo et al. (2020), based on the Extreme Experience Scale (EACA) (Perez-Sales et al., 2013). An *extreme experience* is defined as an event charged with emotions and experienced individually, and in this context, it must be a direct consequence of the armed conflict. The EX2 consists of 18 items divided into two dimensions: direct extreme experiences (dEX2) with twelve items and indirect extreme experiences (iEX2) with six items. The first dimension identifies personal physical situations such as death or illness. The second seeks to capture situations in which the person could have witnessed extreme events to close friends, family, or people with whom they have an emotional bond. This approach allows for a more precise understanding of the mental conditions associated with armed conflict. The scale has a dichotomous response (yes or no); each affirmation adds a unit to the score, and a score greater than 2,5 indicates a high level of exposition.

2.2 Methods

2.2.1 EEG Acquisition and Preprocessing

The IAT task was synchronized with EEG recordings acquired with a 64-channel Biosemi ActiveTwo with a sampling frequency of 2048 Hz. Electrodes were placed according to the international 10–20 system using quick caps, and the impedances were kept below 10 kΩ. The EEG device and laptops worked with batteries while recording to avoid major electrical problems.

EEG recordings were preprocessed using `MNE` for `Python` (Gramfort, 2013). The original signals were high-pass filtered at 1 Hz with a zero-phase shift FIR filter. Continuous EEG data was partitioned into epochs ranging from 200 ms before stimulus onset to 800 ms after it. Epochs were baseline corrected using the−200 to 800 ms window, downsampled to 256 Hz, and offline re-referenced using the Electrode Standardization Technique (REST) (Yao, 2001). Bad channels were automatically detected and interpolated by the `pyprep` library (Bigdely-Shamlo et al., 2015). An Independent Component Analysis (ICA) was performed to remove electrooculography (EOG). Bad trials were automatically corrected by the `Autoreject` library (Jas et al., 2017). For ERP analyses, EEG recordings were low-pass filtered at 30 Hz with a zero-phase shift FIR filter, and for EEG-based functional connectivity analyses, they were low-pass filtered with a cutoff frequency of 40 Hz. After that, epochs were separated according to experimental tasks (Congruent or Incongruent Blocks).

2.2.2 Event-Related Potential Analyses – ERP

An ERP is a measured brain response, recorded by EEG, resulting from presenting a specific stimulus or event that occurs in the external world or within the brain itself (Picton et al., 1995). These potentials provide a noninvasive means to evaluate the activity of the human brain as it perceives stimuli, makes decisions, and controls behaviors.

ERP waveforms are composed of positive and negative deflections, known as components, with characteristic peaks and latencies that reflect neural activity associated with cognitive processes such as perception, attention, memory, and decision-making (Luck, 2014). Usually, components are labeled with “P” or “N” to indicate positive and negative peaks, respectively, and a number that indicates the position of a component within the

waveform. Alternatively, the number may indicate the component's latency in milliseconds, e.g., N170 for a negative peak around 170 ms.

There are two standard procedures for identifying ERP components. The first consists of a priori selecting the electrodes and time windows to be analyzed (based on some hypothesis), which is acceptable for confirmatory analyses. However, to ensure the reliability of the results, the guidelines of the Society for Psychophysiological Research (Keil et al., 2014) should be followed, and the research should be preregistered, as described in (Paul et al., 2021). The second procedure is to select components from visual inspection of the evoked grand averages, but this procedure inflates Type I errors (Luck and Gaspelin, 2017).

An alternative approach involves performing data-driven studies using a massive univariate or similar analysis. In this type of analysis, a vast number of univariate tests (e.g., t–test or ANOVA) are performed at all available time instants and electrodes (in our case, 256-time instants and 64 electrodes), which reduces the possibility of false discoveries, something inherent in performing many hypothesis tests. The general pipeline for conducting many univariate tests is shown in Algorithm 1.

Data: Pre-processed EEG
Result: Statistics
Assemble the trials of the two (or more) experimental conditions into a single set;
for $i \leftarrow 0$ **to** 1000 **do**
  Random partitioning: randomly extract as many trials from this combined data set as in condition one and place those trials in subset 1. Place the remaining trials in subset 2;
  Compute the test statistics in this random partition;
end
Construct a histogram of the test statistics;
From the results observed in line 1, find the proportion of random partitions that resulted in a test statistic larger than the observed one. This proportion corresponds to the p–value ;

Algorithm 1: General pipeline for univariate tests (Maris and Oostenveld, 2007)

If the p–value obtained with Algorithm 1 is less than the critical alpha level (usually ,05), conclude that the data in the two experimental conditions are significantly different. As the neighboring electrodes and time instants should exhibit a similar behavior, Groppe et al. (2011) proposed a modification of Algorithm 1, known as cluster-based permutation,

as in Algorithm 2.

Data: Pre-processed EEG
Result: Statistics
t–scores (or other test statistics) are computed for each time point and sensor of interest;
All t–scores that do not exceed some threshold (e.g., the t–score corresponding to an uncorrected p–value of 5 %) are ignored;
The remaining t–scores are grouped into clusters by grouping t–scores at adjacent time points and sensors;
The t–scores in each cluster are summed to produce a cluster-level t–score. This score is taken as the "mass" of the cluster ;
The most extreme cluster-level t–score across permutations of the data is used to derive a null hypothesis distribution ;
The p–value of each cluster is derived from its ranking in the null hypothesis distribution ;
The p–value of the entire cluster is assigned to each cluster member and reflects an adjustment for multiple comparisons. The adjusted p–value for multiple comparisons of tests not assigned to a cluster is one ;

Algorithm 2: Cluster-based permutation (Groppe et al., 2011)

The cluster-based permutation method has two arbitrary parameters: the neighbor definition and the t–score threshold, which determine the size and number of clusters (Groppe et al., 2011). Smith and Nichols (2009) developed a method known as TFCE (Threshold-Free Cluster Enhancement) in which the threshold value does not have to be predetermined but is calculated using a ROC-based optimization technique. This method is implemented in `MNE` and was used for this analysis. The definition of neighboring electrodes is done using an adjacency matrix.

Massive univariate analyses (Maris and Oostenveld, 2007) are popular because they allow data-driven exploratory analyses with high statistical power and low Type I errors and can even be used in confirmatory and multifactorial studies (Fields and Kuperberg, 2020). However, similar to classical statistical approaches, it is crucial to use caution when interpreting the outcomes, specifically about the precise location or duration of the identified effects. (Sassenhagen and Draschkow, 2019).

2.2.3 Bayesian Inference Methods

Bayesian inference methods are a viable alternative to classical null hypothesis significance testing (NHST) approaches to address concerns around their reliability and reproducibility (Keysers et al., 2020; Wagenmakers et al., 2016). Bayesian inference can be used to determine whether an effect exists and its size (van Doorn et al., 2021). As a result, the Bayes Factor serves to quantify the comparative predictive efficacy of the null and alternative hypotheses. According to Eq. (1), it indicates the extent to which confidence in the relative plausibility of the hypotheses must be reevaluated in light of the data. The Bayes Factor is a statistical measure used in Bayesian hypothesis testing to assess the strength of evidence for one hypothesis over another. It quantifies the ratio of the likelihood of the data under one hypothesis compared to the likelihood under another. In the context of Bayes factors, the hypotheses are typically the null hypothesis ($\mathcal{H}_0$) and the alternative hypothesis ($\mathcal{H}_1$).

$$\mathcal{BF}_{10} = \frac{p\left(D \mid \mathcal{H}_1\right)}{p\left(D \mid \mathcal{H}_0\right)}, \tag{1}$$

where $p(D \mid \mathcal{H}_1)$ can be interpreted as evidence in favor of the alternative hypothesis, and $p(D \mid \mathcal{H}_0)$ as evidence in favor of the null hypothesis (Etz and Vandekerckhove, 2018). The subscript in the Bayes factor notation indicates which hypothesis the data support. $\mathcal{BF}_{10}$ indicates the Bayes factor in favor of $\mathcal{H}_1$ over $\mathcal{H}_0$. A Bayes Factor greater than 1 suggests evidence in favor of the alternative hypothesis, while a Bayes Factor less than 1 suggests evidence in favor of the null hypothesis. In contrast, $\mathcal{BF}_{01} = 1/\mathcal{BF}_{10}$ indicates the Bayes Factor in favor of $\mathcal{H}_0$ over $\mathcal{H}_1$. The larger the Bayes Factor, the stronger the evidence in favor of the selected hypothesis (van Doorn et al., 2021). This approach allows researchers to update their beliefs about hypotheses as new data become available, making it a valuable tool for model comparison and hypothesis testing in Bayesian statistics.

In the Bayesian framework, variance analyses involve comparing linear regression models. These models consist of predictors representing factors and interactions between components and are contrasted against a null hypothesis model. The null hypothesis model lacks predictors but includes the intercept term (Wetzels et al., 2012). An improvement to the standard Bayesian model for ANOVA is to parameterize the model in terms of effect size, which makes priors independent of the units of measurement of the predictor variables, shrinks them, reduces the model error, and simplifies the specification of informative priors (Rouder et al., 2017).

The Bayes factors derived from comparing each potential model with the null model will provide insight into the model that most effectively explains the observed data. Examining each element allows one to determine its inclusion within the model (Wagenmakers et al., 2018). Post hoc analyses are conducted if the factor is deemed suitable for inclusion in the model. These analyses aim to identify significant differences among levels while accounting for multiple comparisons, adjusting the priors using the methodology outlined by Jeffreys (1938) or Westfall et al. (1997).

When the primary objective of the analysis is to determine the size of the effect, it is essential to visually represent the posterior distribution and compute the $x\,\%$ Credible Interval (CrI) (van Doorn et al., 2021). According to Bayes'Theorem, the posterior distribution reflects the likelihood of the parameter values after the a priori knowledge is updated with the data. The $x\,\%CrI$ contains the $x\,\%$ percentage of the mass of the posterior distribution. Two popular ways of creating a $x\,\%CrI$ are the highest density CrI, which is the narrowest interval containing the specified mass, and the central CrI, which is created by cutting off $\frac{100-x}{2}$ from each of the tails of the posterior distribution (van Doorn et al., 2021).

Testing and estimation are not mutually exclusive and can be used in sequence. For example, one can first use a Bayes factor hypothesis test to determine the effect and then estimate the effect size by calculating $95\,\%CrI$. In our analyses, we use Bayesian inference to determine the existence or not of an effect and determine the size of the effect, if it exists (van Doorn et al., 2021).

2.2.4 EEG-based Functional Connectivity Analyses

Functional connectivity refers to a collection of methodologies rooted in graph theory to quantify the degree of synchronization among multiple brain regions and offer valuable insights into how these regions interact. This approach works under the assumption that the synchronization of several regions of the brain, rather than isolated regions acting alone, is necessary to perform various cognitive activities (Poli et al., 2015). Functional connectivity analyses can be performed using data acquired from various neuroimaging techniques such as EEG, magnetoencephalography, or functional magnetic resonance imaging (fMRI). However, in this section, our focus is on functional connectivity strategies using EEG signals. EEG-based functional connectivity analyses can be approached through two distinct methodologies: one at the level of sensors, which typically involves electrodes placed on the

scalp to measure electrical activity directly, and the other at the level of sources, where the aim is to infer the neural sources responsible for the recorded electrical activity.

The statistical correlations between signals captured on different sensors are assessed at the sensor level. This approach's spatial resolution is constrained due to the deep brain sources from which the recorded signals originate and their mixing on the scalp. Consequently, it becomes challenging to readily discern the precise neural sources responsible for the recorded signals. The abovementioned methodology is frequently employed to examine the interconnections between brain areas concerning collective electrical functioning. The source-level approach involves using sensors to capture data that allow the estimation of neural activity in distinct brain areas, sometimes referred to as "sources". Inverse modeling strategies infer the neural sources responsible for the signals detected on the scalp's surface. Functional connectivity examines statistical associations among estimated brain activities from different sources in this scenario. Connectivity analysis conducted at the source level provides a superior spatial resolution to the sensor level. This methodology enables the identification of distinct brain regions that contribute to the detected signals, thus improving the comprehension of the findings. Examining source-level functional connectivity provides helpful information on communication patterns in various brain regions for neuronal activity. Due to the absence of a head model, our investigation primarily focused on assessing functional connectivity at the sensor level.

In addition to the distinction between sensor and source-level analysis, functional connectivity methods also vary depending on the direction of interaction. They can be categorized as directed or non-directed, including model-free and model-based approaches (Bastos and Schoffelen, 2016). One of the key differences lies in whether the metrics measure the direction of interaction. Non-directed functional connectivity captures the interdependence between signals without considering the direction of influence. In contrast, directed measures aim to establish statistical causation based on the principle that causes precede their effects. Directed and non-directed connectivity measures can be classified into model-free and model-based approaches.

EEG-based functional connectivity analysis could be performed in both time and frequency domains, but preferred the second option due to its ability to capture dynamic changes in brain activity (Fornito et al., 2016). In this thesis, we used non-directed measures to study the frequency domain. This choice was motivated by the need to gain insight into the synchronization patterns of multiple brain regions at different frequencies. Furthermore, we considered the limitations of low spatial resolution at the sensor level,

which would invalidate the causal relationships identified by directed estimations. Hence, the following section will outline the necessary techniques for conducting an EEG-based functional connectivity study using non-directed measurements in the frequency domain.

Spectrum Estimation

The representation of a signal conveyed across a set of observations and frequency bands in the frequency domain approximates the amplitude and phase of oscillations. Identifying rhythmic components and phase differences is typical of the results obtained by transforming signals into the frequency domain through Fourier decomposition, wavelet analysis, or Hilbert transformation. In conventional approaches, where single window functions such as rectangular or Gaussian tapers are applied to preprocess the data before the Fourier transform, issues such as spectral leakage and reduced frequency resolution often arise, especially when analyzing short or non-stationary data segments(Thomson, 1982). These limitations underscore the importance of using the multitaper method, which addresses these inconveniences, providing more accurate and robust spectral estimates, particularly in analyzing complex and dynamic signals.

The multitaper method, first described by Thomson (1982) and often used in neurophysiology (Mitra and Pesaran, 1999), is a spectral analysis technique used in analyzing time series data. It estimates the power spectral density (PSD) and assesses the signal's frequency content. The key idea behind the multitaper method is to reduce bias and variance in the spectral estimation compared to traditional single-taper methods, such as the periodogram. These benefits are achieved using multiple orthogonal or nearly orthogonal taper functions, also known as Slepian taper functions (Slepian, 1978), to compute the spectral estimate. Each taper is essentially a weighted version of the data, and they are designed to minimize spectral leakage and maximize energy concentration within a specified bandwidth. Time series data are multiplied, or tapered, by these tapering functions, resulting in multiple tapered data sets. A Fourier transform is further applied separately to each of these tapered data sets, yielding multiple spectral estimates. Finally, the power spectral densities obtained from each tapered data set are averaged or combined to create a final estimate of the PSD. This averaging helps reduce the variance and improve the accuracy of the spectral estimate (Thomson, 1982).

According to Babadi and Brown (2014), the multitaper approach has numerous noteworthy benefits. First, it improves the frequency resolution by allowing each taper to collect distinct signal frequency components, thus facilitating a more comprehensive

examination of the spectrum content. Furthermore, the technique enhances the reliability and robustness of the results by mitigating variation by implementing several orthogonal tapers and the subsequent averaging of spectral estimations. Furthermore, this methodology reduces spectral leakage, leading to a more precise detection of spectral maxima. One additional advantage is the adjustable bandwidth, which allows for customization of the number of tapers used to suit the specific attributes of the data and the desired trade-off between frequency precision and reduction of variation. Using more tapers results in enhanced resolution but at the expense of heightened volatility. On the contrary, using fewer tapers decreases variance but simultaneously restricts resolution.

Finally, the multitaper approach is beneficial for performing statistical hypothesis testing. It facilitates evaluating the extent to which a particular frequency component deviates considerably from the prevailing background noise.

Sinchronization Measurement

In frequency-based brain connectivity, many metrics can be used to obtain particular views on the interaction and synchronization of brain areas. Frequency domain connectivity checks how well phase differences between oscillating elements fit together, which could mean that different groups of neurons work together in a functionally significant way (Bastos and Schoffelen, 2016). These metrics can provide valuable information on the brain's functional organization and help identify communication patterns between different regions. Researchers can better understand how different brain areas work together to perform various cognitive tasks or processes by analyzing the coherence of phase disparities.

Phase-based synchrony measures aim to capture particular characteristics of the probability distribution of cross-spectral densities for a single observation. The cross-spectral density can be obtained by taking the complex conjugate of the spectral representation of one signal and multiplying it by the spectral representation of the other signal. This cross-spectral density is equivalent to the frequency domain representation of the cross-covariance function. In other words, it measures the uniformity of the phase-difference distribution. The fundamental concept is that the weighted sum of the frequency contents will not be zero if observations of the phase difference between two oscillatory signals are consistent. On the contrary, it will approach zero when the phase differences among individual observations are uniformly distributed across 0 to 360 degrees. (Bastos and Schoffelen, 2016).

Connectivity estimates in neuroscience involve using phase-based metrics, which consist of many useful tools such as coherence, phase locking value, and phase lag index. These metrics provide distinct perspectives on brain communication and coordination. Coherence is a widely used metric that measures the consistency and strength of phase relationships between neural oscillations at specific frequencies, reflecting the degree of synchronization between brain regions (Bastos and Schoffelen, 2016). The coherence coefficient between the estimated spectrum of two signals Z_1 and Z_2 is a normalized value between 0 and 1 and can be calculated as follows (Vinck et al., 2011):

$$C \equiv \frac{E\{X\}}{\sqrt{E\left\{M_1^2\right\} E\left\{M_2^2\right\}}}, \tag{2}$$

Where $E\{X\}$ denotes the expected value of X, $X \equiv Z_1 Z_2^*$, being Z_2^* the complex conjugate of Z_2, $M_1 \equiv |Z_1|$ and $M_2 \equiv |Z_2|$.

Higher coherence suggests stronger synchronization and more consistent phase coupling, indicating robust functional connectivity. In contrast, lower coherence values may imply weaker or more variable phase relationships, potentially indicating less effective neural communication or less coherent functional networks.

Coherence only indicates a linear correlation between the signals, which mixes the phase and amplitude correlations. In other words, coherence measures the degree to which two signals are synchronized or share similar patterns. It does not provide information about the signals'specific phase or amplitude relationships. Therefore, while coherence provides valuable information on the overall synchronization between signals, more is needed to characterize the underlying relationships fully. To solve this issue, Lachaux et al. (1999) proposed measuring phase synchronization using only the relative phase between the signals. In pursuit of this objective, the researchers implemented the phase locking value (PLV) index, which the following Eq gives. (3), and it is defined as the resultant circular length of the relative phases.

$$P \equiv |E\{\exp(i\Theta)\}| \tag{3}$$

When considering PLV, it is important to address various potential challenges. Problems to be addressed cover noise sources, the influence of volume conduction on source activity, the presence of a common reference, and potential biases associated with sample

size (Vinck et al., 2011). In response to the need for more robust and reliable phase synchronization indices, Nolte et al. (2004) proposed using imaginary coherence (ImC), which is a conservative approach to assess phase synchronization and is defined as follows:

$$ImC \equiv E\{Im\{C\}\} \tag{4}$$

ImC captures meaningful phase relationships while mitigating some of the limitations of PLV. However, Stam et al. (2007) pointed out a drawback about the effectiveness of ImC in detecting synchronization if two sources are in-phase or in-phase opposition. To overcome this limitation, they proposed a new measure called the phase lag index (PLI), defined as:

$$\Psi \equiv |E\{\mathrm{sgn}(\mathfrak{I}\{\mathrm{X}\})\}| \tag{5}$$

The phase lag index (PLI) quantifies the degree to which the phase differences between signals from two sensors are not equally likely at a specific frequency, regardless of the size of these phase differences. However, the sensitivity of the PLI to noise and volume conduction could be impeded because of the discontinuity present in this index. This discontinuity arises when minor disturbances cause phase lags to transition into phase leads and vice versa (Vinck et al., 2011). Addressing these issues, Vinck et al. (2011) introduced the weighted phase lag index (WPLI), which offers improved capacity to detect true changes in phase synchronization while mitigating the influence of common noise sources and changes in phase coherency. WPLI is calculated as follows:

$$\Phi \equiv \frac{|E\{\mathfrak{I}\{\mathrm{X}\}\}|}{E\{|\mathfrak{I}\{\mathrm{X}\}|\}} = \frac{|E\{|\mathfrak{I}\{\mathrm{X}\}|\,\mathrm{sgn}(\mathfrak{I}\{\mathrm{X}\})\}|}{E\{|\mathfrak{I}\{\mathrm{X}\}|\}}. \tag{6}$$

The WPLI extends the PLI by weighting the contribution of the observed phase leads and lags by the magnitude of the imaginary component of the cross-spectrum. In this way, it alleviates the discontinuity mentioned above. The WPLI exhibits two primary benefits compared to the PLI: increased susceptibility to additional uncorrelated noise sources and enhanced ability to identify real changes in phase synchronization (Vinck et al., 2011; Bastos and Schoffelen, 2016).

Connectivity Matrix Pruning

In functional connectivity analysis, the first stage is constructing a connection matrix by computing one of the connectivity metrics mentioned above for every pair of electrodes. Subsequently, these computed values are arranged in a matrix format, whereby each element denotes the magnitude or extent of the connection between two particular electrodes. To remove spurious or noisy links and highlight the essential topological properties of the network, it is common practice to prune the resulting weighted graph by removing some edges while maintaining a high level of connectivity.

This pruning process is typically based on specific threshold criteria, such as selecting only the strongest connections above a particular threshold value. Various thresholding methods have been proposed in the literature, including hard thresholding (Poli et al., 2015), density-based or proportional thresholding (van den Heuvel et al., 2017), shuffle methods (Kamiński et al., 2001), double thresholding (Boschi et al., 2021), adaptive thresholding (Wang et al., 2019), and deep learning-based methods (Fakhari et al., 2023). However, there currently needs to be a consensus on the optimal thresholding method (Zakharov et al., 2021). One notable concern is that the results could vary significantly depending on the specific thresholding technique and threshold used, as shown by previous studies conducted by Garrison et al. (2015) and Zakharov et al. (2021).

In recent years, the Minimum Spanning Tree (MST) has emerged as a popular alternative to thresholding approaches. An MST is a fundamental concept in graph theory and network analysis. It is a tree structure that spans all nodes in a connected, undirected graph while minimizing the sum of edge weights. In other words, an MST is a subnetwork that connects all nodes without forming cycles and has the smallest possible total edge weight (Kruskal, 1956; Prim, 1957). The MST analysis assumes that all nodes are connected in the original network, and each connection has a unique weight, usually representing the strength of connectivity (Tewarie et al., 2015). These assumptions allow for creating a meaningful and interpretable hierarchy within the network.

There are two ways to find the MST of a weighted graph (Kruskal, 1956; Prim, 1957). Kruskal's algorithm focuses on edge sorting and component connectivity, and Prim's algorithm emphasizes node proximity and the growth of the MST from an initial seed. Kruskal's method sorts edges by weight, starting with the smallest. The algorithm iterates through edges, adding them to the MST if they do not create cycles. Kruskal's algorithm continues until all nodes are connected in the MST, producing the minimum total edge weight (Stam et al., 2014).

The output of constructing an MST from a connected weighted graph with distinct weights is unique. The importance of this uniqueness lies in the fact that it eliminates the necessity of choosing an arbitrary threshold to establish the graph's structure (Tewarie et al., 2015). As its topology is not dependent on the absolute values or distribution of the weights in the original network but instead on their ordering, the MST is immune to scaling effects (Jackson and Read, 2010). Furthermore, the MST remains unaffected by any weight transformation that maintains weight order (Dobrin and Duxbury, 2001).

MST is advantageous, compared to the analysis of the entire weighted graph, as it focuses on the most critical subgraph and avoids biases arising due to weight differences between different graphs (van Dellen et al., 2018). Furthermore, the MST method has better sensitivity to slight differences in the brain network, providing a new tool for research on complex brain networks (van Diessen et al., 2016). When building the MST from the weights in G_w, the resulting MST is binary, with the edges existing or not existing and having no weights (Stam et al., 2014). Analysis of binary connectivity matrices is mainly concerned with understanding the topological patterns of connections between nodes, regardless of variations in their weight (Fornito et al., 2016). By focusing on the connections present and where they are placed, scientists can gain insight into the organization of the brain network and how different regions communicate with each other (Rubinov and Sporns, 2010).

Network Topology Measures

The MST building process is then followed by a topological study, which involves the identification of network topologies that effectively represent the underlying phenomenon. Topology measures are network-invariant functions that measure relevant topological features such as connectivity and sparsity. Such measures can be taken at the local or global level. Local measures of individual network elements, such as nodes or links, typically quantify connectivity profiles associated with these elements and reflect how these elements are embedded in the network. Global measures, on the other hand, provide a holistic view of the entire network by summarizing its overall connectivity and structure. These measures allow comparing networks or tracking network topology changes over time Rubinov and Sporns (2010).

Topological measures in brain connectivity networks play a crucial role in assessing the balance between segregation and integration of information processing within the complex web of neural connections. Functional segregation refers to the ability of the brain

to process information in densely interconnected specialized groups of brain regions. In functional networks, clusters indicate an organization of statistical dependencies indicative of segregated neural processing. This concept is essential to understanding how the brain processes information and how different brain areas work together to perform complex tasks (Rubinov and Sporns, 2010).

In contrast, functional integration is how quickly it combines specialized information from distributed brain regions. Integration measures estimate the ease with which different regions of the brain communicate. These measures are based on the concept of a path, a sequence of nodes, and links between different regions that represent potential routes of information flow. In functional networks, integration represents statistical associations, which makes interpretation more complex (Rubinov and Sporns, 2010).

Regarding the MST topology, there are two extreme shapes: paths and stars (Stam et al., 2014). In the first one, all nodes are connected to two other nodes, except those at either end, with one single link. Nodes with only one link in a tree are called leaf nodes, and the number of such nodes in a tree is known as the leaf number. Thus, a path has a leaf number of two. In the latter scenario, only one central node exists as the connection point for every other node, thus having a leaf number of N. It is possible to characterize the MST topology according to the leaf number, as suggested by Boersma et al. (2013). A more regular network with higher clustering and longer path lengths corresponds to more line-like MSTs with longer diameters and smaller leaf numbers. Instead, more random networks show low clustering, short path lengths, and the corresponding MSTs'shorter diameters and higher leaf numbers.

In this thesis, interpreting topological measures in the context of brain functional connectivity is essential to understanding the organization of neural networks and their relevance to cognitive processes, such as characterizing Colombian armed conflict actors. Thus, we calculate the following topological measures from the MST matrix.

Leaf fraction: The proportion of leaf nodes in the tree to the total number of nodes. A lower value of the leaf fraction indicates a less centralized network topology, and a high leaf fraction means that communication is strongly dependent on the nodes of the hub (Blomsma et al., 2022). The minimum value of this measure is $2/M$, where M is the number of links, and the maximum is 1. Tewarie et al. (2015) found that the MST leaf fraction shows a strong negative linear relationship with the clustering coefficient, meaning it could be considered a segregation measure. These nodes may

represent regions responsible for specialized or distinct functions within the brain (Bullmore and Sporns, 2009).

Diameter: The longest of the shortest path lengths between any pair of nodes in the network. This measure is usually normalized by dividing it by the number of links. An increase in diameter means a decrease in global efficiency, while a low diameter indicates a more efficient information flow between brain regions (Blomsma et al., 2022). Tewarie et al. (2015) found that the MST diameter was linearly scaled with the length of the path network and that the two are positively correlated, which means that the diameter could be considered an integration measure. In cognitive processes, this could characterize the need for information integration in various brain regions (Bullmore and Sporns, 2009).

Mean eccentricity: The eccentricity of a node is the longest optimal path from this node to any other node (Stam et al., 2014). The mean eccentricity represents the average value of the eccentricity of all nodes (Cao et al., 2020). A high mean eccentricity indicates a centralized network in which most nodes are relatively close to each other (Blomsma et al., 2022). In brain networks, centralization can reflect efficient information flow and rapid integration of various cognitive functions (Bullmore and Sporns, 2009).

Maximum degree: The maximum degree is the highest number of connections to a single node, highlighting the network's hubs. The nodes with the maximum degree represent highly connected hubs that play a crucial role in facilitating communication between brain regions (Cao et al., 2020). The presence of hubs in a functional network may correspond to regions responsible for key cognitive functions (Bullmore and Sporns, 2009).

Maximum betweenness centrality: Betweenness centrality measures the shortest paths that pass through a node, revealing the critical nodes for the information flow (Fornito et al., 2016). Nodes with higher values act as bridges connecting disparate regions of the network. These nodes are essential for integrating information and facilitating communication between specialized brain areas.

Tree hierarchy: This measure quantifies the trade-off between large-scale integration in the MST and the overload of central nodes (Cao et al., 2020). It characterizes a hypothetical, optimal topology of an efficient organization while preventing information overload of central nodes (van Dellen et al., 2018). A well-organized hierarchy suggests

a structured information flow and clear cognitive processing hierarchies. Anomalies in the hierarchy could mean cognitive process changes or brain function alterations (Bullmore and Sporns, 2009).

2.2.5 Interpretable Machine Learning Models

This section will concisely overview this thesis's machine learning models employed in Chapter 5. In addition, a concise description of SHAP, the method used to convert opaque ML models into interpretable ones, will be provided.

ML Classification Models. In this section, we discuss machine learning classification models. Since characterizing groups of people is a supervised classification problem, we will provide a brief overview of the models tested in this thesis.

Logistic Regression

Logistic Regression is a linear model initially developed for binary classification, yet it can also be modified to deal with multiclass problems. It models the relationship between the input features and the log-odds of the binary outcome using a linear equation to predict the likelihood of an event occurring. The strength of logistic regression lies in its simplicity and interpretability. However, one significant drawback of the logistic regression model is the assumption of linearity between features and the log-odds of the output, which limits its ability to handle interactions between features effectively and can lead to suboptimal or inaccurate predictions (Hastie et al., 2017). In this thesis, we tuned the regularization type, L1 or L2, and the regularization strength to regulate the size of the coefficients.

Support Vector Machines (SVM)

The SVM model aims to find the optimal hyperplane that maximally separates the data classes, chosen based on the data points closest to this hyperplane, known as the support vectors. The distance between the support vectors and the hyperplane is called the margin, and the objective of SVM is to maximize it. In cases where data are not linearly separable, the SVM uses a kernel function to project data into higher-dimensional spaces where they could become linearly separable (James et al., 2021). For this model, we opt for tuning the regularization parameter to gain control over the balance between margin separation and misclassifications. In addition, we perform an iterative process for

evaluating different kernels and their corresponding hyperparameters.

k-Nearest Neighbors (kNN)

kNN is a non-parametric algorithm that classifies an instance based on the majority class of its k nearest neighbors in the feature space. Given a new observation, kNN goes through the data set to find the k training examples that are closest to the point and returns the output value that has the majority among those k neighbors. The algorithm is non-parametric, so it made no assumptions about data distribution. It is also lazy learning, as no explicit training phase is required. However, kNN can be computationally intensive during the prediction time, especially with large data sets (Lindholm et al., 2022). The selection of the number of neighbors, the distance metric, and the weight function used in the prediction were made as part of the tuning process.

Random Forests

Random Forest is an ensemble of decision trees in which each tree is trained on a subset of the data and features, and the final classification is done by majority vote. A single decision tree is sensitive to noise in its training data and can easily overfit. Random forests mitigate this by training multiple trees by bootstrapping samples and aggregating results in a process known as Bagging. Furthermore, random forests add another layer of randomness by limiting the number of features considered to split at each node, further increasing the diversity of individual trees and reducing overfitting (Hastie et al., 2017). For the tuning process, we selected the number of trees in the forest and the cost-complexity pruning (CCP) alpha, a complexity parameter that controls the trade-off between the number of nodes of each tree and its predictive accuracy.

Gradient Boosting

Gradient Boosting is an ensemble ML technique that sequentially builds predictive models by combining multiple weak learners. It emphasizes correcting errors made by previous models, gradually improving the predictive accuracy. Furthermore, Gradient Boosting also aims to reduce bias, making it a powerful tool for many predictive tasks. However, it can be more prone to overfitting if not carefully tuned (Hastie et al., 2017). In our case, we opt to tune the same hyperparameters, namely the number of estimators and CCP alpha, in the random forest.

Extreme Gradient Boosting (XGBoost). XGBoost is an optimized Gradient Boosting

algorithm with enhancements for performance and speed. One of the main reasons for its effectiveness is its ability to calculate the second-order gradient of the loss function, that is, the Hessian, leading to more accurate step sizes in the boosting process (Chen and Guestrin, 2016).XGBoost exhibits several advantages that separate it from other Gradient Boosting algorithms. It consistently delivers enhanced model accuracy with minimal tuning, operates efficiently due to its well-optimized implementation, and remains robust against overfitting through features like regularization, cross-validation, and tree pruning. In this thesis, we focus on tuning key hyperparameters of this ML model. These included the number of boost rounds, the step size shrinkage to prevent overfitting, the maximum depth of decision trees, the fraction of features used per tree, and the fraction of samples used per tree.

Feature Selection Feature selection is an essential step in ML to improve model performance by eliminating irrelevant or redundant features, reducing overfitting, and enhancing generalization. By performing such a process, one can often build more straightforward, faster, and more interpretable models that perform as well or even better than models built using all the features (Duboue, 2020).

In this work, we used various feature selection methods to extract the most informative attributes for our analysis. First, for the Logistic Regression with both L1 and L2 regularization, we process retained features with coefficients exceeding a designated threshold parameter, which we set at 1.5 times the mean of all feature coefficients. Additionally, for the SVM and Random Forest models, we preserve features whose score ranks surpassed a threshold set at 1.5 times the mean of all score ranks. Further refinement was achieved through filtering techniques, which involved selecting features based on p-values derived from an ANOVA test with the labels of the factor. We adhered to the conventional threshold for statistical significance, keeping features with p-values below 0.05. Lastly, we consider mutual information (MI) by selecting features greater than 0.1 when assessed against the label. MI quantifies the extent to which one random variable provides information about another, allowing us to retain features that offer valuable information for our analysis (Chandrashekar and Sahin, 2014).

SHAP for Interpretability of Opaque ML Models One of the main benefits of using ML models is that they have a high predictive capacity, meaning they can make accurate predictions about new data based on what they have learned from previous data. However, they are opaque, making it challenging to understand the process behind their predictions

or identify the relative importance of different features in influencing these predictions. The attributes above tend to increase along with the complexity of the models, resulting in a trade-off between performance and interpretability (Yarkoni and Westfall, 2017).

In recent years, there has been an increase in data analysis methods aimed at enhancing the interpretability of complex machine learning models such as permutation feature importance (Ribeiro et al., 2016), local interpretable model-agnostic explanations (LIME) (Fisher et al., 2019), and SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017). Consequently, obtaining models with high predictive capacity and interpretability is now feasible. SHAP has become a popular choice for model interpretation due to its ability to provide consistent and meaningful feature attributions for a wide range of ML models. It enhances model transparency and helps to understand the driving factors behind predictions, facilitating trust and accountability in AI and ML applications.

SHAP is an advanced technique that explains the output of complex machine learning models, such as ensemble methods, neural networks, and gradient boosting methods. Its functioning principle is rooted in game theory and the concept of Shapley values(Shapley, 1953). The key idea behind SHAP is to attribute the contribution of each feature to the prediction for a specific instance. The interpretability ability is achieved through the following:

Cooperative game theory: SHAP borrows the concept of cooperative game theory, where "players"(features) cooperate to produce a "payoff"(the model's prediction) in different combinations.

All possible combinations of features: SHAP considers all possible combinations of features, from no features to all features. Each combination represents a "game"where features cooperate to predict an outcome.

Shapley values: For each feature, SHAP computes its Shapley value, which quantifies the average contribution of that feature to all possible combinations. Calculate how much adding the feature to a set contributes to the prediction, considering all possible orders in which the features could join.

Consistency and Fairness: SHAP ensures that Shapley values satisfy properties like consistency and fairness, making them a stable and reliable way to attribute feature contributions. The Shapley values provide an accurate decomposition of the model's

prediction for each instance, enabling an understanding of which features are driving the prediction and to what extent.

Local Interpretability: SHAP values can explain model predictions per-instance basis, offering insights into how specific features influenced a particular prediction.

Global Interpretability: By examining Shapley values in multiple instances, you can understand feature importance at a global level, identifying which features are generally more influential.

The advantages of SHAP include its solid theoretical foundation from game theory, allowing it to explain predictions for any ML model. Shapley values for a single instance can explain a prediction and be assembled into global explanations for a complete data set. Additionally, SHAP has a fast implementation, making it easy to get Shapley values for global explanations such as feature contributions, dependence, and dependence plots. SHAP values also fulfill desirable properties such as local accuracy, missingness, and consistency (Lundberg et al., 2020). Their disadvantages include the potential for misinterpretation of Shapley values for new data if other explanations are used. Moreover, all global explanations in SHAP require computing Shapley values for a large sample (Guidotti et al., 2018).

2.3 Summary

In this chapter, we presented the materials and methods common to several of the following chapters. We described the sample of former armed conflict actors who underwent behavioral tests and electroencephalographic recordings. Next, we gave a brief description of the behavioral tasks performed by the participants, expanding on the description of the IAT, which is designed to measure prejudice between groups of actors. Subsequently, we presented the procedure used to preprocess the EEG and obtain the ERPs, which are the basis of all the analyses performed. After that, the Bayesian inference hypothesis testing method used in Chapters 3 and 4 is explained. Finally, we described the procedure used to obtain the network measures needed to perform the EEG-based functional connectivity analyses presented in Chapter 4, which are part of the features used in the ML models presented in Chapter 5. Finally, the ML models tested in Chapter 5 to characterize the group of participants are briefly described, and the technique used to make these models interpretable is also explained.

3 A Bayesian Approach to Event-Related Potentials

This chapter introduces a new methodology for analyzing Event-Related Potentials (ERPs) by combining Bayesian inference models and massive univariate methods. We demonstrated that our approach is more sensitive and reliable than conventional methods using null hypothesis significance testing or massive univariate methods.

3.1 Literature Review

As previously indicated in Chapter 1, two methodologies exist to analyze ERP.

In the first methodology, the ERP components to be analyzed are previously selected, a measurement is taken of this component - the most common being mean amplitude, peak amplitude, and latency (time instant where the peak amplitude is located) - and finally, NHST is made to these measurements to test whether there are statistically significant differences between experimental conditions or groups of participants. There are several issues with this approach: i) ERP components vary in space and time (Fields and Kuperberg, 2020), and by doing so, they may neglect effects that have not been previously reported (Luck, 2014). ii) Choosing electrodes and time windows a priori is subjective (Keil et al., 2014). In addition, volume conduction, that is, how electromagnetic signals propagate through the brain, introduces variability by allowing multiple components to reach all sensors simultaneously, leading to misinterpretations (Nunez et al., 2006). iii) It inflates Type I errors and can undermine the validity and reliability of the findings (Luck and Gaspelin, 2017).

In addition to issues with conventional ERP analysis approaches, there is growing criticism of the suitability and reliability of NHST and the validity of the effects found using it (Keil et al., 2014; Luck and Gaspelin, 2017). In this regard, Cumming (2014) illustrates that when replicating an experiment with a different random sample, the p–values and confidence intervals are not identical and often contradictory, so they directly recommend not using these tests, an opinion shared by other authors (Nuzzo, 2014). Furthermore, a group of more than 200 researchers conducted a meta-study in which the results showed that psychological studies are prone to reproducibility problems (Open Science Colaboration, 2015).

The second methodology consists of performing massive univariate statistical analy-

ses on the EEG signals, as indicated in Algorithm 1, and from these analyses, finding the existing clusters, which are groups of neighboring electrodes where significant differences are observed in consecutive samples (Groppe et al., 2011). These methods do not require one to pre-select which groups of electrodes or time windows are to be analyzed, as this is done automatically. However, they do not give statistical inference about the location and extent of the effects (Sassenhagen and Draschkow, 2019), so it should be complemented with some statistical inference method.

So far, only ERP-IAT studies using the first methodology have been conducted, which have found effects on components N2 (Chen et al., 2018; Healy et al., 2015), P3a (Healy et al., 2015; Portengen et al., 2022), P3b (Chen et al., 2018), N400 (Healy et al., 2015; Williams and Themanson, 2011), and LPP (Forbes et al., 2012). Nevertheless, the results obtained are only sometimes consistent. For example, at N2, Chen et al. (2018) found effects by block type in the occipital region, Healy et al. (2015) found effects in the interaction block type $\times$ level of the IAT effect in the occipital region, and a negative correlation between the IAT score and the mean amplitude of this component, while Williams and Themanson (2011) did not find effects. Similarly, at N400, Healy et al. (2015) found effects on the interaction block type $\times$ level of the IAT effect in the central and parietal regions, Williams and Themanson (2011) found effects on the interaction block type $\times$ level of the IAT effect in the frontal region, but Chen et al. (2018) did not find effects.

Furthermore, no effects have been found for other components such as N1 (Williams and Themanson, 2011), P1 (Chen et al., 2018), and P2 (Healy et al., 2015; Williams and Themanson, 2011). However, the literature in social cognition indicates that these components are modulated by social categorization processes, which is a precursor of prejudice and what the IAT task measures (Amodio and Cikara, 2021)

Beyond the general issues inherent in traditional ERP analysis methodologies, these inconsistencies in the ERP-IAT analysis results could have two additional causes: EEG preprocessing variations (Healy et al., 2015; Endendijk et al., 2019), and the IAT stimulus dependency (Nosek et al., 2007; Tosi et al., 2018). In the preprocessing stage, for example, while some authors do baseline correction (Chen et al., 2018; Williams and Themanson, 2011), others do not and instead apply a bandpass filter between 4 and 30 Hz (Healy et al., 2015; Endendijk et al., 2019), which fixes the issue of pre-stimulus activity extending into early periods of the ERP waveform, overlapping notably with the P1, issue that not correct well the baselining process, and allows to analyze early ERP stimulus-locked components effectively, but in return strongly attenuates the late ERP components (N400 and LPP). It

is also clear that the type of stimulus conditions the signals; thus, for example, Portengen et al. (2022) works with two types of stimuli (text and images) and defines different time windows for each.

3.2 Methodology

Based on the mentioned limitations, it is crucial to consider alternative approaches that can provide information about the spatial and temporal characteristics of the effects. Following this line, we proposed a novel methodology in which we first discover potential ERP components using massive univariate methods and then measure effect sizes using Bayesian inference.

One distinction between the initial phase of the proposed methodology and conventional massive univariate analysis methods is the retention of samples with a p-value below .5, as opposed to .05. By employing this threshold, we would retain samples in which the probability of discovering an effect is equal to that of the probability of its absence. Moreover, this threshold will be used in the second stage to construct the a priori distribution of the Bayesian inference model.

Once we identified the clusters of interest, we calculated the mean amplitude of the signals for the congruent and incongruent blocks and their difference. Next, the following statistical analyses were performed for each discovered cluster:

- **Mixed ANOVA by Bayesian inference** taking the type of block (congruent or incongruent) as the within-subjects factor and the IAT effect group (negative, neutral, positive) as a between-subjects factor. The prior of Cohen's d effect size is uniform, with a scale factor of 0,2, corresponding to an informative prior assuming a small effect size. The prior is consistent with the p–value used in the first part of the methodology as a threshold for cluster discovery. Furthermore, this analysis was also performed with a prior of 0,5, equivalent to a medium effect size, and with a prior of 0,8, equivalent to a large effect size, to evaluate how sensitive the model is to priors.
- **One-way ANOVA** takes the difference between congruent and incongruent blocks as a dependent variable, depending on where the cluster to be analyzed came from, and the IAT effect group as the factor. The prior of Cohen's d effect size is uniform, with a scale factor of 0,2, corresponding to an informative prior assuming a small effect size.

- ***t*–test for post hoc analysis** for those factors or interactions considered relevant in the models, adjusting the priors according to the procedure proposed by Westfall et al. (1997) to correct for multiple comparisons. With this, a prior with a scaling factor of 0.2 for a 2×3 repeated measures ANOVA results in a prior with a scaling factor of 0,710 to be used in the post hoc analyses. Sensitivity analyses are also performed, which allow us to determine how robust the Bayes factors are to different parameterizations of the prior distribution.

For statistical analysis, we grouped the participants into negative, neutral, and positive levels of the IAT effect, replicating the work of Healy et al. (2015), which is somewhat suggested by the results of Williams and Themanson (2011). The reasoning for this grouping is that we expect ex-combatants to have a positive score, victims a negative score, and civilians a neutral one. We hypothesize that such a distribution among actors (ex-combatants, unaffected civilians, and victims) will contribute to explaining the electrophysiological drivers of prejudice in communities against ex-combatants.

We used the *K-means* algorithm for such a grouping with the initialization method *K-means++* and Euclidean distance. This unsupervised learning algorithm partitions a data set into groups by finding centroids that minimize the sum of squared distances between the data points and their respective centroids. *K-means++* is an improvement over the traditional *K-means* algorithm, designed for more accurate and stable cluster assignments. It ensures that the initial selection of cluster centroids is spread out, reducing the chances of converging to suboptimal solutions. This approach leads to faster convergence and better clustering than the standard *K-means* algorithm (Arthur and Vassilvitskii, 2007).

The following limits were obtained for each group: negative IAT effect: $[-0{,}7057, -0{,}2114]$, neutral: $[-0{,}2114, 0{,}1849]$, positive: $[0{,}1849, 0{,}8680]$. Nineteen subjects comprised the negative IAT effect group, 37 in the neutral group, and 29 in the positive group. Figure 2 shows the densities of these three groups.

3.3 Results with Traditional Analyses

3.3.1 NHST Analyses

The preliminary selection of electrodes and time intervals, following established protocols or testable hypotheses, constitutes the conventional approach to ERP analysis.

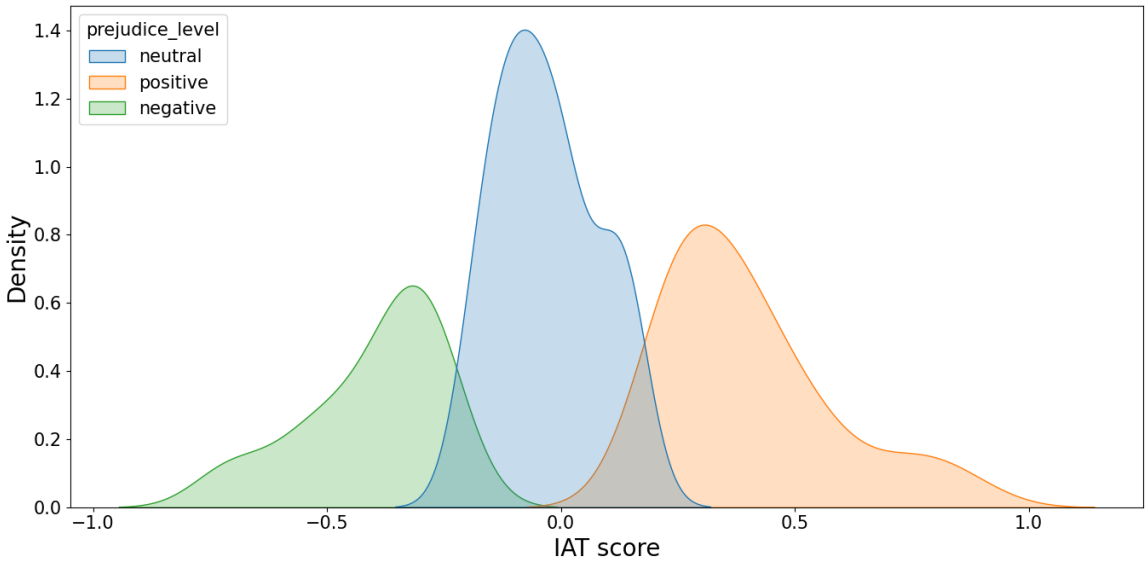


Figura 2

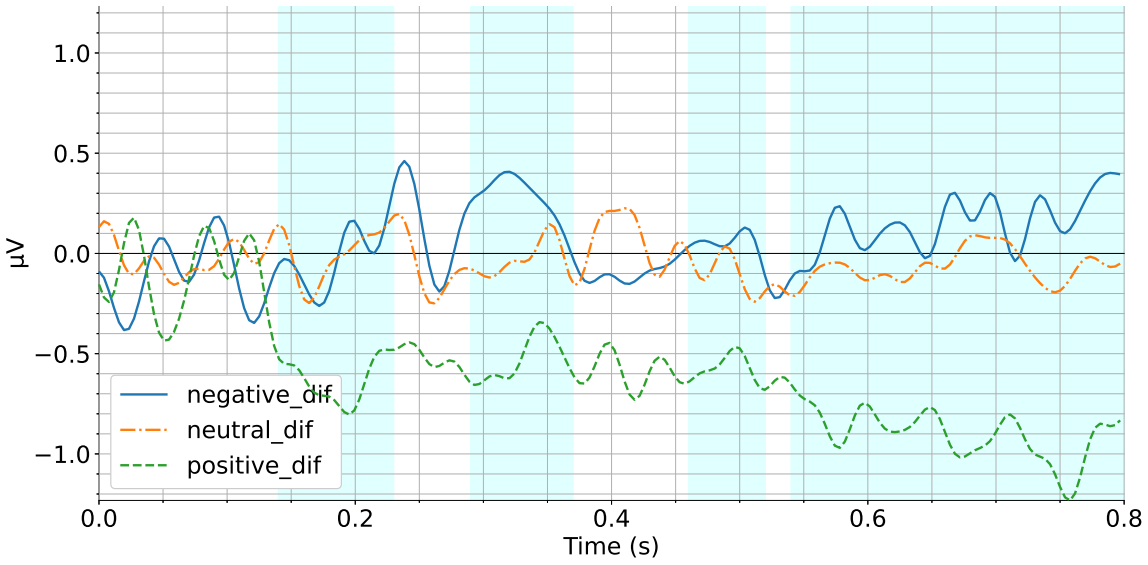
IAT score densities, grouped by IAT effect

In this regard, we selected three electrodes from the central frontal region (FC1, FC2, and FZ), where previous studies had found effects (Forbes et al., 2012; Healy et al., 2015; Hurtado et al., 2009) to illustrate the conventional analysis using NHST.

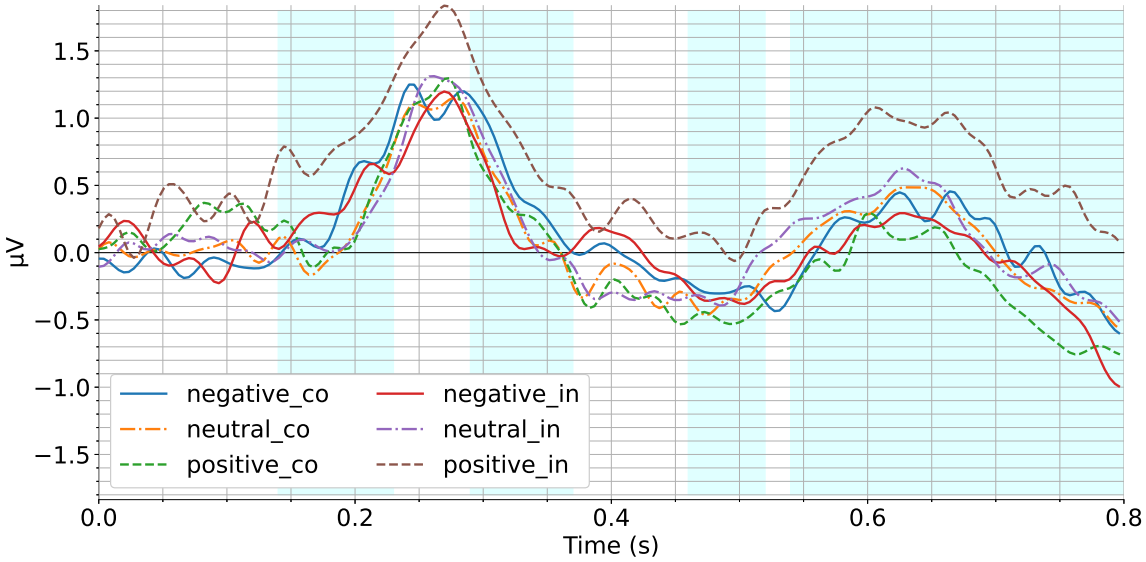
We identified four intervals of interest shown in Figure 3. Their respective descriptive statistics are summarized in Table 3.

Statistical analyses in every time window show the following results.

Time window 1 (140 to 250 ms): There are significant effects on the interaction block type × IAT effect group ($F(2, 82) = 4{,}604, p = {,}013, \eta^2 = {,}030$) and on the difference between congruent and incongruent blocks by the IAT effect group ($F(2, 82) = 4{,}604, p = {,}013, \eta^2 = {,}101$); posthoc analyses resulted in the mean amplitudes of the incongruent block of subjects in the positive IAT effect group being more positive than the mean amplitudes of the congruent block of subjects in this same group ($t = -3{,}563, p_{tukey} = {,}008, d = -0{,}703, CI = [-1{,}126, -0{,}092]$), and than the mean amplitudes of incongruent block of subjects in the neutral IAT effect group ($t = -3{,}079, p_{tukey} = {,}030, d = -0{,}764, CI = [-1{,}304, -0{,}020]$), and that the differences in the mean amplitudes of congruent and incongruent blocks are more negative for subjects in the positive IAT effect group than the differences for the negative IAT effect subjects ($t = 2{,}435, p_{tukey} = {,}045, d = 0{,}719, CI = [0{,}013, 1{,}311]$), and than the differences of the neutral IAT effect subjects ($t = 2{,}744, p_{tukey} = {,}020, d = 0{,}681, CI =$



(a) *ERP congruent minus incongruent block*



(b) *ERP congruent and incongruent block*

Figura 3

Time windows selected by visual inspection to make NHST of the difference between combined ERP of blocks (obtained by subtracting the incongruent block from the congruent block). We selected the time intervals where one signal was appreciably different from the others.

Tabla 3

Descriptive statistics of the mean amplitudes of each time window found by block type and IAT effect group, shown as M(SD).

Time window	Block type	IAT effect group		
		negative	neutral	positive
1	Differences	0,053(0,93)	0,017(0,748)	−0,609(1,099)
	Congruents	0,497(0,759)	0,306(0,917)	0,341(0,759)
	Incongruents	0,445(0,934)	0,289(0,751)	0,951(1,043)
2	Differences	0,268(1,075)	−0,027(0,767)	−0,54(1,328)
	Congruents	0,51(0,811)	0,309(0,944)	0,298(0,886)
	Incongruents	0,242(1,017)	0,335(0,845)	0,838(1,205)
3	Differences	0,049(0,712)	−0,095(0,923)	−0,58(1,768)
	Congruents	−0,288(0,865)	−0,355(1,189)	−0,48(1,138)
	Incongruents	−0,337(0,79)	−0,26(1,074)	0,1(1,412)
4	Differences	0,144(0,677)	−0,061(0,92)	−0,907(1,329)
	Congruents	0,058(1,229)	0,059(1,3)	−0,226(0,94)
	Incongruents	−0,086(1,07)	0,12(1,241)	0,681(1,181)

[0,082, 1,172]).

Time window 2 (290 to 370 ms): There is a significant effect on the interaction block type × IAT effect group ($F(2, 82) = 3{,}722, p = {,}028, \eta^2 = {,}026$); no significant effects were found in the posthoc analyses.

Time window 3 (460 to 520 ms): There are no effects in the block type factor ($F(1, 82) = 2{,}214, p = {,}141, \eta^2 = {,}008$), in the group factor ($F(1, 82) = 0{,}154, p = {,}858, \eta^2 = {,}003$), or in the block type × IAT effect group interaction ($F(2, 82) = 1{,}840, p = {,}165, \eta^2 = {,}013$). There are no effects in the difference between congruent and incongruent blocks ($F(2, 82) = 1{,}840, p = {,}165, \eta^2 = {,}043$).

Time window 4 (540 to 800 ms): There are significant effects on the factor block type ($F(1, 82) = 5{,}520, p = 0{,}021, \eta^2 = 0{,}012$), on the interaction block type × IAT effect group ($F(2, 82) = 7{,}690, p < 0{,}001, \eta^2 = 0{,}034$), and on the difference between congruent and incongruent blocks by the IAT effect group ($F(2, 49{,}770) = 6{,}514, p = 0{,}003, \eta^2 = 0{,}158$); posthoc analyses resulted in the mean amplitudes of incongruent block of subjects in the positive IAT effect group being more positive than the mean amplitudes of congruent block of subjects in this same group ($t = -4{,}708, p_{tukey} < 0{,}001, d = -0{,}769, CI = [-1{,}489, -0{,}324]$), and that the differences in the mean amplitudes of congruent and incongruent blocks are more negative for subjects in the positive IAT effect group than the differences for the negative IAT effect subjects ($t = 3{,}433, p_{tukey} = {,}003, d = 1{,}013, CI = [0{,}320, 1{,}782]$), and than the differences of the neutral IAT effect subjects ($t = 3{,}290, p_{tukey} = {,}004, d = 0{,}816, CI = [0{,}232, 1{,}460]$).

In summary, we found effects in time windows 1 and 4, whereas no effects were found in windows 2 and 3. In the time window 1, which corresponds to early ERP components of positive polarity, the ERPs of the incongruent block of the subjects in the positive IAT effect group are larger (more positive) than the incongruent block of the other groups and than the congruent block of this same group. We also found that while the mean amplitudes of the incongruent block in the subjects in the positive group are larger than the mean amplitudes of the congruent block, the mean amplitudes of the congruent block are larger than the mean amplitudes of the incongruent block in the other groups. In the time window 4, which corresponds to late ERP components of positive polarity, something similar to that described for time window 1 was found.

3.3.2 Massive Univariate Analyses

Since the `MNE` algorithm considers the cluster as a set of two or more adjacent points in time or space, and a significant effect should have a longer duration and extent (Luck, 2014), a sieve was made to retain only those clusters of duration equal to or greater than 40 ms, and with an extent of two or more electrodes, following the usual practice found in the literature review regarding the minimum size of regions of interest (i.e., (Healy et al., 2015; Portengen et al., 2022; Chen et al., 2018)). Finally, we found two clusters by applying this methodology (see Figure 4), based on the suppose that a real effect should have a minimum extension, if not it is noise.

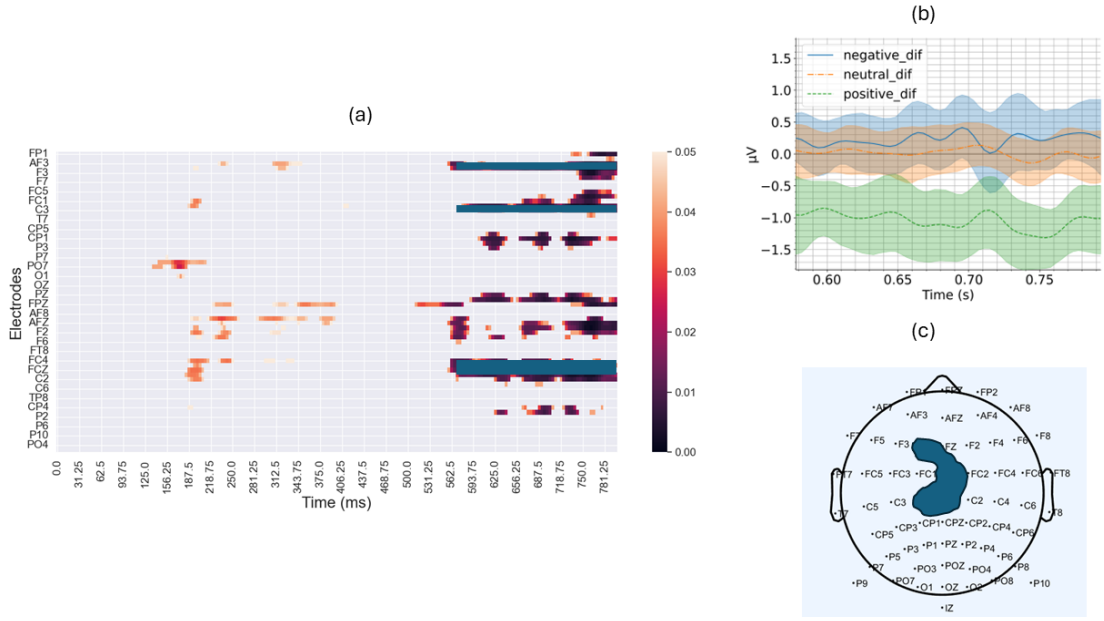


Figura 4

Clusters found by massive univariate analysis, setting the decision threshold at a p-value of ,05. Figure (a) shows the raster plot, highlighting by rectangles the electrodes (F1, FCZ, CZ, and C1) and time window (from 578 to 800 ms) that make up cluster 2. Figure (b) shows the signals resulting from averaging the difference between the ERPs of the congruent and incongruent blocks for each group, including the confidence intervals. Figure (c) shows the location of the electrodes that make up the cluster.

Table 4 presents the description of the clusters found and the descriptive statistics of the mean amplitudes of each block type by the IAT effect group of subjects.

Tabla 4

Descriptive statistics of the mean amplitudes of each cluster found by block type and IAT effect group - M(SD).

Cluster id.	block type	IAT effect group		
		Negative	Neutral	Positive
1	Differences	0,025(1,116)	0,384(1,097)	−0,152(1,357)
	Congruents	0,023(0,862)	−0,304(0,823)	0,199(0,983)
	Incongruents	−0,002(0,697)	−0,688(0,943)	0,351(1,067)
2	Differences	0,225(0,832)	0,013(0,823)	−1,047(1,282)
	Congruents	0,042(1,215)	0,051(1,376)	−0,364(0,939)
	Incongruents	−0,183(1,099)	0,038(1,194)	0,683(1,17)

Once the clusters were identified, we performed NHST analyses on them. Detailed results of these analyses are presented as follows:

Cluster 1 (electrodes: C3, CP1; time window: 343.75 to 386.72 ms): We found significant effects on IAT effect ($F(2, 82) = 10{,}488, p < {,}0001, \eta^2 = {,}126$); posthoc analyses revealed that the mean amplitudes of blocks of subjects in the neutral IAT effect group are more negative than the mean amplitudes of blocks of subjects in the positive IAT effect group ($t = -4{,}485, p_{tukey} < {,}001, d = -0{,}842, CI = [-1{,}191, -0{,}351]$), and that the mean amplitudes of blocks of subjects in the negative IAT effect group ($t = -2{,}589, p_{tukey} = {,}030, d = -0{,}533, CI = [-0{,}985, -0{,}028]$).

Cluster 2 (electrodes: F1, FCZ, CZ, C1; time window: 578.12 to 792.97 ms): We found significant effects on the factor block type ($F(1, 82) = 5{,}668, p = {,}020, \eta^2 = {,}012$), on the interaction block type × IAT effect ($F(2, 82) = 12{,}398, p < {,}001, \eta^2 = {,}051$), and on the difference between congruent and incongruent blocks by the IAT effect group ($F(2, 45{,}160) = 9{,}560, p < {,}001, \eta^2 = {,}232$); posthoc analyses resulted in the mean amplitudes of incongruent block of subjects in the positive IAT effect group being more positive than the mean amplitudes of congruent block of subjects in this same group ($t = -5{,}609, p_{tukey} < {,}001, d = -0{,}883, CI = [-1{,}612, -0{,}483]$), and that the differences in the mean amplitudes of congruent and incongruent blocks are more negative for subjects in the positive IAT effect group than the differences for the negative IAT effect subjects ($t = 4{,}287, p_{tukey} < {,}001, d = 1{,}265, CI = [0{,}564, 1{,}980]$), and than the differences of the

neutral IAT effect subjects ($t = 4{,}254, p_{tukey} < ,001, d = 1{,}055, CI = [0{,}465, 1{,}656]$).

In summary, we found effects in the left central parietal region in a late time window where ERPs have negative polarity. In this region, the ERPs of subjects in the neutral IAT effect group are larger (more negative) than those in the other two groups. We also found effects in the left central frontal region in a late time window where ERPs have positive polarity. In this region, the mean amplitudes of the incongruent block of the subjects in the positive IAT effect group are larger (more positive) than the mean amplitudes of the incongruent block of the other groups and the congruent block of this same group. Finally, while the mean amplitudes of the incongruent block in the subjects in the positive group are larger than the mean amplitudes in the congruent block, in the other groups, mean amplitudes of the congruent block are larger than the mean amplitudes of the incongruent block.

3.4 Results with Proposed Methodology

3.4.1 Discovering of Clusters

We applied the proposed approach to compare between the IAT effect groups the mean amplitudes of the congruent block, the mean amplitudes of the incongruent block, and the differences between the mean amplitudes of the congruent and incongruent blocks (see Figure 5).

As in the conventional method, only clusters with a duration greater than 40 ms and an extension of two or more electrodes were selected, resulting in the 9 clusters shown in Table 5. The grand averages of these clusters are shown in Figure 6.

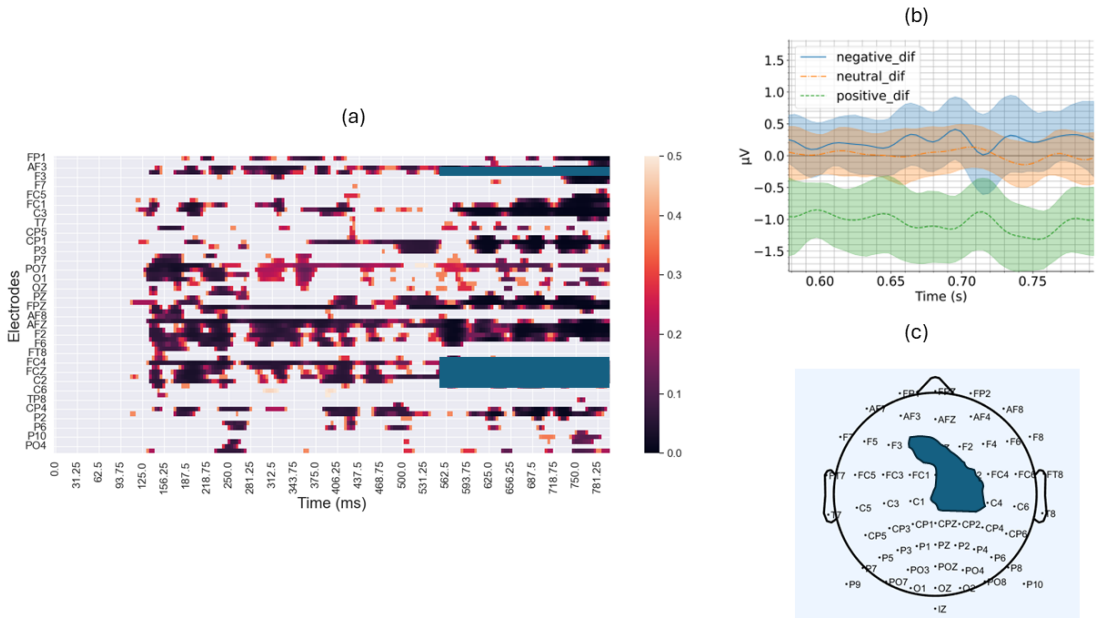
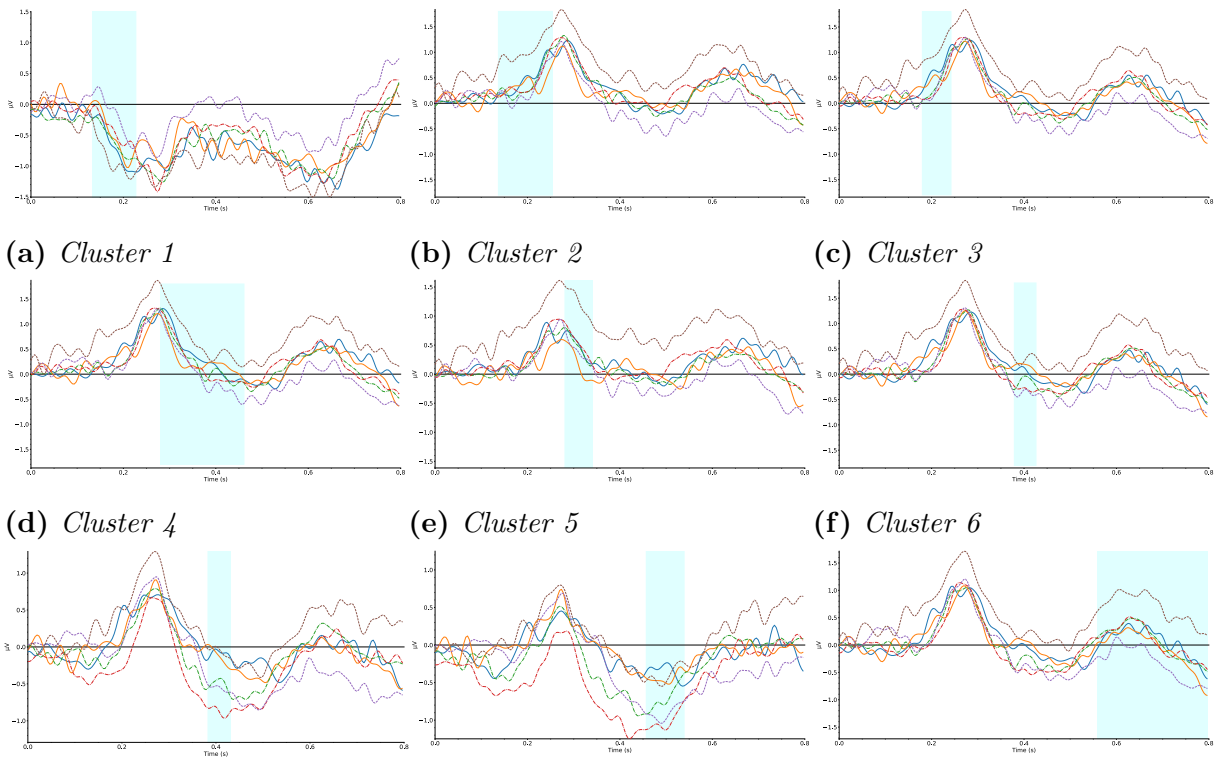


Figura 5

Raster plots of clusters found by massive univariate analysis, setting the decision threshold at a p-value of ,5. Figure (a) shows the raster plot, highlighting by rectangles the electrodes (F1, FZ, FCZ, FC2, CZ, and C2) and time window (from 558 to 800 ms) that make up cluster 9. Figure (b) shows the signals resulting from averaging the difference between the ERPs of the congruent and incongruent blocks for each group, including the confidence intervals. Figure (c) shows the location of the electrodes that make up the cluster.



(a) *Cluster 1* **(b)** *Cluster 2* **(c)** *Cluster 3*

(d) *Cluster 4* **(e)** *Cluster 5* **(f)** *Cluster 6*

(g) *Cluster 7* **(h)** *Cluster 8* **(i)** *Cluster 9*

Figura 6

Grand averages of clusters found in the first stage of the proposed methodology. The shaded area indicates the cluster time window. Conventions: Solid blue line: congruent block - negative IAT score group. Solid yellow line: congruent block - negative IAT score group. Green dash-dot line: congruent block - neutral IAT score group. Red dash-dot line: incongruent block - neutral IAT score group. Purple dashdot line: congruent block - positive IAT score group. Brown dashed line: incongruent block - positive IAT score group.

Tabla 5

Clusters found in the first step of the proposed methodology.

Cluster id.	Electrodes	Time window (ms)	Effects detected in blocks:
1	P9, PO7	132,812 – 226,562	Differences
2	AF4, F2, F4, FC4	136,719 – 253,906	Differences
3	FC1, FCZ, FC2, FC4, F2, F4	179,688 – 242,188	Differences
4	AF4, AFZ, FPZ	281,25 – 339,844	Differences
5	FC2, FC4, F2, F4, FZ	281,25 – 460,938	Differences
6	FC2, FC4, FCZ	378,906 – 425,781	Differences
7	CZ, C2, C1, CP1	382,812 – 429,688	Differences
8	CPZ, CP1, CP2	457,031 – 539,062	Differences
9	FC2, FCZ, CZ, C2, FZ, F1	558,594 – 796,875	Differences

3.4.2 Extended Analyses of One Cluster

Given the extent of the analyses performed for each cluster, we only present the complete results for cluster 9. A summary of the results for all clusters and the statistical analyses that support these results are presented in Section 3.4.3.

For cluster 9, composed of electrodes FC2, FCZ, CZ, C2, FZ, and F1, in the time window between 559 and 797 ms, the comparison of repeated measures ANOVA model provided strong evidence that the model that best predicts the data is the complete model with the two factors and the interaction between factors (see Table 6).

Furthermore, the effects analysis (see Table 7) gives strong evidence in favor of including the two factors and the interaction. These results are similar with different priors, indicating that it is a robust model. With these results, we can proceed to post-hoc analyses. In this table, BF_{inc} shows the evidence favoring including the respective predictor.

The analysis of the posterior distributions of the parameters, as well as the interaction plot depicted in Figure 7, and the credibility intervals for each group shown in Table 8,

Tabla 6

Comparisons of models for cluster 9 with Bayesian Repeated Measures ANOVA.

Models	BF_m prior		
	r scale = 0,2	r scale = 0,5	r scale = 0,8
block type + IAT effect + block type * IAT effect	264,542	205,903	114,967
block type	0,034	0,052	0,092
block type + IAT effect	0,016	0,013	0,034
Null model (incl. subject and random slopes)	0,007	0,009	0,009
IAT effect	0,003	0,002	0,003

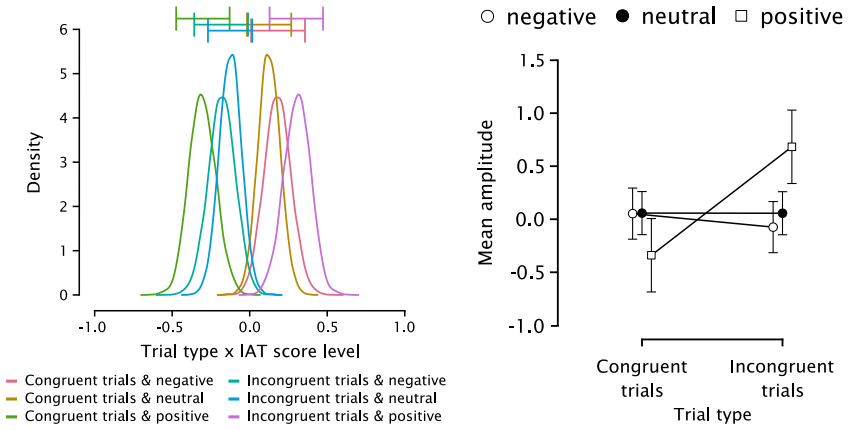
Note. BF_m shows how much the data have changed the prior model odds. All models include subject and random slopes for all factors of repeated measures.

Tabla 7

Analysis of effects for cluster 9 with Bayesian Repeated Measures ANOVA.

Effects	BF_{inc} prior		
	r scale = 0,2	r scale = 0,5	r scale = 0,8
block type	259,286	171,542	73,099
IAT effect	64,616	40,606	21,025
block type * IAT effect	264,542	205,903	114,967

Note. BF_{inc} shows the evidence favoring including the respective predictor.



(a) **(b)**

Figura 7

Plots of (a) posterior distributions of the groups formed by the block type × IAT effect interaction, and (b) block type × IAT effect group interaction plots including 95 % *credible intervals. Note that the incongruent block measures of the positive IAT effect subjects are significantly more positive than those of the other groups.*

indicate that the mean amplitudes of the incongruent block for subjects in the positive IAT effect group are significantly more positive both the congruent block within the same group and the incongruent block within the neutral and negative groups.

The results of the posthoc analyses (see Table 9) indicate that there is decisive evidence supporting the hypothesis that the mean amplitudes of incongruent block of subjects in the positive group are more positive than those of congruent block of subjects in the same group. There is also moderate evidence supporting the hypothesis that the mean amplitudes of the incongruent block of subjects in the positive group are more positive than those of the incongruent block of subjects in the negative and neutral groups.

The sensitivity analysis of Figure 8 indicates that the model is not robust when it compares the incongruent-positive and incongruent-neutral groups. The inference made is weak because it depends on the prior selection.

On the other hand, the Bayesian one-way ANOVA performed on the difference between congruent and incongruent blocks showed that the model that includes the IAT effect factor as a predictor explains the data better than the null model, which, in other words, means that there is decisive evidence in favor of the alternative hypothesis and that the model is robust concerning the selection of priors (view Table 10).

Tabla 8

Descriptives of posterior distributions, Bayesian Repeated Measures ANOVA for cluster 9.

block type	IAT effect level	*M*	*SD*	*N*	95 % Credible Interval	
					Lower	Upper
congruents	negative	0,053	1,176	19	−0,514	0,620
	neutral	0,058	1,314	37	−0,380	0,496
	positive	−0,340	0,948	29	−0,700	0,021
incongruents	negative	−0,074	1,006	19	−0,559	0,411
	neutral	0,057	1,212	37	−0,347	0,460
	positive	0,683	1,152	29	0,245	1,121

Tabla 9

Bayes Factors of post hoc comparisons in cluster 9.

Bayesian Test	Measure 1	Measure 2	BF_{-0}	error %
Paired Samples T-Test	Congruent positive	Incongruent positive	282,879	2.570×10^{-5}
Independent Samples T-Test	Incongruent neutral	Incongruent positive	3,274	4.882×10^{-5}
Independent Samples T-Test	Incongruent negative	Incongruent positive	4,932	7.133×10^{-5}

Note. BF_{-0} is the Bayes factor of the alternative hypothesis versus the null hypothesis, the alternative hypothesis being that measure 1 is smaller than measure 2. The percentage error indicates how much the results can fluctuate since these analyses are based on a numerical algorithm such as Markov chain Monte Carlo (MCMC).

Tabla 10

Model comparisons for cluster 9 with Bayesian one-way ANOVA.

Models	BF_M prior		
	r scale = 0,2	r scale = 0,5	r scale = 0,8
IAT effect	257,781	399,925	369,558
Null model	0,004	0,003	0,003

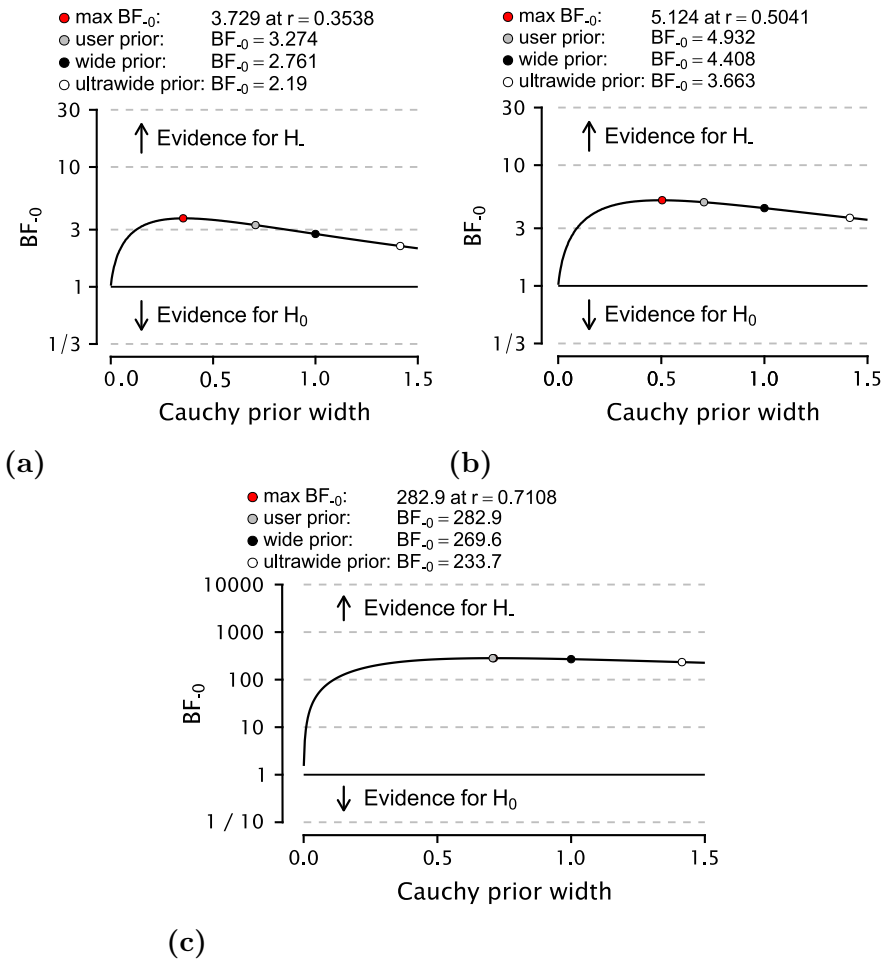


Figura 8

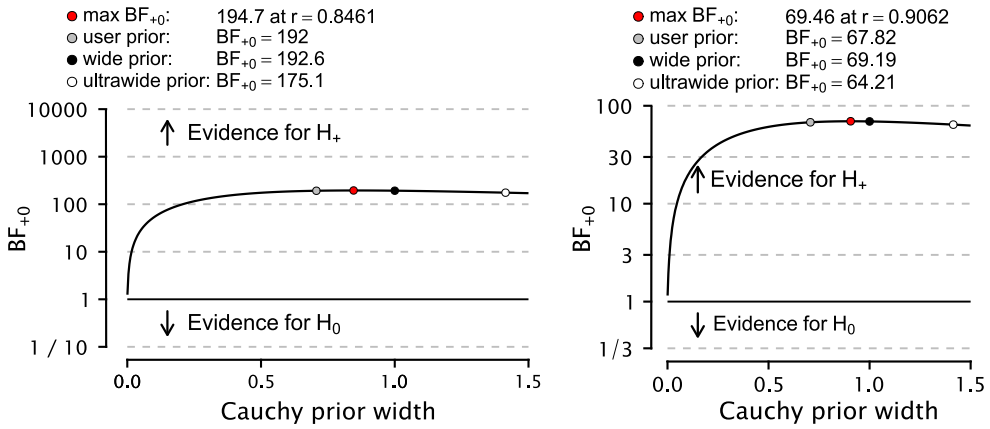
Sensitivity analyses of post-hoc comparisons repeated measures ANOVA cluster 9 for (a) Incongruent block, Positive vs. Neutral, (b) Incongruent block, Positive vs. Negative, and (c) Congruent vs. Incongruent block, Positive. In all figures, ***max*** $\mathbf{BF_{-0}}$ *is the maximum Bayes factor that can be obtained with prior distribution whose parameter is the indicated value of r,* ***user prior*** *indicates the Bayes factor obtained with the prior distribution used,* ***wide prior*** *indicates the Bayes factor that would be obtained with an uninformative prior distribution, and* ***ultrawide prior*** *indicates the Bayes factor that would be obtained with a non-informative prior distribution.*

Tabla 11

Descriptives of posterior distributions, Bayesian one-way ANOVA.

IAT effect level	*M*	*SD*	*N*	95 % Credible Interval	
				Lower	Upper
negative	0,127	0,707	19	−0,213	0,468
neutral	0,001	0,861	37	−0,286	0,288
positive	−1,022	1,286	29	−1,511	−0,533

Post hoc comparisons show that the measure in the positive group is more negative than in the neutral and negative groups. The sensitivity analysis shows this result is robust (Figure 9). Observation of the credibility intervals of the groups corroborates this statement (Table 11).



(a) **(b)**

Figura 9

Sensitivity analyses of post hoc comparisons one-way ANOVA, cluster 9 for (a) Positive vs. Neutral IAT score groups and (b) Positive vs. Negative IAT score groups. The same conventions in Fig. 8 stand.

The above results show that for this cluster, in subjects of the positive IAT effect group, the average amplitudes of the incongruent block are significantly more positive than those of the congruent block. In the other groups, the mean amplitudes in the two types of blocks tend to be similar. In the incongruent block, the average amplitudes of the positive group are more positive than those of the other groups.

3.4.3 Cluster Analysis Summary

In summary, we found the following:

1. In the left parietal region between 132 and 227 ms, the mean amplitudes of the incongruent block of the positive IAT score group are more negative than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. In Figure 10, it can be seen that the topographic map of the incongruent block of the positive score group at 190 ms (Subfigure 10d) is a darker blue in this region, concerning the topographic maps of the other mentioned block×group interactions, which denotes its greater negativity.

2. In the right frontal region between 137 and 254 ms, the mean amplitudes of the incongruent block of the positive IAT score group are more positive than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. On the other hand, in the midline frontal region between 180 and 242 ms, the mean amplitudes of the incongruent block of the positive IAT score group are more positive than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. In Figure 10, it can be seen that the topographic map of the incongruent block of the positive score group at 190 ms (Subfigure 10d) is a darker reed in these regions, concerning the topographic maps of the other mentioned block*group interactions, which denotes its greater positivity.

3. In the right prefrontal and frontal regions between 281 and 340 ms, the mean amplitudes of the incongruent block of the positive IAT score group are more positive than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. In Figure 10, it can be seen that the topographic map of the incongruent block of the positive score group at 290 ms (Subfigure 10d) is a darker reed in these regions, concerning the topographic maps of the other mentioned block×group interactions, which denotes its greater positivity.

4. In the right fronto-central region between 340 and 461 ms, the mean amplitudes of the incongruent block of the positive IAT score group are more positive than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. In Figure 10, it can be seen that the topographic map of the incongruent block of the positive score group at 400 ms (Subfigure 10d) is a darker red in this region, concerning the topographic maps of the other mentioned block×group interactions, which denotes its greater positivity.

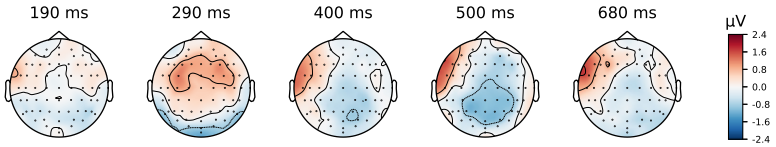
5. In the left central parietal region between 383 and 430 ms, the mean amplitudes of the incongruent block of the neutral IAT score group are more negative than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. In Figure 10, it can be seen that the topographic map of the incongruent block of the neutral score group at 400 ms (Subfigure 10e) is a darker blue in this region, concerning the topographic maps of the other mentioned block×group interactions, which denotes its greater negativity.

6. In the central parietal region between 457 and 539 ms, the mean amplitudes of the incongruent block of the neutral IAT score group are more negative than those of the congruent block of this same group. In contrast, the mean amplitudes of the incongruent block of the positive IAT score group are more positive than those of the congruent block of this same group. This effect can be seen in Figure 10 by comparing the topographic maps of these groups at 500 ms (Subfigure 10e versus Subfigure 10b, and Subfigure 10d versus Subfigure 10a).

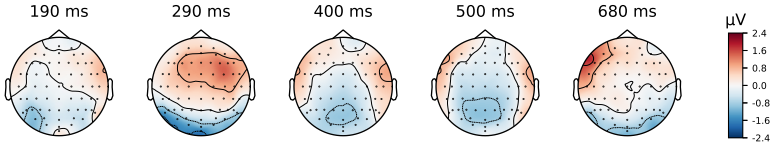
7. In the fronto-central region between 559 and 797 ms, the mean amplitudes of the incongruent block of the positive IAT score group are more positive than those of the congruent block of this same group and those of the incongruent blocks of the other two groups. In Figure 10, it can be seen that the topographic map of the incongruent block of the positive score group at 680 ms (Subfigure 10d) is a darker red in this region, concerning the topographic maps of the other mentioned block×group interactions, which denotes its greater positivity.

The statistical analyses that support these statements are presented below.

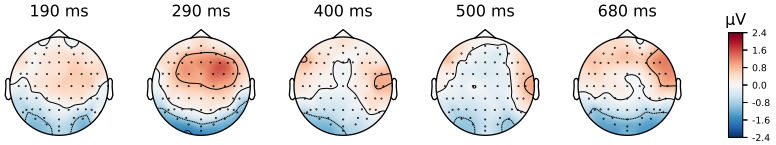
Cluster 1: There is decisive evidence (BF = 132,827) that the mean amplitudes of the incongruent block of the positive IAT effect group (CrI = [−1,543, −0,387]) are more negative than the mean amplitudes of the congruent block of the same group (CrI = [−0,682, 0,232]). There is decisive evidence (BF = 478,679) that the mean amplitudes of the differences among congruent and incongruent blocks of the positive IAT effect group (CrI = [0,264, 0,819]) are more positive than the mean amplitudes of the differences among congruent and incongruent blocks of the neutral group (CrI = [−0,629, −0,113]). There is strong evidence (BF = 9,361) that the mean amplitudes of the differences among congruent and incongruent blocks of the positive IAT effect group (CrI = [0,264, 0,819]) are more positive than the mean amplitudes of the differences among the congruent and incongruent blocks of the negative group (CrI = [−0,474, 0,118]).



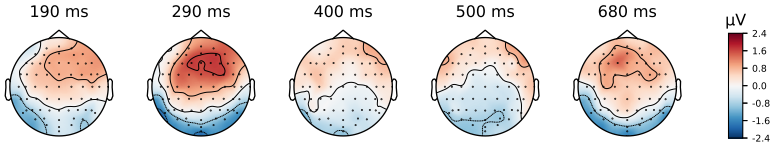
(a) *Topographic map of congruent block, positive IAT score group*



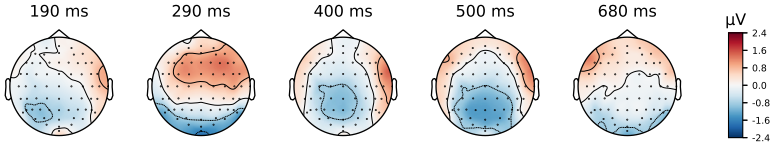
(b) *Topographic map of congruent block, neutral IAT score group*



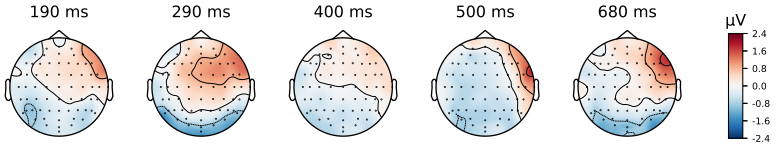
(c) *Topographic map of congruent block, negative IAT score group*



(d) *Topographic map of incongruent block, positive IAT score group*



(e) *Topographic map of incongruent block, neutral IAT score group*



(f) *Topographic map of incongruent block, negative IAT score group*

Figura 10

Topographic maps of clusters found with the proposed methodology. According to clusters found, these are the topographic maps of each block×group's interaction at selected times.

Cluster 2: There is decisive evidence (BF = 882,991) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}688, 1{,}421]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [0{,}002, 0{,}569]$). There is strong evidence (BF = 20,128) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}688, 1{,}421]$) are more positive than the mean amplitudes of the incongruent block of the neutral group ($CrI = [0{,}117, 0{,}669]$). There is moderate evidence (BF = 9,448) that the mean amplitudes of the incongruent block of the positive group ($CrI = [0{,}688, 1{,}421]$) are more positive than the mean amplitudes of the incongruent block of the negative group ($CrI = [-0{,}016, 0{,}719]$). There is strong evidence (BF = 53,400) that the mean amplitudes of the differences among congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}101, -0{,}437]$) are more negative than the mean amplitudes of the differences among congruent and incongruent blocks of the neutral group ($CrI = [-0{,}273, 0{,}295]$). There is strong evidence (BF = 19,832) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}101, -0{,}437]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the negative group ($CrI = [-0{,}341, 0{,}602]$).

Cluster 3: There is strong evidence (BF = 40,607) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}647, 1{,}472]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [0{,}065, 0{,}708]$). There is strong evidence (BF = 23,345) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}647, 1{,}472]$) are more positive than the mean amplitudes of the incongruent block of the neutral group ($CrI = [0{,}024, 0{,}612]$). There is strong evidence (BF = 44,324) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-0{,}772, -0{,}177]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the neutral group ($CrI = [-0{,}057, 0{,}474]$). There is strong evidence (BF = 14,898) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-0{,}772, -0{,}177]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the negative group ($CrI = [-0{,}049, 0{,}580]$).

Cluster 4: There is moderate evidence (BF = 9,768) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}359, 1{,}337]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [-0{,}345, 0{,}340]$). There is moderate evidence (BF = 3,546) that the mean amplitudes of

the incongruent block of the positive IAT effect group ($CrI = [0{,}359, 1{,}337]$) are more positive than the mean amplitudes of the incongruent block of the negative group ($CrI = [-0{,}212, 0{,}422]$). There is moderate evidence ($BF = 6{,}088$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}383, -0{,}317]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the neutral group ($CrI = [-0{,}270, 0{,}277]$). There is strong evidence ($BF = 13{,}546$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}383, -0{,}317]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the negative group ($CrI = [-0{,}268, 0{,}502]$).

Cluster 5: There is moderate evidence ($BF = 26{,}580$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}778, 1{,}743]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [0{,}253, 0{,}983]$). There is moderate evidence ($BF = 6{,}606$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}778, 1{,}743]$) are more positive than the mean amplitudes of the incongruent block of the neutral group ($CrI = [0{,}454, 1{,}089]$). There is moderate evidence ($BF = 4{,}774$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}778, 1{,}743]$) are more positive than the mean amplitudes of the incongruent block of the negative group ($CrI = [0{,}077, 0{,}983]$). There is strong evidence ($BF = 18{,}514$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}113, -0{,}173]$) are more negative than the mean amplitudes of the differences among congruent and incongruent blocks of the neutral group ($CrI = [-0{,}293, 0{,}249]$). There is strong evidence ($BF = 25{,}116$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}113, -0{,}173]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the negative group ($CrI = [-0{,}177, 0{,}889]$).

Cluster 6: There is moderate evidence ($BF = 24{,}732$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [-0{,}163, 0{,}757]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [-0{,}792, 0{,}006]$). There is moderate evidence ($BF = 3{,}568$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [-0{,}163, 0{,}757]$) are more positive than the mean amplitudes of the incongruent block of the neutral group ($CrI = [-0{,}663, 0{,}042]$). There is strong evidence ($BF = 24{,}770$) that the mean amplitudes of the

differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}127, -0{,}253]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the neutral group ($CrI = [-0{,}198, 0{,}533]$).

Cluster 7: There is moderate evidence ($BF = 3{,}543$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [-0{,}615, 0{,}492]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [-1{,}051, -0{,}087]$). There is moderate evidence ($BF = 3{,}231$) that the mean amplitudes of the incongruent block of the neutral IAT effect group ($CrI = [-1{,}249, -0{,}502]$) are more negative than the mean amplitudes of the congruent block of the same group ($CrI = [-0{,}870, -0{,}137]$). There is moderate evidence ($BF = 7{,}694$) that the mean amplitudes of the incongruent block of the neutral IAT effect group ($CrI = [-1{,}249, -0{,}502]$) are more negative than the mean amplitudes of the congruent block of the positive group ($CrI = [-0{,}615, 0{,}492]$). There is moderate evidence ($BF = 8{,}084$)that the mean amplitudes of the incongruent block of the neutral IAT effect group ($CrI = [-1{,}249, -0{,}502]$) are more negative than the mean amplitudes of the congruent block of the positive group ($CrI = [-0{,}508, 0{,}251]$). There is moderate evidence ($BF = 3{,}568$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [-0{,}163, 0{,}757]$) are more positive than the mean amplitudes of the incongruent block of the neutral group ($CrI = [-0{,}663, 0{,}042]$). There is strong evidence (($BF = 36{,}058$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-0{,}963, -0{,}051]$) are more positive than the mean amplitudes of the differences among the congruent and incongruent blocks of the neutral group ($CrI = [0{,}037, 0{,}706]$).

Cluster 8: There is strong evidence ($BF = 15{,}234$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}048, 0{,}013]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the neutral group ($CrI = [-0{,}001, 0{,}639]$).

Cluster 9: There is decisive evidence ($BF = 282{,}879$) that the mean amplitudes of the incongruent block of the positive IAT effect group ($CrI = [0{,}245, 1{,}121]$) are more positive than the mean amplitudes of the congruent block of the same group ($CrI = [-0{,}700, 0{,}021]$). There is moderate evidence ($BF = 3{,}274$) that the mean amplitudes of the incongruent block of the positive group ($CrI = [0{,}245, 1{,}121]$) are more positive than the mean amplitudes of the incongruent block of the neutral group ($CrI = [-0{,}347, 0{,}460]$). There is moderate evidence ($BF = 4{,}932$) that the mean amplitudes of the incongruent

block of the positive group ($CrI = [0{,}245, 1{,}121]$) are more positive than the mean amplitudes of the incongruent block of the negative group ($CrI = [-0{,}559, 0{,}411]$). There is decisive evidence ($BF = 192{,}0$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}511, -0{,}533]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the neutral group ($CrI = [-0{,}286, 0{,}288]$). There is strong evidence ($BF = 67{,}82$) that the mean amplitudes of the differences among the congruent and incongruent blocks of the positive IAT effect group ($CrI = [-1{,}511, -0{,}533]$) are more negative than the mean amplitudes of the differences among the congruent and incongruent blocks of the negative group ($CrI = [-0{,}213, 0{,}468]$).

Accounting for all clusters, we found the following effects: (i) In a negative polarity segment in the left occipital parietal region between 133 and 227 ms. (ii) In a segment of positive polarity in the right frontal region between 137 and 254 ms, in the frontal central region between 180 and 242 ms, and in the right frontal and prefrontal regions between 281 and 340 ms. (iii) In a segment of negative polarity in the prefrontal region between 340 and 461 ms, in the right frontal region between 379 and 426 ms, and in the central parietal region between 383 and 430 ms. (iv) In a segment of positive polarity in the central parietal region between 457 and 539 ms, and in the frontal central region between 559 and 797 ms. Some of these effects may be part of the same ERP component, but determining that is beyond the scope of this work.

3.5 Discussion

With the proposed methodology, we found effects by levels of prejudice in the parietal-occipital region between 133 and 227 ms in a segment of negative polarity (cluster 1), consistent with findings of Chen et al. (2018) in the N2 component. In contrast, this component did not have any effects with the other methods. Concerning the P3 component, we found effects in the interaction block type × IAT effect group in the frontal, central region (clusters 4 and 5), consistent with the results of Healy et al. (2015) and Portengen et al. (2022). We also found an effect with the traditional methodology (time window 2) consistent with one result of the new methodology (cluster 5). We did not find effects with the massive univariate methodology.

About the N400 component, we found effects in the interaction block type × IAT effect group in the frontal and prefrontal regions (clusters 5 and 6), which is consistent

with the results of Williams and Themanson (2011), and in the central parietal region (clusters 7), which is consistent with findings in Healy et al. (2015). We also found an effect with the methodology based on massive univariate analyses (cluster 1) and the traditional methodology in the N400 component (time window 3). Concerning the LPP component, we found effects in the interaction block type × IAT effect group in the central parietal region (clusters 8 and 9), consistent with Williams and Themanson (2011) and Forbes et al. (2012). We also found effects in the LPP component, both with the traditional methodology (time window 4) and the methodology based on massive univariate analyses (cluster 2).

Finally, we found that in the frontal, central, and parietal regions, there are effects in the interaction block type × IAT effect group in a positive polarity segment that could be assimilated with a P3 component (clusters 2 and 3). Other authors could not find it in their ERP-IAT studies, but the effects of this component have been reported in other social cognition tasks (Amodio and Cikara, 2021; Portengen et al., 2022). We also found an effect in this component with the traditional methodology (time window). However, because the effect is more substantial on other electrodes, it is not well observed on the pre-selected ones.

Regarding the novelty of our methodology, to our knowledge, only one study has combined massive univariate methods and Bayesian inference methods (Schindler et al., 2018) to test the effects on ERP components. Our proposal differs from theirs, as we modify the parameters of the massive univariate analysis to decrease Type II errors and find additional potential effects verified by a Bayesian analysis.

Our thesis's proposed methodology outperformed the traditional one based on an a priori selection of ERP components. Additionally, it allowed us to find effects in ERP components other than the pre-selected ones. Moreover, the novel methodology provides higher precision in the location of ERP components because it identifies the groups of electrodes and time windows with the most potent effects. For example, in the case of the P3 components, it was possible to establish that the effects are more substantial in the left frontal and parietal regions and not in the central frontal region. It could be argued that in the traditional methodology, many regions of interest covering a large region of interest could be pre-selected, but this would add a factor to the analysis that would increase the risk of increasing type I or type II errors, depending on the strategy adopted to deal with the problem of multiple comparisons, as explained by Luck and Gaspelin (2017). As expected, this strategy could not accurately locate the effect obtained with the new methodology.

The proposed methodology outperformed the data-driven one based on a massive univariate analysis, where only three effects were identified, suggesting that the clustering threshold ($p < ,05$) is too restrictive. However, this threshold could not be extended in the univariate analysis because it would go against the frequentist approach on which it is based, in which this value is the criterion for accepting or rejecting an effect. With the Bayesian approach proposed by us, we can extend this threshold because we are only using it to establish an a priori distribution of the parameters of our hypotheses. Thus, the threshold of 0,5 that we use is interpreted as meaning that the probability that an effect exists in that cluster (alternative hypothesis) is at least equal to the probability that it does not exist (null hypothesis). Therefore, the data will tell which hypothesis is more probable.

These results suggest that our novel methodology has a higher sensitivity than traditional ones and that it is possible to find more true effects with the same data without paying the cost of losing specificity because the Bayesian analysis filters out non-existent effects, as evidenced by the results of several clusters.

The following chapter will cover a different method for analyzing EEG signals, employing the frequency domain. ERP analysis has provided information on the activation patterns of particular brain regions during specific time intervals. Subsequent analyses will provide information on the reconfiguration of the participants'brain networks in response to the tasks they are performing.

4 A Novel Methodology to Perform EEG-based Functional Connectivity Analyses

This chapter presents an EEG-based functional connectivity analysis to characterize prejudice. For this purpose, we compared seven topology measures from brain networks in six different frequency bands and two different experimental conditions. We used Bayesian inference methods for statistical analyses and found biases toward victims from the same victims and biases of civilians towards ex-combatants. For the connectivity measures, we found differences in two different frequency bands. Finally, we found a slight correlation between the behavioral score to measure prejudice and two measures of the theta-band network.

4.1 Literature Review

Prejudice is an individual attitude towards a group and its members that creates or maintains hierarchical status relationships between groups and leads to discriminatory judgments and actions. Research on the neurological basis of prejudice has elucidated how prejudice is formed, represented in the mind, expressed behaviorally, and reduced (Rösler and Amodio, 2022).

Studies with EEG-ERP and fMRI have shown that social identity is quickly recognized in the brain; thus, in just 100 ms (N100), a person has already categorized another based on race, ethnicity, gender, or even arbitrary social categories (Ito and Urland, 2003; Rösler and Amodio, 2022). After this comes a top-down processing of social identity between 180 and 200 ms (P200), which depends on the person's explicit and implicit goal-directed goals (Amodio and Cikara, 2021; Volpert-Esmond and Bartholow, 2019).

Three regions were observed to be involved in these two processes: the orbitofrontal cortex (OFC), the fusiform gyrus (FG), and the anterior temporal lobe (ATL). The FG is involved in the visual processing of faces, and the ATL retrieves social-conceptual associations related to perceived characteristics such as social categories (e.g., stereotypes, person knowledge). The OFC may then use such social-conceptual information to implement top-down visual predictions that modulate FG representations of faces in line with those predictions. This network supports rapid and flexible integration of bottom-up facial cues and higher-order social cognitive processes (Freeman and Johnson, 2016).

Depending on the task, activations have been observed around 260 ms originating in the dorsal anterior cingulate cortex (dACC), thought to be related to conflict and response selection processes (Amodio and Cikara, 2021) and in the medial prefrontal cortex (MPFC), a region related to mentalization (Molenberghs and Louis, 2018). Activations have also been observed around 450 to 600 ms (P300) in sorting behaviors involving rapid response evaluation (Molenberghs and Louis, 2018), in a vast area involving the anterior cingulate cortex (ACC), a region usually related to conflict monitoring, and the prefrontal cortex (PFC), which in this case is associated with regulative control (Ito and Bartholow, 2009).

Other studies of prejudice other than race have shown that other regions, such as the amygdala (Molenberghs and Louis, 2018), striatum, and insula (Amodio et al., 2014), are also involved in prejudice. It has also been found that the more "real" the prejudice (e.g., prejudice based on ethnicity, nationality, or politics), the broader the brain areas involved (Saarinen et al., 2021). From all this, it concluded that there are two groups of brain regions involved in prejudice: those where it originates, which include the fusiform gyrus, amygdala, ATL, striatum, and insula, and those where it is regulated or attempted to be reduced, which include the ACC and PFC (Rösler and Amodio, 2022).

Although functional connectivity based on EEG has already been used in studies of mental illness (Stam et al., 2014) – such as epilepsy (van Diessen et al., 2016), Alzheimer's disease (Cui et al., 2018), depression (Bankwitz et al., 2023), and bipolar disorder (Zhang et al., 2022) – and emotional processing (Cao et al., 2020), to our knowledge, it has not been used in the study of social behaviors. This technique could provide additional information on how these brain regions are reconfigured in prejudice events.

Previous studies using EEG-based functional connectivity suggest that training in social cognition can improve emotional recognition and modify brain connectivity (Quintero-Zea et al., 2019; Valencia et al., 2020; Trujillo et al., 2017). Using this technique to characterize prejudice among actors in the Colombian armed conflict is expected to provide valuable information that may help us design more specific psychosocial intervention strategies to mitigate prejudice.

4.2 Methodology

For statistical analyses, the IAT test scores and network measures were compared among the groups described in the subsection 2.1.1.

The spectrum of the signals was calculated using the multitaper method. We use the Weighted Phase Lag Index (Vinck et al., 2011) to build the connectivity matrix. We built connectivity matrices for each subject's ERP and IAT block (congruent and incongruent) on the following frequency bands: Delta (0 - 4 Hz), Theta (4 - 7 Hz), Alpha (7 - 13 Hz), Beta1 (13 - 20 Hz), Beta2 (20 - 30 Hz), and Gamma (30 - 40 Hz). Estimating the spectrum and constructing the connectivity matrices were done in the `MNE` package V. 0.23 (Gramfort, 2013).

Once the connectivity matrices were obtained, we extracted the MST using the Kruskal method (Kruskal, 1956). Then, for each subject per block and frequency band, the following global network measures were calculated: leaf fraction, diameter, mean eccentricity, maximum degree, maximum betweenness centrality, and tree hierarchy. The extraction of MST and the calculation of network measures were done with the `NetworkX` package V. 2.7 (Hagberg et al., 2008).

With the IAT task scores, ANOVA and post hoc Bayesian inference t-tests were performed to test whether there were statistically significant differences between groups. On the other hand, the following analyses were performed to test whether there are differences among groups in the network measures.

- Mixed ANOVA by Bayesian inference taking as a within-subjects factor the block type (congruent or incongruent) and as a between-subjects factor the network measure. As a priori parameter, we will consider a small effect size (Cohen's d = 0,25).
- One-way ANOVA, taking the difference between the congruent and incongruent network measures block as a dependent variable. As a priori parameter, we considered a small effect size (Cohen's d = 0,25).
- t–Post hoc analysis for those factors or interactions considered relevant in the models, adjusting the priors according to the Westfall et al. (1997)'procedure to correct for multiple comparisons.

Finally, we performed correlation analyses between IAT scores and network measures in which effects were found to determine whether there is a relationship between behavioral and electrophysiological outcomes.

4.3 Results

4.3.1 Behavioral Results

For the IAT D-score, the best-fitting model explaining the observed mean amplitudes includes measures of the actors'group (BF_M = 2,466, $BF_{nullmodel}$ = 0,405). There is strong evidence in favor of the hypothesis that the IAT D-scores of the victims tend to be more positive than those of civilians (BF = 10,838) and ex-guerrillas (BF = 8,977). Table 12 displays the descriptive statistics and 95 % credible intervals for each actor group's IAT D-score posterior.

Tabla 12

Descriptives - IAT score

				95 %CrI	
Actors'group	Mean	SD	N	Lower	Upper
civilians	−0,121	0,355	15	−0,317	0,076
ex-guerrillas	−0,031	0,233	22	−0,135	0,072
ex-paramilitaries	0,049	0,381	31	−0,091	0,189
victims	0,185	0,252	23	0,076	0,294

4.3.2 Functional Connectivity Results

Statistically significant differences were found in the theta and beta2 bands. No effect was found in any other band. The effects found are presented below.

Theta band (4 - 7 Hz)

We found significant differences in mean diameters and eccentricity between congruent and incongruent blocks. For diameter, the model that best explains the data is the one that includes the actor group as a factor (BF = 11,663), and the post hoc analyses showed that there is moderate evidence in favor of the hypothesis that victims'differences are more positive than the ex-guerrillas'differences (BF = 5,020) and civilians'differences (BF = 3,703). We also found moderate evidence that the differences of ex-paramilitaries are more positive than the differences of ex-guerrillas (BF = 6.854)

and the differences of civilians (BF = 4,847). Finally, we found moderate evidence that the ex-paramilitariesánd victims'differences are similar (BF = 0,285) and that the civiliansánd ex-guerrillas'differences are similar (BF = 0,323). The pointplot (Figure 11(a)) shows that for victims and ex-paramilitaries, the mean diameters of the congruent block are larger than the ones of the incongruent block. In contrast, for civilians and ex-guerrillas, the mean diameters of the congruent block are smaller than those of the incongruent one.

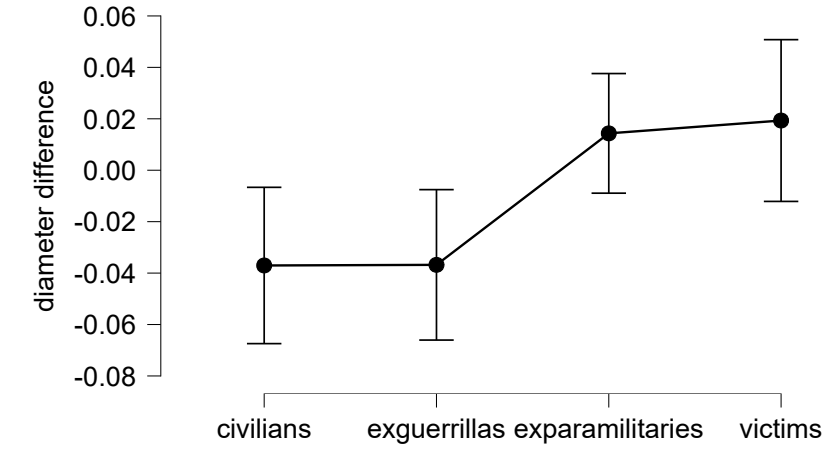
For eccentricity, the model that best explains the data is the one that includes the actor group as a factor (BF = 3,038). Post hoc analyses showed moderate evidence in favor of the hypothesis that ex-paramilitaries'differences are more positive than ex-guerrilla ones (BF = 3,255). We also found moderate evidence that the differences between ex-paramilitaries and victims are similar (BF = 0,292) and that the differences between civilians and ex-guerrillas are similar (BF = 0,329). The interaction diagram (Figure 11(b)) shows that for victims and ex-paramilitaries, the mean diameters of the congruent block are larger than those of the incongruent block. In contrast, for civilians and ex-guerrillas, the mean diameters of the congruent block are smaller than those of the incongruent one.

Beta2 band (20 - 30 Hz)

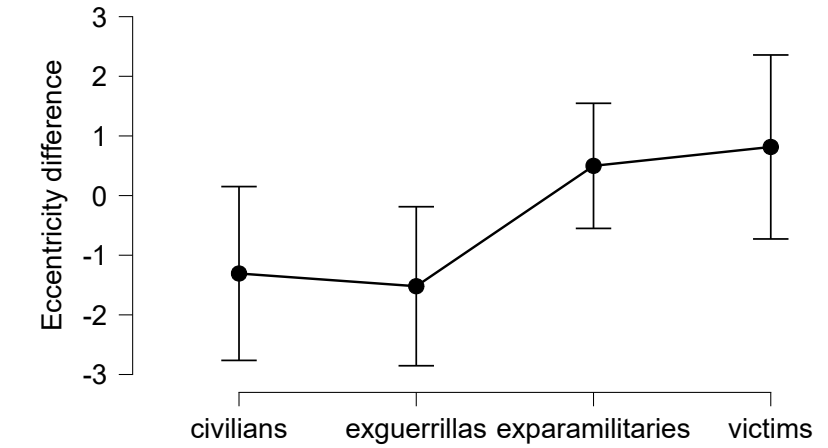
We observed effects in the mean diameter and mean eccentricity (and their differences) between congruent and incongruent blocks. For diameter, the model that best explains the data is the one that includes the actor group as a factor (BF = 5,230), and the post hoc analyses showed that there is moderate evidence in favor of the hypothesis that ex-paramilitary differences are more negative than the ex-guerrilla differences (BF = 7,849) and victims'differences (BF = 6,427). Additionally, we found moderate evidence that the differences between the ex-guerrillas and victims are similar (BF = 0,295).

Concerning the mean diameter in the incongruent block, the model that best explains the data is the one that includes the actor group as a factor (BF = 48,582), and the post hoc analyses showed that there is strong evidence in favor of the hypothesis that ex-paramilitary diameters are more positive than ex-guerrillas'diameters (BF = 29,619) and victims'diameters (BF = 44,733). Additionally, we found moderate evidence that the diameters of the ex-guerrillas and victims are similar (BF = 0,322).

The pointplot (Figure 12(a)) shows that for victims and former guerrillas, the mean diameters of the congruent block are bigger than the mean diameters of the incongruent one. In contrast, for civilians and ex-paramilitaries, the mean diameters of the congruent block are smaller. Regarding eccentricity, the model that best explains the data is the one



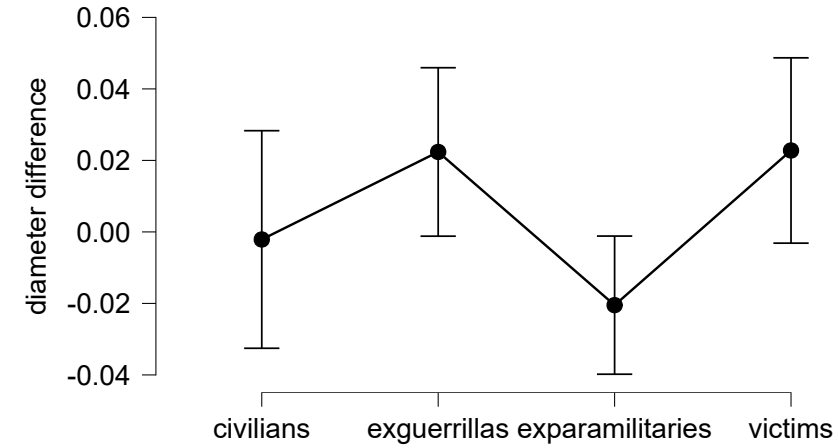
(a)



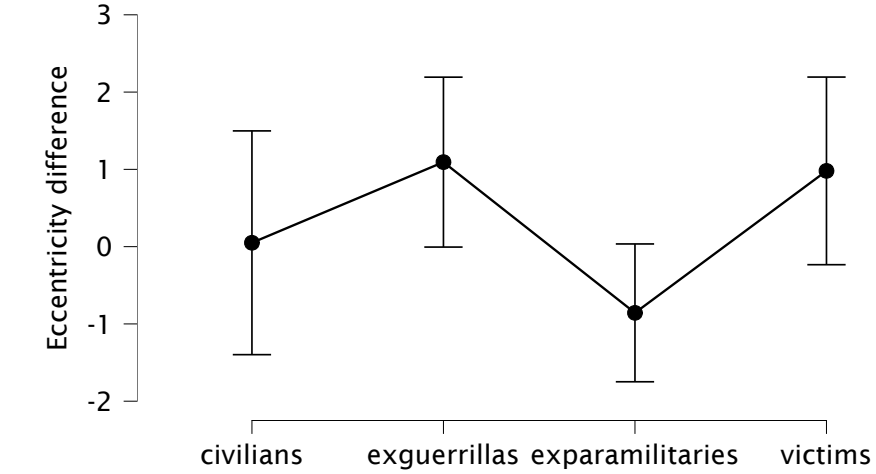
(b)

Figura 11

Pointplots of differences between (a) diameter and (b) eccentricity means of congruent and incongruent blocks in the theta band. The plots include 95 % *credible intervals. For victims and ex-paramilitaries, the mean diameters and eccentricities of the congruent block are larger. As a result, the differences tend to be positive. In contrast, the differences tend to be negative for civilians and ex-guerrillas.*



(a)



(b)

Figura 12

Pointplots of (a) differences between diameter means of congruent and incongruent blocks, and (b) differences between eccentricity means of congruent and incongruent blocks in beta2 band. The plots include $95\,\%CrI$*. For victims and ex-guerrillas, the mean diameters and eccentricities of the congruent block are larger. As a result, the differences tend to be positive. Conversely, for ex-paramilitaries, the differences tend to be negative. Finally, for civilians, the differences tend to be zero.*

that includes the actor group as a factor ($BF = 3{,}455$), and post hoc analyses showed that there is moderate evidence in favor of the hypothesis that the differences of ex-paramilitaries are more positive than the differences of ex-guerrillas ($BF = 7{,}001$) and the differences of victims ($BF = 3{,}894$). We also found moderate evidence that the differences between ex-guerrillas and victims are similar ($BF = 0{,}297$).

For the mean eccentricity in the incongruent block, it was found that the model that best explains the data is the one that includes the actor group as a factor ($BF = 16{,}856$), and post hoc analyses showed that there is strong evidence in favor of the hypothesis that ex-paramilitarieséccentricities are more positive than the ones of ex-guerrillas ($BF = 10{,}099$) and victims ($BF = 23{,}325$). Additionally, we found moderate evidence that the diameters of the ex-guerrillas and victims are similar ($BF = 0{,}334$).

The interaction plot (Figure 12(b)) shows that for victims and ex-guerrillas, the mean eccentricities of the congruent block are larger than those of the incongruent one. In contrast, for civilians and ex-paramilitaries, the mean eccentricities of the congruent block are smaller than the incongruent block.

4.3.3 Comparison Between Behavioral and Connectivity Analyses

The correlation analysis between the functional connectivity variables and the IAT test score, shown in Table 13, did not yield promising results. The only variable that shows a weak relationship with the IAT score is the difference in mean diameters between the congruent and incongruent blocks in the theta band.

4.4 Discussion

This chapter outlines our proposed EEG-based functional connectivity analysis approach at the sensor level. Our methodology incorporates REST for EEG re-referencing, minimizing connectivity pattern distortion compared to conventional methods (Chella et al., 2016; Zhang et al., 2020). We also utilize ML-based techniques for detecting noisy channels and artifacts, offering superior reliability over visual or statistical methods (Bigdely-Shamlo et al., 2015; Jas et al., 2017). Again, our approach involves spectrum estimation using multitaper for enhanced frequency resolution (Babadi and Brown, 2014), synchronization measurement via WPLI for greater robustness against noise and biases (Vinck

Tabla 13

Bayesian Pearson correlations

Behavioral variable	Connectivity variables	Pearson's r	BF_{10}
IAT score	- Diameter differences in theta band	**0,274**	**3,928**
	- Eccentricity differences in theta band	0,251	2,240
	- Diameter differences in beta2 band	0,084	0,179
	- Eccentricity differences in beta2 band	0,098	0,200
	- Diameter incongruent block in beta2 band	−0,144	0,327
	- Eccentricity incongruent block in beta2 band	−0,173	0,495

et al., 2011; Bastos and Schoffelen, 2016), and connectivity matrix refinement using MST to overcome the inconsistencies of thresholding methods (Garrison et al., 2015; Zakharov et al., 2021). Furthermore, between-group comparisons were made using Bayesian Inference hypothesis testing methods, as with ERP analyses, which are more reliable than NHST (Keysers et al., 2020; Wagenmakers et al., 2016). Because of the combination of all these elements, we consider that the methodology used represents a significant advance over the methodologies traditionally used for these analyses.

We used this methodology to characterize an IAT synchronized with EEG and designed to measure prejudice among former actors of the Colombian armed conflict. To our knowledge, this is the first study to use this technique to characterize neural activity related to prejudice. It could provide relevant information about how different brain regions are synchronized in time and phase domains when social bias events occur and whether different patterns of neural rhythms can be observed in them.

The behavioral results indicate that civilians are prone to be more prejudiced toward combatants than victims. Ex-guerrillas and ex-paramilitaries appear to exhibit no prejudice

toward victims or themselves, and victims are more inclined to be prejudiced toward their group. Although this seems to go against common sense, many reported cases of prejudice towards the same group exist. For example, March and Graham (2014) showed in a study that Hispanic women are biased toward their group concerning the white women's group; Newheiser et al. (2014), and Gedeon et al. (2021) have reported cases in which children of low economic status show bias against their group in favor of children of high economic status.

In the theta band, the mean diameter and eccentricity of the congruent block of victims and ex-paramilitaries tended to be greater than those of the incongruent block. Greater diameters and eccentricities are indicative of a more efficient network in transmitting information (van Dellen et al., 2018), given that the MST tends to be more centralized (star-like shape) and less linear-like (Stam et al., 2014; Tewarie et al., 2015). Theta rhythms have been associated with executive function and cognitive control (Cavanagh and Frank, 2014). These rhythms have been detected to increase when a surprising or unpleasant stimulus is received (Smit et al., 2023).

Considering that the congruent block of IAT associates combatants with positive valence stimuli and victims with negative valence stimuli, and the incongruent block does the opposite (a more detailed explanation could be seen in Section 2.1.2 and (Baez et al., 2020)), a reasonable explanation for what has been observed is that for ex-paramilitaries and victims, the configuration of the incongruent block task is more opposed to their preconceptions than the configuration of the congruent block task, which would reflect a prejudice towards victims. On the other hand, given that the diameter and eccentricity of the congruent block of civilians and ex-guerrillas tend to be smaller than those of the incongruent block, it could be inferred that they make a greater effort to respond in the congruent block than the incongruent block and that this reflects a prejudice towards combatants.

In the beta2 band, the mean diameter and eccentricity of the congruent block of victims and ex-guerrillas tended to be larger than in the incongruent block, which, following the same reasoning of the previous paragraph would indicate that for these groups, the networks of the congruent block are more efficient than those of the incongruent one. Böttcher et al. (2023) state that beta rhythm is mainly associated with sensorimotor processing. Beste et al. (2023) found that the amplitude of beta oscillations in sensorimotor areas decreases just before and during movement execution. Conversely, an increase in beta amplitude above baseline levels is observed after movement execution, known as

post-movement event-related synchronization –ERS. Beta oscillations tend to vary during movement. Generally, movements decrease beta activity, while successful movement cancellation typically increases beta activity. Therefore, beta ERS is believed to reflect an active inhibition of the motor cortex by somatosensory feedback. For victims and ex-guerrillas, lower efficiency in the incongruent block in this band might reflect greater inhibition to respond correctly to the association among combatants and negative valence stimuli, or victims and positive valence stimuli. In contrast, the lower efficiency in the congruent block to the incongruent block observed in ex-paramilitaries might reflect greater inhibition to respond to the association among ex-combatants and positive valence stimuli, or victims and negative valence stimuli.

In resume:

- Victims generate more prejudice toward themselves and try to regulate this prejudice more, which should cause them to take longer to answer the incongruent block trials than congruent block trials. These effects would lead to positive D-scores, which is consistent with the IAT test results ($CrI = [0{,}076, 0{,}294]$).
- Ex-paramilitaries generate more prejudice toward victims. However, they tend to regulate the prejudice toward combatants more, which would balance out the response times in the congruent and incongruent block trials. These effects would lead to D-scores close to zero, which is consistent with the IAT test results ($CrI = [-0{,}091, 0{,}189]$).
- Ex-guerrillas generate more prejudice towards combatants but tend to regulate the prejudice towards victims. Again, the net result is that the response times in the two blocks are similar, which would lead to IAT scores close to 0, which is indeed the case ($CrI = [-0{,}135, 0{,}072]$).
- Civilians generate more prejudice towards combatants and similarly regulate prejudice towards victims and combatants. These effects would lead to longer response times in congruent block trials than in incongruent block trials, which should produce negative IAT scores. Again, the test result is consistent with this reasoning ($CrI = [-0{,}317, 0{,}076]$).

Given that we have previously conducted socio-cognitive interventions with former actors of the Colombian armed conflict, designed to improve emotional processing (Valencia et al., 2020) and reduce aggressiveness (Trujillo et al., 2017), the results of this research are valuable input for customizing and improving these strategies, always aiming to reduce prejudice and achieve reconciliation between former enemies (Ugarriza et al., 2019).

In the following chapter, we will use the electrophysiological measures obtained through EEG-based functional connectivity, explained in this chapter, along with a set of demographic and behavioral variables, to characterize the groups of former actors of the Colombian armed conflict using an interpretable ML model.

5 Interpretable Machine Learning Model to Characterize Colombian Armed Conflict Actors

In this chapter, we present an interpretable ML (iML) model to characterize our participants. For this purpose, we fit and evaluated nine classification models and selected one using a nested cross-validation process. We incorporated feature selection, preprocessing, and hyperparameter tuning procedures in the inner loop. Then, we used SHAP to perform global (at the group level) and local (at the individual participant level) interpretability analyses. We found that five features out of 128 evaluated are sufficient to discriminate between groups and determine whether a participant should be reclassified.

5.1 Literature Review

In recent years, ML-based methods in psychology have become very popular. By performing a simple search in Scopus, it can be found that 2823 scientific papers have been published between 2019 and October 2023, with a growing trend covering practically all subfields of psychology (see Figure 13).

Typically, psychology researchers are interested in more than just making predictions about a dependent variable. They usually want to know which predictor variables influence the dependent variable and how they interact with each other (Henninger et al., 2023). Although the ML models with the best predictive performance are not directly interpretable, a set of techniques allows one to add a layer of interpretability. Despite this, iML research in psychology remains rare and in the last five years, only seven publications related to this topic have been published in Scopus. One of these is a tutorial explaining how to use iML in psychology (Pargent et al., 2023), and another points out the opportunities that iML models have in psychology, but also the risks of misinterpreting the results due to poor data preparation (Henninger et al., 2023). Two more investigations use iML models to study amnesia: Martínez-Florez et al. (2021) used ensemble and SHAP models to find the most important features to distinguish between healthy cognition and amnestic Mild Cognitive Impairment (aMCI), and Kang et al. (2021) used iML models (unknown which ones) to find predictors of Amyloid-β ($A\beta$) assessment in aMCI. On the other hand, Kim et al. (2023) used ensemble models and the permutation importance method that comes with implementing these to find predictors of suicide in South Korean women, and Lu et al. (2023) used ensemble models and SHAP to study possible predictors of aggres-

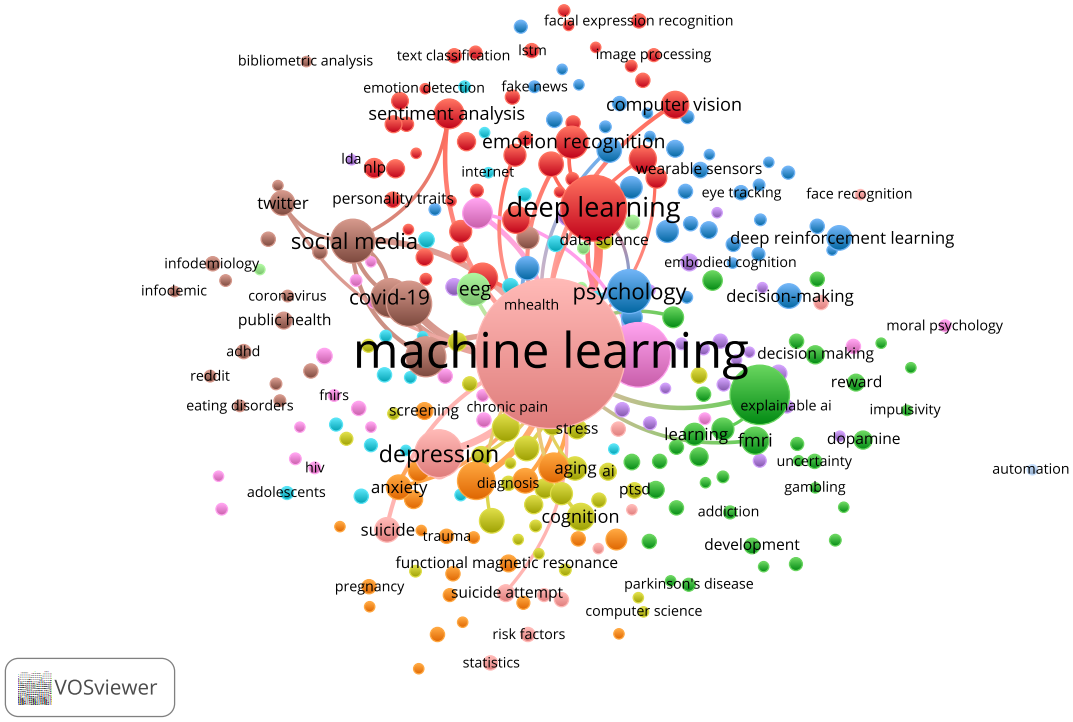


Figura 13

Trending topics on ML for psychology research.

sion in people who undergo drug detoxification treatment in China. Finally, Gyorda et al. (2023) used ensemble and SHAP models to assess whether treatment response to a digital intervention for Generalized Anxiety Disorder (GAD) can be accurately predicted using baseline characteristics.

From this, it can be concluded that the combination of ensemble modeling plus SHAP is preferred for constructing iML. However, to our knowledge, no research has yet taken advantage of SHAP's capabilities to do local analysis and identify subject-by-subject whose characteristics make them outsiders within their group. This kind of analysis can be precious to psychologists because it allows them to decide what to do with these subjects, considering that membership in a particular group often determines their treatment.

5.2 Methodology

The complete workflow of the methodology used is shown in Figure 14. The methodology has three stages: data preparation, best model selection, and best model interpretation. The procedures and techniques used at each stage will be explained below.

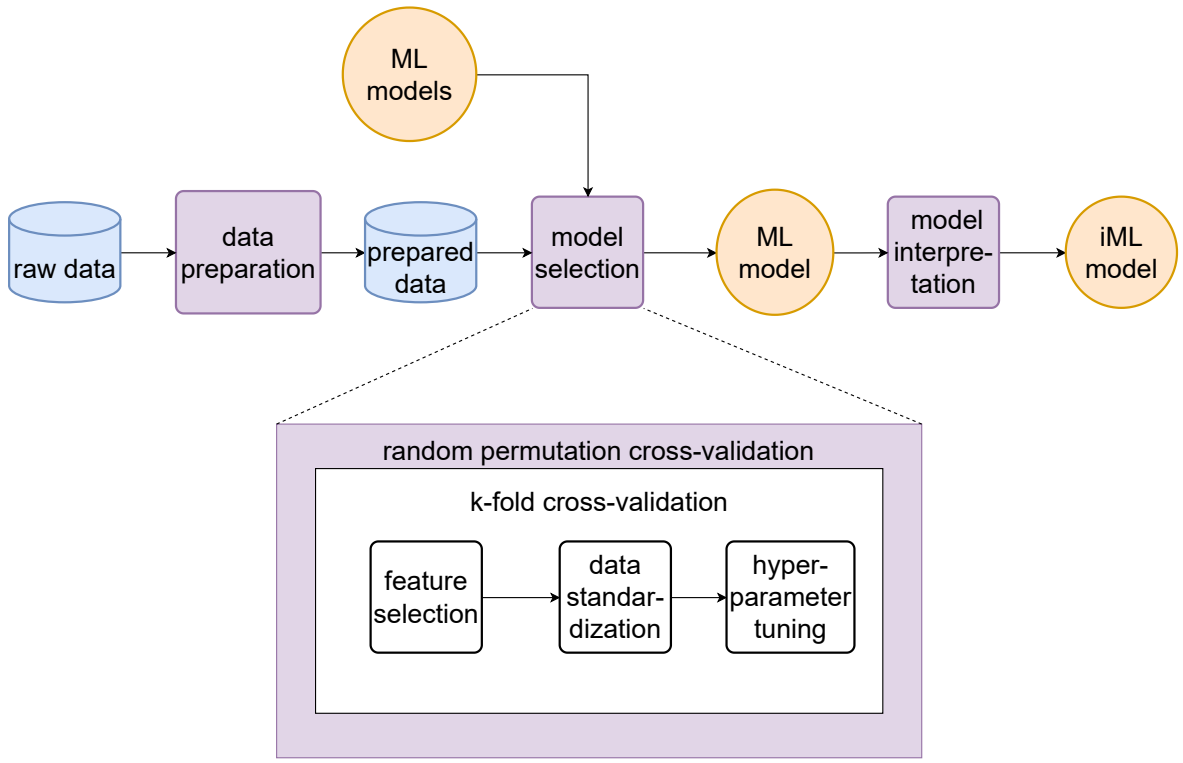


Figura 14

Methodology flowchart. The methodology has three stages: data preparation, best model selection, and best model interpretation.

5.2.1 Data Preparation

We constructed a data set of 128 variables that contains all the functional connectivity measures based on EEG analyzed in Chapter 4, the results of the behavioral tests described in Section 2.1.2, and the demographic variables of the participants. Table 14 shows the list of characteristics that comprised the data set.

Subjects with missing data were eliminated, so the final data set comprised 88 records. Categorical variables were coded using the one-hot technique. One-hot encoding is a technique used to represent categorical variables as binary vectors. For each unique category in the original variable, a new binary (0 or 1) column is created in the transformed dataset, where a "1" denotes the presence of the category and a "0" denotes its absence. This method effectively transforms non-numeric data into a format that machine learning algorithms can understand. However, it can substantially increase the dimensionality of the dataset if the categorical variable has many unique categories (Zheng and Casari, 2018).

Cuadro 14: List of dataset features

Demographical features	Behavioral features		EEG-based functional connectivity features		
	Test	Subscale / subtest	Measure	Band	Block
Actors group	IAT	IAT	Leaf fraction	Delta	Congruent
Age	IRI	PT, FS, EC, PD	Diameter	Theta	Incongruent
Years of schooling	IMA	IMA	Mean eccentricity	Alpha	Differences between congruent and incongruent block
Gender	RPQ	AR, AP	Maximum degree	Beta1	
Laterality	EVEA	AL, TD, IH, AN	Maximum betweenness	Beta2	
	EX2	EX2, self identification as a victim, exposure level	Tree hierarchy	Gamma	
Total features: 5	Total features: 15		Total features: 108		

We worked with unstandardized data for the ensemble ML models because it is considered that these models are insensitive to data standardization (James et al., 2021). For the other ML models, we tested the following three methods:

1. Discrete data were scaled to a [0, 1] range, and continuous data were scaled to zero mean, unit variance (Pedregosa et al., 2011).

2. Discrete data were scaled to a range [0, 1], and continuous data were standardized using the Yeo-Johnson (Yeo and Johnson, 2000) power transform method.

3. The data was not standardized.

5.2.2 Selection of the Best ML Model

We used the Random Permutation cross-validation (CV) technique in the outer loop with ten splits in a ratio of 80/20. Random Permutation CV (often termed "Shuffle & Split") is a model validation technique in which the data set is randomly shuffled and then split into train and test subsets. This procedure can be repeated multiple times, leading to different possible train-test split combinations. It is an alternative to more traditional K-Fold cross-validation. The process can be repeated n times, producing n different train-test splits, ensuring that the model is trained and tested on diverse data combinations, helping to understand the variability in performance (James et al., 2021).

We utilized the k-Fold CV technique on the inner loop with five folds in a 80/20 ratio. k-Fold CV is a popular model validation technique used to assess the performance of machine learning algorithms. In this method:

- The dataset is randomly partitioned into k equally (or nearly equally) sized subsamples or "folds."
- Of the k folds, $k-1$ folds are used for training the model, and the remaining single fold is used as a validation set to assess performance.
- This process is repeated k times, with each fold used exactly once as a validation set.
- The k results of the folds can then be averaged (or combined otherwise) to produce a single estimation of the model performance.

k-Fold CV is a powerful and widely used technique, offering a more robust estimation of model performance than a single train-test split. It is especially valuable when the available data is limited, and you want the most comprehensive assessment of the potential performance of a model (James et al., 2021).

To avoid data leakage (Lones, 2021), we build a pipeline consisting of feature selection, data standardization, and hyperparameter tuning processes in each fold.

We selected the $F1_{weighted}$ score to select the best ML model and tune the hyperparameters. This metric calculates the $F1$ score for each class individually and then takes a weighted average of those scores, considering the number of true instances for each label. The $F1$ score is the harmonic mean of precision and recall, with a higher score signifying a better balance between the two (Sokolova and Lapalme, 2009). Given precision P and recall R, the $F1$ score is:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{7}$$

Where:

$$P = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{8}$$

and

$$R = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{9}$$

For a multi-class problem, the $F1_{weighted}$ score is:

$$F1_{weighted} = \sum_{i=1}^{C} w_i \times F1_i \tag{10}$$

where C is the number of classes, $F1_i$ is the $F1$ score for class i, and w_i is the weight for class i.

The $F1$ score offers several advantages over other metrics. Unlike accuracy, which can be misleading, the $F1$ score considers false positives and negatives, providing a more accurate evaluation of the model performance. Moreover, the $F1_{weighted}$ score is helpful in multiclass scenarios, as it considers class distribution, ensuring frequent ones do not overshadow performance in underrepresented classes. This attribute is particularly beneficial when classes are imbalanced, as it prevents a model from being considered "good" based solely on its ability to predict the majority class (Sokolova and Lapalme, 2009).

The tuning of hyperparameters was done using a tool based on Bayesian optimization, which is computationally more efficient and less time-consuming than brute force methods, such as random, grid, and sequential search(Sandha et al., 2020). In summary, Bayesian optimization fits a probabilistic model to capture the relationship between hyperparameter settings and their measured performance; it then uses this model to select

the most promising hyperparameter setting (trading off the exploration of new parts of the space vs. exploitation in known promising regions), evaluates that hyperparameter setting, and updates the model with the result and iterates (Feurer et al., 2015).

5.2.3 Feature Selection

To select the most relevant features, we followed a two-step process. First, we used the techniques described in Section 2.2.5 and retained the features selected by at least three of these methods. Later, we performed a hierarchical clustering based on Ward's method (Ward Jr, 1963) to control multicollinear features. It is crucial in our study because when multicollinear features exist, it can be difficult to determine each feature's contribution to the model's performance, and the model may lose interpretability.

Ward's method (or Ward's minimum-variance method) is a criterion applied in hierarchical clustering. It aims to minimize the variance within clusters. When clustering with this method, the pair of clusters that leads to the minimum increase in total variance within the cluster after merging will be merged at each step (Ward Jr, 1963). The metric of this method is Ward's distance (or the increase in the error sum of squares). The following formula shows that this metric is calculated between two clusters, A and B:

$$D(A,B) = \frac{n_A n_B}{n_A + n_B} d^2\left(C_A, C_B\right) \tag{11}$$

Where:

n_A and n_B are the numbers of observations in the clusters A and B, respectively.

$d^2\left(C_A, C_B\right)$ is the squared Euclidean distance between the centroids C_A and C_B of the clusters A and B.

With this method, we selected non-collinear features as follows:

- We calculated the absolute Spearman rank-order correlations of the features chosen in the first step.
- We measured the Ward distance among the features of the Spearman correlations.

- We build a dendrogram, a tree-like diagram that displays the sequence and results of hierarchical clustering. Visualizes the arrangement of clusters produced by the corresponding merges and splits, allowing an observer to reconstruct the history of those merges or splits (Murtagh and Legendre, 2014).
- We set a threshold of 0,75 to prune the dendrogram. We selected this value because it filtered out the highly correlated features, and the model performance did not degrade appreciably.
- We preserve only one feature for each survival branch.

5.2.4 Interpretation of the ML Model Selected

Since the best models were tree-based ML models, we used TreeSHAP, a variant of SHAP specialized in these models (Lundberg et al., 2018). The explication of TreeSHAP is published online (Molnar, 2022). This algorithm is faster than other SHAP algorithms but only works with tree-based ML models.

We used summary graphs for global analyses, while for local analyses, we used force plots (Molnar, 2022) from the `SHAP` package (version 0.41.0).

The summary plot combines feature importance with feature effects. Each point on the summary plot is a Shapley value for a feature and an instance. The feature importance determines the position on the y-axis; on the x-axis, it is determined by the Shapley value. The color represents the value of the feature from low to high. Overlapping points are jittered in the y-axis direction, so we understand the distribution of the Shapley values per feature. Features are ordered according to their importance (Molnar, 2022).

A SHAP force plot visualizes the contribution of each feature to a particular prediction, making it easier to understand which features push the prediction in one direction or another. The elements of a force plot are (a) Base Value: This is the starting point of the plot and represents the model's average prediction for all instances. In other words, if it had no specific information about the current instance and relied only on the average, this is the prediction it would get. (b) Features: Each feature that affects the prediction is shown as a force that pushes the prediction to increase (positive) or decrease (negative). The size and direction of the arrow/line for each feature represent the magnitude and direction of its impact on the prediction. (c) Output value: The end of the plot shows the final prediction

for observation, considering all the contributions of features (Molnar, 2022).

For one observation, the force plot begins by showing the base value of the prediction, based on the relative frequency of the class. After that, it considers each feature separately and adds its effects to the base value one by one. Features that increase the prediction are shown in red, whereas those that tend to decrease it are displayed in blue. Combining all these effects will arrive at the final prediction value.

5.3 Results

5.3.1 Selection of the Best ML Model and Feature Selection

A total of 12 models were tested: three ensemble models with unstandardized data, i.e., Random Forest, Gradient Boosting, XG Boosting, and the three other models, i.e., Logistic Regression, SVM, kNN, with the three different standardization procedures described in Section 5.2.1. Table 15 shows the cross-validation and test scores of all trained models sorted by test score. The ensemble models make up most of the best models, and the difference between the first three places is minor. Therefore, any of these models can be selected. We chose the random forest model, with hyperparameters ccp_alpha =1.01×10^{-4} and $n\_estimators = 88$.

We obtained the 25 pre-selected features shown in Figure 15 for the model selected. After that, we obtained 16 selected features with the set threshold. Selected features are listed in Table 16.

5.3.2 Interpretation of Selected ML Model

Global analyses

We made global analyses with each group's summary plots, shown in Figure 16. Next, we performed NHST and chi-squared tests on each group's most important features to understand why they are critical for the classifier's performance on prediction.

The summary plot of civilians (Figure 16a) shows that the two most important features to predict whether a subject in the sample is a civilian are the EX2_score and victims_self_no. For the model, a subject with a low EX2_score and who self-identifies as a

Tabla 15

Scores of classification models with selected features

Model	F1 weighted CV score M(SD) %	F1 weighted test score M(SD) %
Random Forest	80.4(3.8)	86.7(6.5)
XGBoost	83.9(3.4)	85(9.1)
Gradient Boosting	82.5(3.5)	84.4(9.7)
SVC NP	77.9(3)	83.3(6.4)
SVC PT	81.5(2.7)	82.8(8.9)
SVC SC	79.1(5.2)	82.8(10.0)
Logistic Regression NP	73.7(3.2)	80(9.1)
KNN PT	74.1(4.4)	75.6(7.0)
KNN NP	74.3(4.6)	74.4(9.5)
KNN SC	74.4(4.2)	72.8(8.5)
Logistic Regression PT	45.1(0.0)	61.1(0.0)
Logistic Regression SC	45.1(0.0)	61.1(0.0)

NP: Not standardization. PT: Standardization with Power Transformer. SC: Standardization with Standard Scaler

Tabla 16

Selected features by best ML model

Demographical features	Behavioral features	Functional connectivity features
1. Age (age) 2. Years of schooling (school_years) 3. Female gender (gender_F)	1. EX2 test score (EX2_score) 2. Perspective-taking subscale score of the IRI test (IRI_PT) 3. Empathy concern subscale score of the IRI test (IRI_EC) 4. Personal upset subscale score of the IRI test (IRI_PD) 5. Reactive Aggression subscale score of the RPQ test (RPQ_AR) 6. EVEA Joy subscale score on the EVEA test (AL) 7. EVEA Anxiety subscale score on the EVEA test (AN) 8. Subject self-identifies as a non-victim (victims_self_no)	1. Mean eccentricity in beta2 band incongruent trials (mean_eccentricity_b2i) 2. Mean eccentricity in gamma band congruent trials (mean_eccentricity_gc) 3. Mean eccentricity in theta band congruent trials (mean_eccentricity_tc) 4. Maximum degree in beta2 band incongruent trials (max_degree_b2i) 5. Maximum degree in beta2 band congruent trials (max_degree_b2c)

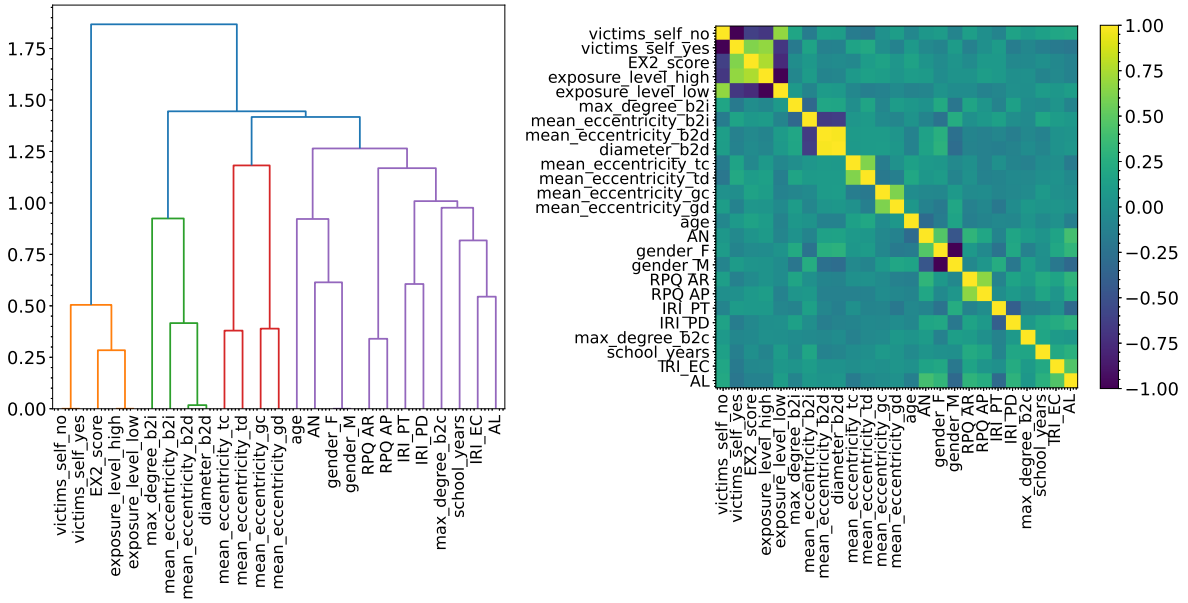


Figura 15

Description of the second step of the feature selection procedure. After having made an initial feature selection, the second step preserves those features that are less correlated with each other, trying not to degrade the performance of the model substantially. The selection was made by performing hierarchical clustering on the Spearman rank-order correlations, shown on the right side of the Figure, using the Ward distance as a metric, obtaining the dendrogram of the left side of the Figure. Next, we picked a threshold of 0.75. Finally, we chose a feature to keep from each cluster.

non-victim is a civilian, which makes perfect sense because, by definition, in this sample, a civilian is someone who has had low exposure to the armed conflict. AL and gender_F are also important features, although less discriminating. Thus, low values in the EX2_score increase the probability that a subject is classified as a civilian by up to ,3. However, given that the base probability of being a civilian is ,159 (14 civilians in a sample of 88 subjects), this may not be sufficient, so having "correct" values (with "correct" values being understood as those that increase the probability of a subject being classified as a civilian) of the following three features (victims_self_no, AL, and gender_F) are also necessary. From this, it can be stated that a civilian is characterized by having a low EX2_score, self-identifying as a non-victim, having a high AL score, and being male.

The summary plot of ex-combatants (Figure 16b) shows that the three most important features are AL, EX2_score, and gender_F. Thus, an ex-combatant is characterized by

having a low score on the AL task, which assesses joy, a score above two on the EX2_scale, indicating that he has had high exposure to violent events, and by being male. In this case, however, given that the probability that the model predicts ex-combatant is a priori very high (53/88 = ,602), which makes it the default prediction, it is more valuable for us to identify which features contribute most strongly to lowering this probability. The summary plot of this group allows us to observe that these features are EX2_score (low scores lower the probability of predicting ex-combatant by up to ,25), AL (low scores lower the probability by up to ,19), gender (being female lowers the probability by up to ,18), IRI_PT (low scores lower the probability by up to ,15), and mean_eccentricity_b2i (low mean eccentricity lowers the probability by up to ,15).

The summary plot of the victims (Figure 16c) shows that the most critical features are gender_F (being female raises the probability by up to ,15, and being male lowers it by up to ,11), AL (a high score raises the probability by up to ,16 and a low score lowers it by up to ,13), the IRI_PT score (a low score raises the probability by up to ,14, and a low score lowers it by up to ,09), and the mean_eccentricity_b2i (a low mean eccentricity raises the probability by up to ,17, and a high mean eccentricity lowers the probability by up to ,09). Thus, according to the ML model, a victim is characterized by being female, having a high AL score, a low IRI_PT score, and a low mean eccentricity in the beta2 band incongruent block.

Finally, to understand why the model identified as the most important features, we performed NHST and chi-square tests on the features EX2_score, victim_self_no, AL, gender_F, IRI_PT, and mean_eccentricity_b2i. The NHST analysis applied to the AL feature revealed that there are significant differences between actors groups ($p - value <$,001, $n_p^2 = 0{,}275$), and posthoc analyses revealed significant differences among ex-combatants and victims ($p - value <$,001, $CI = [-6{,}243, -2{,}402]$) and among ex-combatants and civilians ($p - value = 0{,}007$, $CI = [-5{,}144, -0{,}668]$). A similar analysis performed on the EX2_score feature shows that there are significant differences among groups of actors ($p - value <$,001, $n_p^2 = 0{,}304$), more specifically among civilians and ex-combatants ($p - value <$,001, $CI = [-7{,}036, -3{,}070]$) and among civilians and victims ($p - value <$,001, $CI = [-6{,}491, -1{,}937]$). Regarding the IRI_PT feature, it was found that there are significant differences among groups of actors ($p - value <$,001, $n_p^2 = 0{,}242$), more specifically among ex-combatants and victims ($p - value <$,001, $CI = [3{,}080, 8{,}397]$). Concerning the mean_eccentricity_b2i feature, it was found that there are significant differences among groups of actors ($p - value =$,017, $n_p^2 = 0{,}091$), more specifically among ex-combatants and victims ($p - value =$,014, $CI = [0{,}269, 2{,}887]$). Last, the chi-squared tests proved that

there are significant differences between actors in terms of gender ($p - value <$,001) and whether they identify themselves as victims or not ($p - value <$,001).

Local Analyses

We used force graphs to perform local analyses, as shown in Figure 17. Since we found only one error with the chosen model in the test set (one participant was misclassified as a civilian when, in fact, she was an ex-combatant), we applied the same analysis to subjects who, although correctly classified, had a high probability of being misclassified.

In the case of the subject who was misclassified, when analyzing the most important features when making this prediction, we found the EX2_score, victim_self_no, and AL score (refer to Figure 17a). Upon further inspection, we found that this participant had a very low EX2 score (1), which is more typical of civilians (with a mean score of 0,929 and standard deviation of 1,592), rather than ex-combatants (with a mean score of 5,981 and standard deviation of 2,866). Additionally, the participant identified herself as a non-victim, whereas the vast majority (81,1 %) of ex-combatants identify as victims, while 92,9 % of civilians do not. Furthermore, the participant's AL score was relatively high (8) for an ex-combatant (with a mean score of 3,844 and standard deviation of 3,611) and more in line with the scores of civilians (with a mean score of 6,750 and standard deviation of 2,600). In particular, the participant's gender (female) was also contrary to the prediction since 73 % of the civilians are male.

Among the subjects classified correctly, one was classified as an ex-combatant with a probability of 0,443, while the probability of being misclassified as a victim was 0,386, two relatively close values. The local interpretability analysis (see Figure 17b) shows that the features that most positively influenced this prediction were the IRI_PT score (18), the RPQ_AR score (13), and the mean eccentricity in beta2 band incongruent trials (14,82), which makes sense because these three values are within the typical range of ex-combatants (IRI_PT: $M = 18{,}453, SD = 4{,}177$, RPQ_AR: $M = 16{,}321, SD = 4{,}677$, mean_eccentricity_b2i: $M = 14{,}420, SD = 2{,}142$), while the features most negatively influencing prediction were AL score (8,25), atypical for ex-combatants ($M = 3{,}844, SD = 3{,}611$) and more usual for victims ($M = 7{,}837, SD = 2{,}403$), the gender (female), which is atypical for ex-combatants (20,8 % are female) and typical for victims (91,3 % are female), and the self-identification as a victim (no), which is atypical for ex-combatants (only 18,86 % of them self-identifies as a non-victim) and typical for civilians (92,85 % of them self-identifies as a non-victim).

5.4 Discussion

In this chapter, we presented an iML model to characterize a sample of subjects who were immersed in the Colombian armed conflict, either as combatants, victims, or civilian non-ex-combatants and non-victims, based on a set of demographic, behavioral, and electrophysiological variables. For this purpose, 12 different ML models were evaluated. The model selection was done with a nested CV process, using the $F1_{weighted}$ score as the evaluation metric. Hyperparameter tuning was performed by CV using Bayesian optimization techniques, incorporating feature selection, standardization, and tuning into the pipeline to avoid data leakage. The interpretability of the selected model was done using SHAP, mainly their summary plots and force plots.

Although studies with global analyses of iML models already exist in psychology, to our knowledge, this is the first study in psychology that performs a local interpretability analysis to identify the features that led the model to classify some subjects differently from the original class.

The results revealed that the ensemble models (random forests, gradient boost, and XG boost) perform better than other models (SVM, kNN, logistic regression) and that the standardization technique has little effect on the result. Four behavioral variables (EX2, EVEA-AL, self-identification as victim, and IRI-PT) and one demographic (genre) were also identified as the essential features of the characterization. The first electrophysiological characteristic identified as necessary was the mean eccentricity in the beta2 band incongruent block (mean_eccentricity_b2i), below the mentioned characteristics. This information is valuable for future research with the same population because it gives us insights into which variables are worth capturing and which are not, which, in the long run, may result in resource savings.

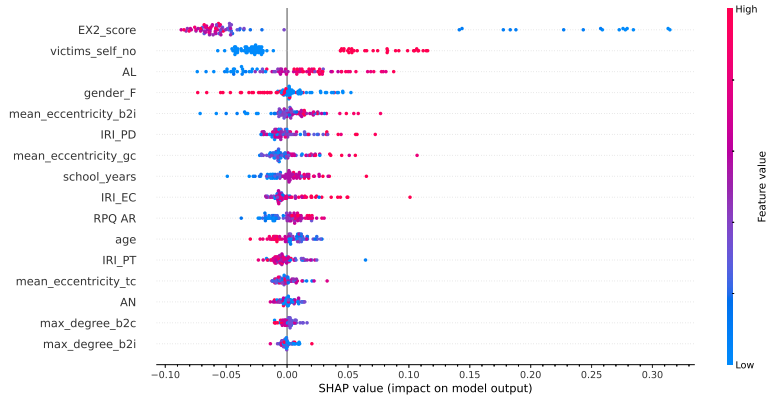
According to these results, it is possible to affirm that ex-combatants are typically characterized by having an AL score lower than 7, an EX2 score higher than 4, being male, having an IRI-PT score higher than 16, and recognizing themselves as victims. A victim is typically characterized as female, having an AL score greater than 7 and an IRI-PT score less than 15. Finally, a civilian is typically characterized as having an EX2 score equal to or below 2 and not recognizing himself as a victim.

Previous work with armed conflict actors has not been focused on their characterization or finding features that would allow discriminating ex-combatants from victims

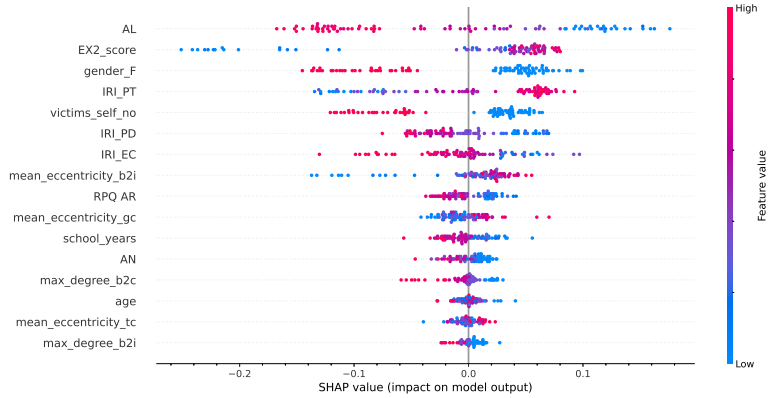
and civilians, but on estimating exposure to the conflict (Cano et al., 2022) and on identifying emotional processing profiles in ex-combatants (Quintero-Zea et al., 2018). Thus, this study is the first to fully characterize the various groups of actors in the Colombian armed conflict.

These findings are crucial in designing personalized training programs to mitigate prejudice among Colombian armed conflict actors. The ability to reclassify subjects based on these thresholds is essential, as it ensures that individuals receive the most appropriate and effective interventions aligned with their group categorization. This research has practical implications for developing tailored interventions in social-cognitive studies.

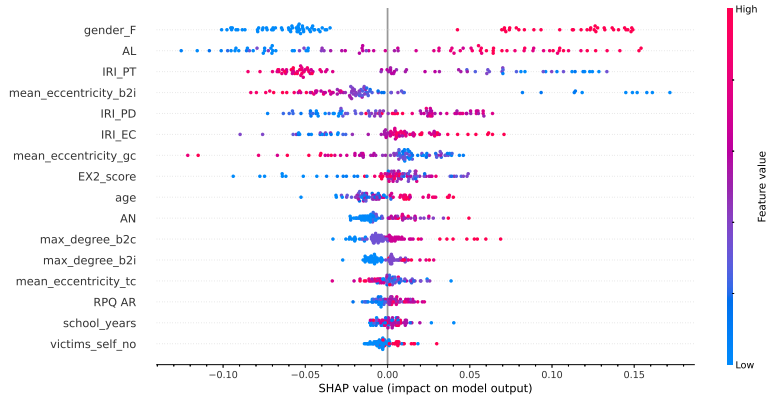
The proposed methodology could be used in psychological research where it is necessary to characterize samples of subjects with a vast number of variables or where these variables have non-linear relationships. Furthermore, the methodology used to perform local interpretability analysis could be used to assess whether a participant might be reclassified based on its scores in the set of most important features of the iML model.



(a) *Shapley values of selected features, civilians group*



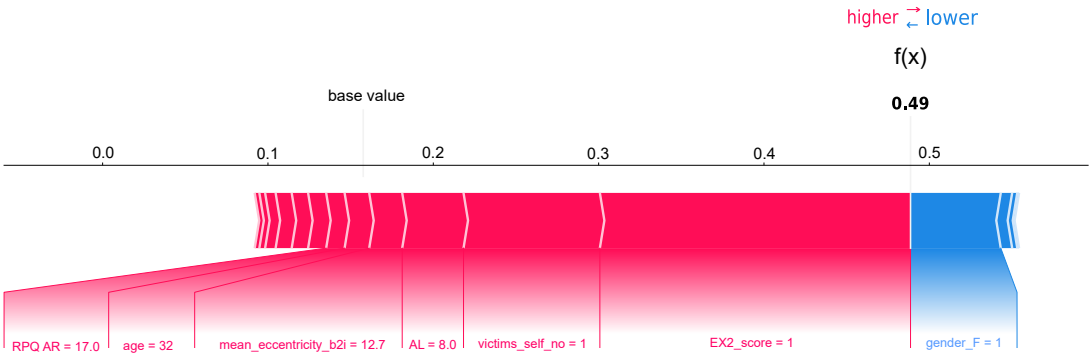
(b) *Shapley values of selected features, ex-combatants group*



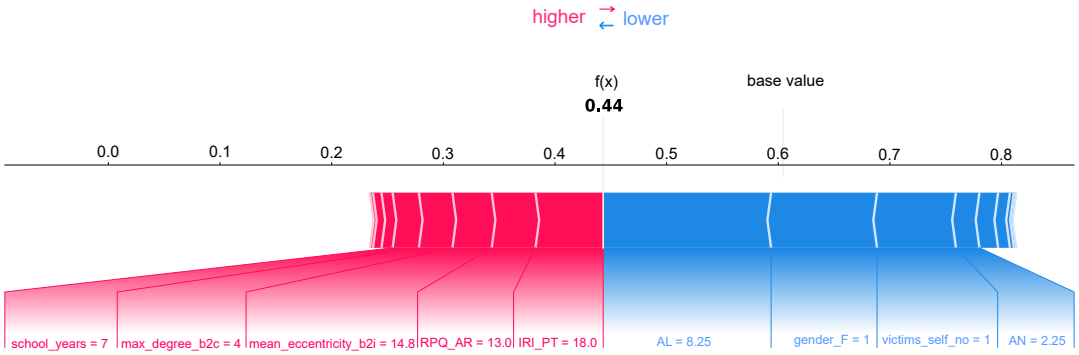
(c) *Shapley values of selected features, victims group*

Figura 16

Summary plots of SHAP analyses by groups. Summary plots show how each feature affects the prediction and its importance. Each dot on the plot is the effect of a feature for each observation. The importance of the feature determines the position on the y-axis, the most important features are at the top, and the size of the effect determines the position on the x-axis. The color shows the value of the feature from low (blue) to high (red) (Molnar, 2022). For example, in the plot for civilians (subfigure (a)), we can see that EX2_score is the most important feature and has a negative relationship with the prediction: lower scores increase the prediction, and higher scores decrease it for this group.



(a) *Features'contribution to the prediction of subject 21137*



(b) *Features'contribution to the prediction of subject 21114*

Figura 17

Force plots of two participants. These plots show how each feature value changes the prediction from the average. The average prediction for all data is the baseline. The plot has arrows for each feature value that move the prediction up (positive) or down (negative) from the baseline. These forces balance each other out at the actual prediction of the data instance (Molnar, 2022).

6 Concluding Remarks

This thesis established that it is feasible to characterize a cohort of former combatants in the Colombian armed conflict based on the demographic variables, EEG recordings, and behavioral tests they completed. To this end, the following tasks were performed:

1. A novel methodology proposed for ERP analyses that utilizes Bayesian inference to test hypotheses and massive univariate statistical methods.
2. A proposed methodology for an EEG-based functional connectivity analysis at the sensor level that aggregates current best practices.
3. A deep examination, on a global and local scale, of the significance of features in an iML model that characterizes our population.

The following section presents some conclusions from these analyses and recommendations for future work.

6.1 ERP Analyses

We introduced a novel and robust methodology for analyzing ERPs by combining massive univariate statistical permutation and Bayesian inference techniques for hypothesis testing. The advantages of this approach over traditional methods are twofold: first, it provides a significantly enhanced precision in localizing ERP components, allowing for more accurate identification of specific effects. Second, by addressing multiple comparison issues, our methodology can pinpoint the exact effects that traditional approaches may struggle to identify. Furthermore, our approach surpasses data-driven methods grounded in massive univariate analysis, demonstrating the flexibility of Bayesian methods in establishing thresholds without compromising statistical rigor. This adaptability results in a more precise detection of effects.

In addition, the study findings shed light on the relationship between prejudice and ERP components. Participants with prejudice toward victims showed larger incongruent block amplitudes in the components P3, N400, and LPP compared to individuals without prejudice or those exhibiting prejudice toward ex-combatants.

The results imply that people with prejudice towards victims allocate more cognitive resources to process incongruent block stimuli, potentially leading to longer reaction times, which is consistent with behavioral outcomes.

This research underscores the importance of understanding the neural correlates of prejudice, contributing valuable insights to cognitive neuroscience and social psychology.

6.2 Functional Connectivity Analyses

To fulfill the second objective, we introduced a cutting-edge methodology that uses state-of-the-art engineering techniques such as REST for EEG re-referencing, machine learning-based techniques to detect noisy channels and artifacts, multitaper for spectrum estimation, WPLI as a measure for synchronization, MST for pruning the connectivity matrix, and Bayesian Inference Hypothesis Testing Methods for statistical analyses. The integration of these elements represents a notable advancement in the methodologies commonly employed in these analyses.

The functional connectivity analysis indicates that civilians and ex-guerrillas tend to generate more prejudice toward armed actors than toward victims and that victims and ex-paramilitaries tend to generate more prejudice toward victims. On the other hand, ex-guerrillas and victims tend to regulate their prejudice toward victims better. Similarly, the ex-paramilitaries seem to regulate their prejudice toward combatants better. Finally, it seems that civilians regulate prejudice toward victims and combatants similarly. All of these results are highly consistent with those of the behavioral test.

The results suggest that the theta-band MST diameter and eccentricity measures are associated with the activity of the prejudice generation network.On the other hand,, these same measures in the beta2 band are associated with the activity of the prejudice regulation network.

6.3 iML Analyses

We developed a novel methodology centered on ML models'global and local interpretability analyses to identify the most critical features in characterizing distinct groups and assessing the feasibility of individual subject reclassification. Although our methodo-

logy was applied to a specific task, it could be used in broader research fields to characterize samples of subjects with many variables or where these variables have non-linear relationships. Furthermore, this methodology could be used to identify features at a high degree of granularity that determine whether an ML model incorrectly predicts a particular subject.

Through our research, we have successfully identified four behavioral variables and one demographic variable as critical elements in characterizing groups of actors within the context of the Colombian armed conflict. This finding holds substantial value for future research within the same population, as it provides essential information on which variables are worthy of capture and which can be omitted, ultimately leading to long-term resource savings.

Also, we determined the threshold values for this small set of critical features that are pivotal in defining a subject's membership within a specific group. The obtained results will allow the customization of training/treatment interventions to the profile of the subject to whom they will be applied, ensuring that individuals receive the most appropriate and effective interventions aligned with their group categorization.

This analysis is crucial in designing personalized social-cognitive training programs to mitigate prejudice among Colombian armed conflict actors. This research advances our understanding of these actor groups'characteristics and has practical implications for developing tailored social-cognitive interventions.

6.4 Future Work

Following the research line described in this thesis, five main lines of projects could be taken up:

a) The findings from our ERP analysis indicate that our proposed methodology exhibits greater sensitivity than traditional approaches, enabling the detection of a higher number of true effects within the same dataset, all while maintaining specificity. This gain is by the Bayesian analysis's ability to effectively filter out non-existent effects, as exemplified by the results across various clusters. Future research should focus on evaluating the methodology under more controlled and rigorous conditions to validate and substantiate these promising outcomes, using, by example, synthetic data. These

controlled experiments will be instrumental in confirming the enhanced sensitivity and specificity of the methodology, ensuring its reliability and applicability in diverse research scenarios.

b) Future research endeavors should consider the exploration of potential electrophysiological correlates of prejudice through measures of EEG-based functional connectivity at the sensor level, like diameter and eccentricity in the theta and beta2 frequency bands. These measures may hold valuable insights into the neural underpinnings of prejudice.

c) An exciting avenue for future research lies in extending functional connectivity analysis to the source level of EEG data. Integrating the findings of source-level connectivity with the previous results of the EEG-ERP analyses, sensor-level functional connectivity, and functional connectivity derived from fMRI would be a significant step forward. This comprehensive approach would offer a more holistic and nuanced understanding of the neurological basis of prejudice, providing a richer perspective on the complex neural dynamics involved. Such research efforts have the potential to shed light on the intricate interplay of neural networks and cognitive processes underlying prejudice, advancing the field of neuroscience and social psychology.

d) For future investigations, a prospective direction involves exploring alternative ML techniques, particularly clustering methods, to group subjects and determining the most suitable method for specific tasks within neuropsychology requiring classification or reclassification. Comparing and evaluating different ML techniques would provide information on their efficacy in characterizing subject groups, potentially enhancing the precision and accuracy of such classifications.

e) Last, a focus area for subsequent research within the research group entails integrating measures of established ERP components to ascertain their importance as characteristics in characterizing groups of individuals who were former actors in the Colombian armed conflict. This approach seeks to determine well-known ERP components'relevance and potential contributions in distinguishing and characterizing these groups. Such an investigation holds promise in uncovering nuanced cognitive and neural markers associated with this population, further enriching our understanding of their psychological and neurological profiles.

6.5 Final Conclusions

The results of the first specific objective show that subjects with prejudice against victims have larger ERPs than subjects who have either prejudice against combatants or have no marked prejudice against either side. On the other hand, the results of the IAT test show that, paradoxically, victims are the group that has more prejudice towards victims. In comparison, civilians have more prejudice towards combatants, and former members of guerrillas and paramilitary groups do not have a marked prejudice towards either victims or combatants. Additionally, the results of the second objective are consistent with the behavioral results and allow us to know that the cognitive processes that produce prejudice – generation and regulation – occur differently in each of the groups of actors, which justifies from the electrophysiological point of view the design of socio-cognitive interventions differentiated by groups of former actors of the armed conflict to reduce prejudice among them. Finally, the results of the third objective provide tools to determine the best intervention strategy subject by subject, beyond their original membership to a former Colombian armed conflict actor group. Additionally, these results allow us to concentrate future efforts on collecting a relatively small set of characteristics of the subjects, avoiding unnecessary efforts to collect others that are not so significant.

All of the above allows us to affirm that the new methodologies developed in this thesis provide a comprehensive view of the phenomenon of prejudice among former actors of the Colombian armed conflict, which would be difficult to achieve by applying only the classical methodologies of ERP analysis and classical multivariate statistical analysis. For example, with the latter, it would have been more challenging to recognize the differences in the generation and regulation processes of prejudice between groups or to identify which characteristics should be considered to decide what type of intervention should be applied to an individual.

In this way, this thesis achieves the objective of characterizing the phenomenon of prejudice among former armed conflict actors from an electrophysiological point of view. However, it goes further because it adds behavioral and demographic variables to the characterization. The integral characterization of prejudice provides researchers with valuable information to improve intervention strategies designed to mitigate prejudice, aiming at the final purpose of achieving reconciliation and rebuilding the social fabric in former zones of armed conflict in Colombia.

Referencias

Amodio, D. M., Bartholow, B. D., and Ito, T. A. (2014). Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. Social Cognitive and Affective Neuroscience, 9(3):385–393.

Amodio, D. M. and Cikara, M. (2021). The Social Neuroscience of Prejudice. Annual Review of Psychology, 72:439–469.

Arthur, D. and Vassilvitskii, S. (2007). K-means++ the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM Symposium on Discrete algorithms, pages 1027–1035.

Babadi, B. and Brown, E. N. (2014). A review of multitaper spectral analysis. IEEE Transactions on Biomedical Engineering, 61(5):1555–1564.

Baez, S., Trujillo, N., Hurtado, E., Ortiz-Ayala, A., Calvache, M. R., Quishpe, R. C., and Ibanez, A. (2020). The Dynamics of Implicit Intergroup Biases of Victims and Ex-combatants in Post-conflict Scenarios. Journal of Interpersonal Violence, pages NP9295–NP9319.

Bankwitz, A., Rüesch, A., Adank, A., Hörmann, C., Villar de Araujo, T., Schoretsanitis, G., Kleim, B., and Olbrich, S. (2023). EEG source functional connectivity in patients after a recent suicide attempt. Clinical Neurophysiology, 154:60–69.

Bar-Siman-Tov, Y. (2004). From conflict resolution to reconciliation. Oxford University Press.

Barnes-Holmes, D., Staunton, C., Barnes-Holmes, Y., Whelan, R., Stewart, I., Commins, S., Walsh, D., Smeets, P. M., and Dymond, S. (2004). Interfacing relational frame theory with cognitive neuroscience: Semantic priming, the implicit association test, and event related potentials. International Journal of Psychology and Psychological Therapy, 4(2):215–240.

Bastos, A. M. and Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls.

Beste, C., Münchau, A., and Frings, C. (2023). Towards a systematization of brain oscillatory activity in actions. Communications Biology, 6(1):137.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Frontiers in Neuroinformatics, 9(JUNE):1–19.

Blomsma, N., de Rooy, B., Gerritse, F., van der Spek, R., Tewarie, P., Hillebrand, A., Otte, W. M., Stam, C. J., and van Dellen, E. (2022). Minimum spanning tree analysis of brain networks: A systematic review of network size effects, sensitivity for neuropsychiatric pathology, and disorder specificity. Network Neuroscience, 6(2):301–319.

Boersma, M., Smit, D. J. A., Boomsma, D. I., De Geus, E. J. C., Delemarre-van de Waal, H. A., and Stam, C. J. (2013). Growing trees in child brains: graph theoretical analysis of electroencephalography-derived minimum spanning tree in 5-and 7-year-old children reflects brain maturation. Brain connectivity, 3(1):50–60.

Boschi, A., Brofiga, M., and Massobrio, P. (2021). Thresholding Functional Connectivity Matrices to Recover the Topological Properties of Large-Scale Neuronal Networks. Frontiers in Neuroscience, 15:1013.

Böttcher, A., Wilken, S., Adelhöfer, N., Raab, M., Hoffmann, S., and Beste, C. (2023). A dissociable functional relevance of theta- and beta-band activities during complex sensorimotor integration. Cerebral Cortex, 33(14):9154–9164.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3):186–198.

Cao, R., Hao, Y., Wang, X., Gao, Y., Shi, H., Huo, S., Wang, B., Guo, H., and Xiang, J. (2020). EEG Functional Connectivity Underlying Emotional Valance and Arousal Using Minimum Spanning Trees. Frontiers in Neuroscience, 14:355.

Cavanagh, J. F. and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences, 18(8):414–421.

Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. Computers and Electrical Engineering, 40(1):16–28.

Chella, F., Pizzella, V., Zappasodi, F., and Marzetti, L. (2016). Impact of the reference choice on scalp EEG connectivity estimation. Journal of Neural Engineering, 13(3):036016.

Chen, L., Zhou, H., Gu, Y., Wang, S., Wang, J., Tian, L., Zhu, H., and Zhou, Z. (2018). The Neural Correlates of Implicit Cognitive Bias Toward Internet-Related Cues in Internet Addiction: An ERP Study. Frontiers in Psychiatry, 9(SEP).

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794.

Comisión de la Verdad (2022). Cifras de la Comisión de la Verdad presentadas junto con el Informe Final.

Cui, X., Xiang, J., Guo, H., Yin, G., Zhang, H., Lan, F., and Chen, J. (2018). Classification of Alzheimer's Disease, Mild Cognitive Impairment, and Normal Controls With Subnetwork Selection and Graph Kernel Principal Component Analysis Based on Minimum Spanning Tree Brain Functional Network .

Cumming, G. (2014). The New Statistics: Why and How. Psychological Science, 25(1):7–29.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. Catalog of Selected Documents in Psychology,, 10(85):1—-17.

de Colombia., C. d. l. R. (2005). Ley 975 de 2005, por la cual se dictan disposiciones para la reincorporación de miembros de grupos armados organizados al margen de la ley, que contribuyan de manera efectiva a la consecución de la paz nacional y se dictan otras disposiciones para acuerdos humanitarios.

Dehghan, P., Alashwal, H., and Moustafa, A. A. (2022). Applications of machine learning to behavioral sciences: focus on categorical data. Discover Psychology, 2(1):22.

Dobrin, R. and Duxbury, P. M. (2001). Minimum spanning trees on random networks. Physical Review Letters, 86(22):5076.

Dovidio, J. F., Hewstone, M., Glick, P., and Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. Prejudice, stereotyping and discrimination, 12:3–28.

Duboue, P. (2020). The art of feature engineering: Essentials for machine learning. Cambridge University Press, Cambridge, 1st edition.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. Journal of human evolution, 22(6):469–493.

Endendijk, J., Spencer, H., Bos, P., and Derks, B. (2019). Neural processing of gendered information is more robustly associated with mothers'gendered communication

with children than mothersímplicit and explicit gender stereotypes. Social Neuroscience, 14(3):300–312.

Etz, A. and Vandekerckhove, J. (2018). Introduction to Bayesian Inference for Psychology. Psychonomic Bulletin and Review, 25(1):5–34.

Fakhari, S. N. S., Ghaderi, F., Tehrani-Doost, M., and Charkari, N. M. (2023). EEG-based brain connectivity analysis in autism spectrum disorder: Unraveling the effects of bumetanide treatment. Biomedical Signal Processing and Control, 86:105054.

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., and Hutter, F. (2015). Efficient and robust automated machine learning. Advances in neural information processing systems, 28.

Fields, E. C. and Kuperberg, G. R. (2020). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. Psychophysiology, 57(2):e13468.

Fisher, A., Rudin, C., and Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. J. Mach. Learn. Res., 20(177):1–81.

Fokkema, M., Iliescu, D., Greiff, S., and Ziegler, M. (2022). Machine Learning and Prediction in Psychological Assessment. European Journal of Psychological Assessment, 38(3):165–175.

Forbes, C. E., Cameron, K. A., Grafman, J., Barbey, A. K., Solomon, J., Ritter, W., and Ruchkin, D. (2012). Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT). Frontiers in Human Neuroscience, 6(NOVEMBER 2012).

Fornito, A., Zalesky, A., and Bullmore, E. (2016). Fundamentals of brain network analysis. Academic Press.

Freeman, J. B. and Johnson, K. L. (2016). More Than Meets the Eye: Split-Second Social Perception. Trends in Cognitive Sciences, 20(5):362–374.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. Human brain mapping, 2(1-2):56–78.

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. Brain Connectivity, 1(1):13–36.

Garcia-Barrera, M. A., Karr, J. E., Trujillo-Orrego, N., Trujillo-Orrego, S., and Pineda, D. A. (2017). Evaluating empathy in colombian ex-combatants: Examination of the internal structure of the Interpersonal Reactivity Index (IRI) in Spanish. Psychological Assessment, 29(1):116–122.

Garrison, K. A., Scheinost, D., Finn, E. S., Shen, X., and Constable, R. T. (2015). The (in)stability of functional brain network measures across thresholds. NeuroImage, 118:651–661.

Gedeon, C., Esseily, R., and Badea, C. (2021). Examining differences in minority versus majority preschoolers on social categorization and perceived intergroup distance. Journal of Community & Applied Social Psychology, 31(1):94–106.

Giraldo, L. S., Aguirre-Acevedo, D. C., Trujillo, S., Ugarriza, J. E., and Trujillo, N. (2020). Validation of the Extreme Experiences Scale (EX2) for armed conflict contexts. Psychiatric quarterly, 91(2):495–520.

Gómez, D., López Hincapié, J. D., Cardona, L. S. G., Ugarriza, J. E., Herrera, E., and Trujillo, N. (2022). Structural analysis of the Reactive-Proactive Aggression Questionnaire in population exposed to armed conflicts. Peace and Conflict: Journal of Peace Psychology, 28(1):34.

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience, 7(7 DEC):267.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. Journal of personality and social psychology, 74(6):1464.

Groppe, D. M., Urbach, T. P., and Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology, 48(12):1711–1725.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A Survey of Methods for Explaining Black Box Models. ACM Comput. Surv., 51(5).

Gyorda, J. A., Nemesure, M. D., Price, G., and Jacobson, N. C. (2023). Applying ensemble machine learning models to predict individual response to a digitally delivered worry postponement intervention. Journal of Affective Disorders, 320:201–210.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In Varoquaux, G., Vaught, T., and Millman, J., editors, Proceedings of the 7th Python in Science Conference, pages 11–15, Pasadena, CA USA.

Hastie, T., Tibshirani, R., and Friedman, J. (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science+Business Media, LLC, New York, NY, USA, 2nd edition.

Healy, G. F., Boran, L., and Smeaton, A. F. (2015). Neural patterns of the implicit association test. Frontiers in Human Neuroscience, 9(NOV).

Henninger, M., Debelak, R., Rothacher, Y., and Strobl, C. (2023). Interpretable Machine Learning for Psychological Research: Opportunities and Pitfalls. Psychological Methods.

Hurtado, E., Haye, A., González, R., Manes, F., and Ibáñez, A. (2009). Contextual blending of ingroup/outgroup face stimuli and word valence: LPP modulation and convergence of measures. BMC Neuroscience, 10:1–21.

Ito, T. A. and Bartholow, B. D. (2009). The neural correlates of race. Trends in Cognitive Sciences, 13(12):524–531.

Ito, T. A. and Urland, G. R. (2003). Race and Gender on the Brain: Electrocortical Measures of Attention to the Race and Gender of Multiply Categorizable Individuals. Journal of Personality and Social Psychology, 85(4):616–626.

Jackson, T. S. and Read, N. (2010). Theory of minimum spanning trees. I. Mean-field theory and strongly disordered spin-glass model. Physical Review E, 81(2):21130.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). An introduction to statistical learning, volume 112. Springer, New York, NY, USA, 2nd edition.

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., and Gramfort, A. (2017). Autoreject: Automated artifact rejection for MEG and EEG data. NeuroImage, 159:417–429.

Jeffreys, H. (1938). Significance tests when several degrees of freedom arise simultaneously. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 165(921):161–198.

Johnstone, J., Gunkelman, J., and Lunt, J. (2005). Clinical database development: Characterization of EEG phenotypes. Clinical EEG and Neuroscience, 36(2):99–107.

Juárez Acosta, F. and Montejo Hernández, M. (2008). Psychometric properties of the situation and Aggressive Behavior Inventory and the motives for Aggression Inventory. Universitas Psychologica, 7(1):149–171.

Kamiński, M., Ding, M., Truccolo, W. A., and Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. Biological Cybernetics, 85(2):145–157.

Kang, S. H., Cheon, B. K., Kim, J.-S., Jang, H., Kim, H. J., Park, K. W., Noh, Y., Lee, J. S., Ye, B. S., Na, D. L., Lee, H., and Seo, S. W. (2021). Machine Learning for the Prediction of Amyloid Positivity in Amnestic Mild Cognitive Impairment. Journal of Alzheimer's disease : JAD, 80(1):143–157.

Kato, K., Kadokura, H., Kuroki, T., and Ishikawa, A. (2018). Event-related synchronization/desynchronization in neural oscillatory changes caused by implicit biases of spatial frequency in electroencephalography. In Lhotska L. Sukupova L., Lackovic I., I. G. S., editor, IFMBE Proceedings, vol 68/2, volume 68, pages 175–178. Springer Verlag.

Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E. S., Luck, S. J., Luu, P., Miller, G. A., and Yee, C. M. (2014). Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology, 51(1):1–21.

Keysers, C., Gazzola, V., and Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nature Neuroscience, 23(7):788–799.

Kim, D., Quan, L., Seo, M., Kim, K., Kim, J.-W., and Zhu, Y. (2023). Interpretable machine learning-based approaches for understanding suicide risk and protective factors among South Korean females using survey and social media data. Suicide and Life-Threatening Behavior, 53(3):484–498.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. Proceedings of the American Mathematical Society, 7(1):48–50.

Lachaux, J., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. Human brain mapping, 8(4):194–208.

Liberman, Z., Woodward, A. L., and Kinzler, K. D. (2017). The origins of social categorization. Trends in cognitive sciences, 21(7):556–568.

Lindholm, A., Wahlström, N., Lindsten, F., and Schön, T. B. (2022). Machine learning: a first course for engineers and scientists. Cambridge University Press, Cambridge, 1st edition.

Lones, M. A. (2021). How to avoid machine learning pitfalls: a guide for academic researchers. arXiv preprint arXiv:2108.02497.

Lu, Z., Xie, C., Liu, N., Xie, Y., and Lu, H. (2023). 'Can we predict aggression?'—Determining the predictors of aggression among individuals with substance use disorder in China undergoing enforced detoxification through machine learning. Journal of Affective Disorders, 320:628–637.

Luck, S. J. (2014). An introduction to the event-related potential technique. MIT press, Cambridge, MA, 2nd edition.

Luck, S. J. and Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology, 54(1):146–157.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1):56–67.

Lundberg, S. M., Erion, G. G., and Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in neural information processing systems, volume 30, pages 4766–4775.

March, D. S. and Graham, R. (2014). Exploring implicit ingroup and outgroup bias toward Hispanics. Group Processes & Intergroup Relations, 18(1):89–103.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods, 164(1):177–190.

Martínez-Florez, J. F., Osorio, J. D., Cediel, J. C., Rivas, J. C., Granados-Sánchez, A. M., López-Pelaéz, J., Jaramillo, T., and Cardona, J. F. (2021). Short-Term Memory Binding Distinguishing Amnestic Mild Cognitive Impairment from Healthy Aging: A Machine Learning Study. Journal of Alzheimer's Disease, 81(2):729–742.

Mitra, P. P. and Pesaran, B. (1999). Analysis of dynamic brain imaging data. Biophysical journal, 76(2):691–708.

Molenberghs, P. and Louis, W. R. (2018). Insights from fMRI studies into ingroup bias. Frontiers in Psychology, 9(OCT):1868.

Molnar, C. (2022). Interpretable machine learning: A Guide For Making Black Box Models Explainable. Independently published, 1st edition.

Murtagh, F. and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? Journal of classification, 31:274–295.

Newheiser, A.-K., Dunham, Y., Merrill, A., Hoosain, L., and Olson, K. R. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. Developmental Psychology, 50(4):1081.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. Clinical neurophysiology, 115(10):2292–2307.

Nosek, B. A., Greenwald, A. G., and Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In Social psychology and the unconscious: The automaticity of higher mental processes, pages 265–292. Psychology Press.

Nunez, P. L., Srinivasan, R., et al. (2006). Electric fields of the brain: the neurophysics of EEG. Oxford University Press, USA.

Nuzzo, R. (2014). Scientific method: statistical errors. Nature, 506(7487):150–152.

Open Science Colaboration, n. (2015). Estimating the reproducibility of psychological science. Science, 349(6251).

para la Paz, A. C. (2016). Acuerdo final para la terminación del conflicto y la construcción de una paz estable y duradera.

Pargent, F., Schoedel, R., and Stachl, C. (2023). Best Practices in Supervised Machine Learning: A Tutorial for Psychologists. Advances in Methods and Practices in Psychological Science, 6(3).

Paul, M., Govaart, G. H., and Schettino, A. (2021). Making ERP research more transparent: Guidelines for preregistration. International Journal of Psychophysiology, 164:52–63.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Perez-Sales, P., Eiroa-Orosa, F., Fernandez, I., Olivos, P., Vergara, M., Vergara, S., and Barbero-Val, E. (2013). La medida del impacto psicológico de experiencias extremas. Cuestionario VIVO. Diseño, validación y manual de aplicación [The measure of the psychological impact of extreme experiences. VIVO Questionnaire. Design, validation and administration manual].

Picton, T. W., Lins, O. G., and Scherg, M. (1995). The recording and analysis of event-related potentials. In Handbook of neuropsychology, volume 10, chapter 1, pages 3 – 73. Elsevier Science Ltd., 1st edition.

Poli, D., Pastore, V. P., and Massobrio, P. (2015). Functional connectivity in in vitro neuronal assemblies. Frontiers in Neural Circuits, 0(OCT):57.

Portengen, C. M., Huffmeijer, R., van Baar, A. L., and Endendijk, J. J. (2022). Measuring the neural correlates of the violation of social expectations: A comparison of two experimental tasks. Social Neuroscience, 17(1):58–72.

Prim, R. C. (1957). Shortest connection networks and some generalizations. The Bell System Technical Journal, 36(6):1389–1401.

Quintero-Zea, A., Trujillo Orrego, N., López, J. D., Rodríguez Calvache, M., Trujillo Orrego, S., Escudero, J., and Parra, M. A. (2019). Neural Reorganization During Emotional Face Processing in Ex-combatants. (In preparation).

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., and Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. Aggressive Behavior: Official Journal of the International Society for Research on Aggression, 32(2):159–171.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should i trust you?."Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pages 1135–1144.

Rösler, I. K. and Amodio, D. M. (2022). Neural Basis of Prejudice and Prejudice Reduction. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 7(12):1200–1208.

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., and Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. Psychological Methods, 22(2):304.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. NeuroImage, 52(3):1059–1069.

Saarinen, A., Jääskeläinen, I. P., Harjunen, V., Keltikangas-Järvinen, L., Jasinskaja-Lahti, I., and Ravaja, N. (2021). Neural basis of in-group bias and prejudices: A systematic meta-analysis. Neuroscience and Biobehavioral Reviews, 131:1214–1227.

Sandha, S. S., Aggarwal, M., Fedorov, I., and Srivastava, M. (2020). Mango: A python library for parallel hyperparameter tuning. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3987–3991. IEEE.

Sanz Fernández, J. (2001). Un instrumento para evaluar la eficacia de los procedimientos de inducción de estado de ánimo: la."Escala de Valoración del Estado de Ánimo"(EVEA). Análisis y Modificación de Conducta, 27:71–110.

Sassenhagen, J. and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology, 56(6):e13335.

Schindler, S., Schettino, A., and Pourtois, G. (2018). Electrophysiological correlates of the interplay between low-level visual features and emotional content during word reading. Scientific Reports, 8(1):1–13.

Schindler, S., Wolff, W., Kissler, J. M., and Brand, R. (2015). Cerebral correlates of faking: Evidence from a brief implicit association test on doping attitudes. Frontiers in Behavioral Neuroscience.

Shapley, L. S. (1953). A value for n-person games. In Contributions to the Theory of Games (AM-28), Volume II, chapter IV, pages 307–324. Princeton University Press, Princeton, NJ.

Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case. Bell System Technical Journal, 57(5):1371–1430.

Smit, D., Trevino, L., Mohamed, S. M. H., and Enriquez-Geppert, S. (2023). Theta power and functional connectivity as neurophysiological markers of executive functions in individuals with cognitive complaints in daily life. Biological Psychology, page 108503.

Smith, S. M. and Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage, 44(1):83–98.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information processing and management, 45(4):427–437.

Sporns, O. (2011). Networks of the Brain. MIT Press, Cambridge.

Stam, C. J., Nolte, G., and Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Human brain mapping, 28(11):1178–1193.

Stam, C. J., Tewarie, P., Van Dellen, E., van Straaten, E. C., Hillebrand, A., and Van Mieghem, P. (2014). The trees and the forest: Characterization of complex brain networks with minimum spanning trees. International Journal of Psychophysiology, 92(3):129–138.

Teige-Mocigemba, S., Klauer, K. C., and Sherman, J. W. (2010). 7. A practical guide to implicit association tests and related tasks. In Handbook of implicit social cognition: measurement, theory, and applications, pages 117–139. The Guilford Press.

Tewarie, P., van Dellen, E., Hillebrand, A., and Stam, C. J. (2015). The minimum spanning tree: An unbiased method for brain network analysis. NeuroImage, 104:177–188.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. Proceedings of the IEEE, 70(9):1055–1096.

Tosi, J. D., Ledesma, R. D., Poó, F. M., Montes, S. A., and López, S. S. (2018). El Test de Asociaciones Implícitas (Implicit Association Test). Una revisión metodológica. Revista Iberoamericana de Diagnóstico y Evaluación-e Avaliação Psicológica, 1(46):175–187.

Trujillo, S., Trujillo, N., Lopez, J. D., Gomez, D., Valencia, S., Rendon, J., Pineda, D. A., and Parra, M. A. (2017). Social cognitive training improves emotional processing and reduces aggressive attitudes in ex-combatants. Frontiers in psychology, 8:510.

Ugarriza, J. E., Villegas, C. A., Trujillo, N., and López, J. D. (2019). Restaurar tejidos sociales en comunidades durante el postconflicto a través de un enfoque de intervención biopsicosocial comprehensivo: estrategias hacia a la construcción de paz en Colombia. Programa de investigación presentado a la convocatoria 852-2019 de Colciencias.

Valencia, S., Trujillo, N., Trujillo, S., Acosta, A., Rodríguez, M., Ugarriza, J. E., López, J. D., García, A. M., and Parra, M. A. (2020). Neurocognitive reorganization of emotional processing following a socio-cognitive intervention in Colombian ex-combatants. Social Neuroscience, 15(4):398–407.

van Dellen, E., Sommer, I. E., Bohlken, M. M., Tewarie, P., Draaisma, L., Zalesky, A., Di Biase, M., Brown, J. A., Douw, L., Otte, W. M., Mandl, R. C., and Stam, C. J. (2018). Minimum spanning tree analysis of the human connectome. Human Brain Mapping, 39(6):2455–2471.

van den Heuvel, M. P., de Lange, S. C., Zalesky, A., Seguin, C., Yeo, B. T., and Schmidt, R. (2017). Proportional thresholding in resting-state fMRI functional connectivity networks and consequences for patient-control connectome studies: Issues and recommendations. NeuroImage, 152:437–449.

van Diessen, E., Otte, W. M., Stam, C. J., Braun, K. P., and Jansen, F. E. (2016). Electroencephalography based functional networks in newly diagnosed childhood epilepsies. Clinical Neurophysiology, 127(6):2325–2332.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., and Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. Psychonomic Bulletin and Review, 28(3):813–826.

Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., and Pennartz, C. M. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. NeuroImage, 55(4):1548–1565.

Volpert-Esmond, H. I. and Bartholow, B. D. (2019). Explicit categorization goals affect attention-related processing of race and gender during person construal. Journal of Experimental Social Psychology, 85:103839.

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kes-

teren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., de Jong, T., van den Bergh, D., Sarafoglou, A., Steingroever, H., Derks, K., Rouder, J. N., and Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. Psychonomic Bulletin and Review, 25(1):58–76.

Wagenmakers, E.-J., Morey, R. D., and Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. Current Directions in Psychological Science, 25(3):169–176.

Wang, H., Sun, Y., Lv, J., and Bo, S. (2019). Random topology organization and decreased visual processing of internet addiction: Evidence from a minimum spanning tree analysis. Brain and Behavior, 9(3):e01218.

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American statistical association, 58(301):236–244.

Westfall, P. H., Johnson, W. O., and Utts, J. M. (1997). A Bayesian perspective on the Bonferroni adjustment. Biometrika, 84(2):419–427.

Wetzels, R., Grasman, R. P. P. P., and Wagenmakers, E.-J. (2012). A Default Bayesian Hypothesis Test for ANOVA Designs. The American Statistician, 66(2):104–111.

Williams, J. K. and Themanson, J. R. (2011). Neural correlates of the implicit association test: Evidence for semantic and emotional processing. Social Cognitive and Affective Neuroscience, 6(4):468–476.

Yao, D. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. Physiological Measurement, 22(4):693–711.

Yarkoni, T. and Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspectives on Psychological Science, 12(6):1100–1122.

Yeo, I. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 87(4):954–959.

Yuste, R. (2015). From the neuron doctrine to neural networks. Nature reviews neuroscience, 16(8):487–497.

Zakharov, I., Adamovich, T., Tabueva, A., Ismatullina, V., and Malykh, S. (2021). The effect of density thresholding on the EEG network construction. In Journal of Physics: Conference Series, volume 1727, page 012009. IOP Publishing.

Zhang, J., Liu, T., Shi, Z., Tan, S., Suo, D., Dai, C., Wang, L., Wu, J., Funahashi, S., and Liu, M. (2022). Impaired Self-Referential Cognitive Processing in Bipolar Disorder: A Functional Connectivity Analysis. Frontiers in Aging Neuroscience, 14:754600.

Zhang, L., Wang, P., Zhang, R., Chen, M., Shi, L., Gao, J., and Hu, Y. (2020). The Influence of Different EEG References on Scalp EEG Functional Network Analysis During Hand Movement Tasks. Frontiers in Human Neuroscience, 14:367.

Zheng, A. and Casari, A. (2018). Feature engineering for machine learning: principles and techniques for dat O'Reilly Media, Inc., Sebastopol,CA,USA, 1st edition.