UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE PSICOLOGÍA

TESIS DOCTORAL

Correlatos neuronales de las emociones sociales de culpa, vergüenza y orgullo

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

José Sánchez García

Director

Manuel Martín-Loeches Garrido

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE PSICOLOGÍA

TESIS DOCTORAL

Correlatos neuronales de las emociones sociales de culpa, vergüenza y orgullo

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Jose Sánchez García

DIRECTOR

Manuel Martín-Loeches Garrido

A mis padres, de donde venimos.

A mi mujer, quienes somos.

A mi hijo, a donde vamos.

A Arutam, fuerza de vida que nos traspasa, tan inefable y esquiva para los necios.

AGRADECIMIENTOS

Aunque en la cultura en la que habito, el éxito es una extraña entidad asociada a la actitud, las agallas y la voluntad de un Yo individual (o acaso a las oportunidades de un Yo con amigos), en el mundo que vivo dentro nada se hace solo, sino que la mayoría de las cosas suceden por sí mismas, lo cual relaja incluso en medio de la guerra y el resto, se hacen apoyándose unos a otros. No hay mucho más espacio para Narciso.

En el caso de este texto, la emoción que surge es aún más extraña, porque resulta sospechoso reclamar autoría alguna sobre aquello que jamás hubiera surgido por talento especial o actitud encomiable, ambas tan individuales y cuestionables a mi entender. Es por tanto más importante en este viaje del héroe que parece una tesis doctoral, agradecer, sincerarse o enraizarse que elevarse o sentirse autor, por mucho que uno tenga que poner el nombre y defender estos estudios, más bien en representación de todos que no solo de sí mismo. Al fin y al cabo el viaje del héroe, suele ser el de alguien un poco bobo, que como Skywalker, tarda 6 películas en conocer a su padre.

En los agradecimientos sucede algo parecido a lo anterior con los esfuerzos. Se pueden agradecer los niveles técnicos o centrarse en los más importantes: los humanos. Para éste, que está de vuelta de muchos viajes en torno a un centro ciertamente esquivo, solo me interesa a estas alturas comprobar cuán bueno es alguien y no tanto sus dotes de memoria aplicada a un software o tarea. Importa más si posee esa inteligencia social que nos trasciende y que brilla en capacidades para respetar o apoyar al otro, en sinceras ausencias de juicio sobre aquel a quien se dirige, en posibilidades de ponerte en la piel del de enfrente o en esa rara habilidad que pocos profesan de confiar tanto, tanto en el otro, que a ojos de Zeus, convierte a Galatea en mujer, ante la pasión de Pigmalión.

Valoro esa humanidad mucho más que hacer de Sísifo cargando estúpidamente una y otra vez la piedra del esfuerzo montaña arriba para demostrar perfeccionismo y voluntad o quien sabe si destino, mientras me temo, creaba hernias discales. O peor aún, hacer de Milarepa subiendo las piedras desde la base para hacer una casa en la cima, que Marpa se encarga de destruir, para que el posterior tibetano más grande de su historia limpie karma y acumule mérito.

Comienzo desde este lugar por tanto, agradeciendo a Manolo, mi director, por muchas características y detalles. Muchas de ellas probablemente escaparán a sus ojos y entendimiento. Sospecho, como ingeniero camuflado, que no soy aquel que esperaba, pero espero no haberle defraudado del todo, siendo quien soy o acaso en quien me estoy convirtiendo. El caso es que cuando uno espera, como dice la tradición, una fricción constante entre director y doctorando, junto con la clásica y advertida depresión a mitad de tesis y una relación típica con un catedrático petulante con corbata, se encuentra con Manolo, el cual escapa a toda clasificación vulgar, incluso al mismísimo Big Five. La realidad es que no recuerdo fricción alguna en estos años, acaso las mías propias con mis desatinos. La depresión no apareció a la mitad, sino en todo el recorrido, porque no hay cosa más inapropiada que hacer una tesis a estas alturas y circunstancias, si bien me motiva como decían de Drake en *100 años de soledad*, vínculos más poderosos que el amor, o sea, el remordimiento. Los supuestos estereotipos de catedrático alfa fueron más bien paciencia, perspectiva, estilo directivo y foco: afortunado soy por tanto. Sobre todo ello Manolo destaca con otra virtud: su sentido del humor, incluyendo una parte para “meterse conmigo” y es que nadie se libra de su creatividad… pero que ha sido una gran enseñanza e incluso aunque él no lo sepa, motivo de práctica introspectiva. Gracias a Manolo, veo que la ciencia es algo más modesto de lo que había conocido e imaginado antes y eso no se enseña, sino que lo vive y te cala. Se realizan hipótesis pequeñas, con diseños simples y conclusiones modestas en donde la estridencia se encuentra siempre ausente. Este camino que él marca tiene ese toque japonés que tanto venero: el placer de lo pequeño, Wabi Sabi, o el silencio, que siempre es más armónico que hablar siendo pensado por uno mismo en atención ausente. Hay otro regalo inesperado: como profuso lector que es, él me ha acercado a autores que por mí mismo hubiera tardado mucho en llegar a ellos o quién sabe si nunca. Cuanto valor tiene esto último. Al colarme en su biblioteca y su entendimiento, abrí huecos insospechados en mi cerebro. Por último, para quienes agradecemos la N400 de la verdad flagrante más que la comodidad del autoengaño, para los que nos resultan insoportables los rodeos, las esquivas extrañas, o el sexto nivel del "mentalizing" de "si me dice B porque le digo A, le contesto C", el ir al grano de Manolo, es un Kyosaku en toda regla. No solo se agradece sino que como en el buen Dojo, el Kyosaku se pide y si no, Manolo igualmente te lo da. La vida tiene sus más y sus menos, pero lo cierto, es que al final acabo siempre a lomos de gigantes. En esta tesis con el director, sin duda también.

Con Pili la cosa no empezó bien. El primer día, me senté en su silla y ocupé su mesa, con la lógica inocente de que estaba vacía y así me dispuse a arrancar con buen pie la tesis doctoral. A los 15 minutos llegó, menos mal que no tenía los pies encima y amablemente me ofreció sentarme enfrente. Claro que antes le expliqué infantilmente: "soy Jose, soy nuevo y vi una mesa vacía". Pili ha sido siempre un punto de inflexión para mí, ella es como esa derivada igual a cero que te dice: ahora te toca subir. Como se puede deducir, al ser varias las ocasiones, la tesis es por tanto un viaje sinusoidal. Con Pili, he podido ver señal en vez de artefactos y sin ella estaría inmerso en un electrodo epiléptico, o como un gallo batiendo alas en un corral para impresionar a las gallinas… mientras pasa el halcón volando camino al bosque. Si Manolo ha aportado dirección, foco, guía, perspectiva y resolución, Pili me ha llenado de estructura, algo que necesito

porque soy de los que ven bosque, o incluso selva, en vez de un árbol. Todo menos vivir en el corral de gallitos y gallinas: Pili me ha ayudado a no estar ahí.
Además de su humanidad debo decir, que comprender sus tesis del *embodied syntax*, es una tarea que te hace sentir de nuevo, que no eres nadie, absolutamente nadie. Si la meditación sirve para que no te creas que eres un Yo, y que por fin descanses en la vida, las tesis de Pili te lo muestran desde otro lugar. Destaco una paradoja más: es la psicóloga sin serlo a la que uno esperaría ir a terapia, por la sencilla razón de que cuando uno sufre de verdad, lo único que necesita alguien no es curación, sino no sentirse juzgado. Qué más decir de Pili…

David, o Da-Big si se quiere utilizar un nombre más apropiado, es ese compañero que en medio del caos te mete a que le acompañes en un Master de Metodología porque así se (mal-) trata a los amigos. Tiene alma de Pascal, cuerpo de luz de Laplace y sin embargo permanece libre y anumérico. Cuando te habla entiendes porqué la criptografía y el mismo Turing, son ciencias non-natas. Más allá de los dones citados, siente la ciencia con mucho rigor. Es metódico, trabaja como un pro y de corazón espero que se cumpla un sueño, donde le veía como un Pope de la neurolingüística mundial. A lo largo de los años ha estado siempre cerca, disponible y valorando mis integrales como hazañas. Lo más importante con Da-big ha sido que se convirtió en alguien en quien puedo confiar. ¿Qué precio tiene eso en esta vida de giros y surcos donde el prefrontal dice una cosa y el hipotálamo la contraria? Para mí muchísimo, porque lo que no suma, resta, y a cierta edad, no está uno para inocentadas y promesas vacuas personales o profesionales. Así que sí, disfruto de sus alegrías y lloro con sus penas, faltaría más. Pero con Da-Big me he sentido como Rinzai, aquel maestro que a la mañana reía a carcajada limpia al despertar y a la noche hacía lo propio, sin motivo aparente, para sorpresa de sus discípulos. Y es que han sido muchísimos los momentos donde nos hemos reído por reír, ni siquiera ya de nosotros mismos o las situaciones, sino quizá como una forma sabia y oculta de parodiar nuestras miserias, presiones y anhelos incumplidos.

Sabela tiene la virtud de ver a los demás fuertes, bien preparados y ciertamente inteligentes. Dado que cualquier camino cuando se empieza te lleva irremediablemente a sentirte tonto, y yo me he sentido dispuesto a reconocerlo sin sufrimiento alguno, Sabela es una luz. No te adula ni te halaga sino que sientes que su mirada sincera está teñida de ver potenciales ocultos, incluso en las cuevas a las que uno mismo ha caído. Me recuerda el ejemplo de C.J. Skender, ese afamado profesor de Duke, citado en el bellísimo libro de Adam Grant, *Give and Take*, donde todos los alumnos alcanzan cotas indescriptibles por la sencilla razón de que Skender ve siempre talento, motivación y potenciales. Cuando Sabela corrige el paper, notas compañía, apoyo y la forma concreta en la que puedes hacerlo mejor. En medio de una maraña de erratas en las revisiones, técnicas y científicas, sus mensajes son neurociencia social pura. Con la misma naturalidad que te acoge, también se abre y comparte intereses, sueños o ideas lo cual de nuevo te ayuda a sentirte uno más. Para muchos el *Openness* suele ser motivo de mofa más que de capacidad: estoy acostumbrado. Pero Sabela no se asusta, ni se posiciona provinciana, más bien escucha, se interesa y participa. Y es que me temo que la curiosidad es una virtud fundamental suya, no solo en la ciencia lo cual es esperable, sino en la vida, rara avis, curiosidad que ella derrocha y contagia.

Laura tiene loco a Alan Baddeley, porque su ancho de banda le permite llevar proyectos, artículos, clases y maratones, sí, sí, maratones como si tal cosa. Esto huele a que además del bucle fonológico y la agenda visoespacial, existe un módulo de tinte Chomskyano, llamado Jiménez-Ortega, encapsulado en ella misma y que desde luego explica FoxP2. Me ha resultado inspiradora su capacidad encomiable y me siento agradecido por sus ánimos en este viaje. Sin duda es un ejemplo a lo *Di Stefano*: roba balones, lanza faltas, mete goles y da pases. Laura es así: dirige tesis, publica artículos, hace proyectos, da clases y habla con las palomas, o hasta según se mire, de ahí lo del FoxP2. Omnipotente no sé, pero multipresente sin duda. Tiene la virtud de los mecánicos de antes: chapa, motor o pintura y eso es una garantía en procesos de eslabones lineales donde más de uno dará error. ¡Mucho valor tiene todo eso!

Paco, además de ser cliente asiduo de restaurantes Michelín, es otro fajador del laboratorio. Su nivel de entendimiento de bucles, algoritmos, triggers y clusters le hacen merecedor de un puesto de honor en Black Mirror Bandersnatch. Sería presuntuoso por mi parte intentar comprender todo eso, me conformo con observar y asentir. Sus esporádicos paseos hacia el WC cuando pasa por tu pantalla suelen ser apocalípticos: se queda mirando, sabe lo que estás haciendo y encima te lo puede explicar.

Javi me sorprendió por su meticulosidad y paciencia, de ahí que le gustara y dominara la programación del lab. Ante él mis propios fallos me generaban menos stress de lo habitual, vete a saber porqué. Lo cierto es que en el primer experimento su ayuda fue muy importante para coger unas mínimas tablas y parecer que haces algo, hasta que Manolo de alguna manera te hacía la pregunta precisa para darte cuenta que no sabías qué estabas haciendo y, como en las pelis de espionaje, tienes unas horas de experiencia y escombros sobre los que rehacer una misión. Esto no es Stanford, pero tiene un valor incalculable.

Susana fue muy buena ayudante y me sentí, aunque fuera por solo unos meses, muy correspondido con todo lo que humildemente pude explicarla y de ahí unas líneas aquí. Qué lindo recordar Wirikuta con una mexicana o cómo la amabilidad que viene del náhualt en la travesía desértica de una tesis que te permite hasta abrazar un cáctus, puede venir de quien te ayuda a colocar un simple electrodo.

Clara ha sido la imagen del "easy going" que tanto aprecio en algunos amigos americanos. Inteligente y reservada, no ha escatimado ni esquivado nunca una buena conversación ni una visión amplia sobre todo aquello que estábamos haciendo, como si pudiera contextualizar en la vida una simple tarde de EEG. Es de esas personas que ha ido más allá de un papel X en un laboratorio Z y a día de hoy, disfruto de su amabilidad y escucho atento sus sueños e ideas para un mundo más sincero y sencillo.

El resto de compañeros que estuvieron o están en el laboratorio como Miguel, Oscar, Esperanza, Alvaro, Rahele, Esther, Marta y otros también han sido parte de mi aprendizaje, reflejo de mis procesos (incluidas mis neurosis) e incluso objeto de mi admiración silenciosa. A algunos de ellos que les tocó recibir algo mío, os recuerdo que el "disclaimer" decía: "se asegura la empatía, no el conocimiento". A tod@s también muchas gracias.

"Dice la razón: busquemos la verdad

Y el corazón: Vanidad.

La verdad ya la tenemos.

La razón: ¡Ay, quien alcanza la verdad!

El corazón: Vanidad.

La verdad es la esperanza.

Dice la razón: Tú mientes

Y contesta el corazón:

Quien miente eres tú, razón,

Que dices lo que no sientes.

La razón: Jamás podremos entendernos, corazón.

El corazón: Lo veremos"

Antonio Machado

INDEX

Neural correlates of the social emotions guilt, shame and pride

ABSTRACT

Brain research on emotions is generally focused on basic emotions such as anger, fear, sadness, joy, contempt, surprise and disgust. These basic emotions, are by far, the most well understood. They are easy to approach, presumably universal and linked to other primates and non human animals. These characteristics have permitted a broader search. Social emotions as guilt, shame, embarrasment and pride, nevertheless, remain largely unexplored, although our life is tainted with them,.

Social emotions are different to basic ones. They require complex cognitive and evaluative aspects, and not only emotional. In experimental settings are usually studied stemming from imagining or remembering situations removed from social reality. Participants imagines him/herself acting in a precise situation, or tries to restore a memory, relatively recent, in which that emotion has been felt in a similar social context. This experimental setting is very different to social life, where social emotions emerge as a reaction to self predictions, precise contexts, standards and norms.

Through three studies, this PhD research shows that social emotions are, spatially and temporally, created in the brain. Previous studies do not generate social emotions in the lab. Our aim has looked for a more ecological setting to study the social emotions of guilt, shame and pride where their emergence is not related to memory or imagination.

Participants performed several games in which the main task was to cooperate or compete with others while their brain activity was measured by electroencephalography (EEG). Event-related potentials were measured in the critical moments of guilt, shame, pride, joy or absent of emotion, according to each experiment.

In Study 1, we manipulated a dot estimation task result with 24 pairs (participant and friend) in which the different results elicited our conditions of interest, guilt and shame. A frontal negativity, replicating and broadening the results of a previous study (Leng et al, 2016), explained the emergence of empathic processes linked to the emotion of guilt and with a presumably medial prefrontal cortex activation related to the frontal negativity..

In Study 2, we focused on shame, trying to separate it from it from guilt, with a similar paradigm (dot estimation task) but creating new roles (adviser and decisor), for every pair of participants. Medial areas were again involved in shame, with an early centro parietal negativity and a lateralized posterior frontal positivity.

In Study 3, we approached shame and pride simultaneously with the dot estimation task and a particular social exposure, different to previous studies. Other students from different universities participated in real time in the dot estimation task simultaneously with our participants. The different experimental design combinations allowed us to graduate the maximal and minimal degrees of pride and shame elicited. Results showed early negativites and late positivities, frontals and centrals, for both emotions.

Results in the three studies performed, consistently show that social emotions can and must be researched in a social context in the lab through ecological settings more similar to daily situations and imply a two-step brain process, early and late, with brain activity linked mainly to medial areas.

Correlatos neuronales de las emociones sociales de culpa, vergüenza y orgullo

RESUMEN

La investigación cerebral tradicional sobre las emociones se centra en las llamadas emociones básicas, siendo ira, miedo, tristeza, alegría y asco las más estudiadas. Su facilidad de estudio, ubicuidad y posible vínculo con otros primates y animales no humanos, han posibilitado una extensa investigación. Las emociones sociales, entre las que destacan principalmente la culpa, la vergüenza, el orgullo y el bochorno, a pesar de que nuestra vida está teñida de ellas, permanecen aún con muchísimos aspectos por explorar.

Estas emociones sociales a diferencia de las básicas, implican aspectos cognitivos y evaluativos complejos, no solo emocionales. En el ámbito experimental se suelen elicitar a partir de situaciones imaginadas o recordadas, en los que el sujeto se visualiza a sí mismo respondiendo en una situación precisa, o trata de rescatar de su memoria una experiencia relativamente reciente en la que haya sentido esa emoción en un contexto social similar. Evidentemente este modo experimental está muy lejos de nuestra realidad social cotidiana en la que se las emociones sociales, surgen in situ, como respuesta a predicciones propias, contextos precisos, normas y acuerdos comunes, que toda vida social requiere.

A lo largo de tres estudios, esta tesis investiga cómo se forman en el cerebro, temporal y espacialmente, estas emociones sociales. A diferencia de la mayoría de los estudios hasta ahora realizados, nuestro interés ha sido generar la emoción social en el laboratorio, en lo que hemos considerado un escenario más ecológico que el recuerdo o la imaginación de las mismas.

Los participantes realizaron diversos juegos en los que la competición y cooperación con otros era la principal característica y al mismo tiempo se registró su actividad eléctrica cerebral mediante un electroencefalograma (EEG). Se calcularon potenciales evento–relacionados (PERs) en los momentos críticos, los cuales, dependiendo del experimento, implicaban una condición de culpa, vergüenza, orgullo, alegría o ausencia de emoción.

En el Estudio 1, se manipuló el resultado de una estimación de puntos realizada por 24 parejas (participante y amiga) en la que los diferentes resultados elicitaban las condiciones de interés, siendo culpa y vergüenza las que fueron objeto de nuestro estudio. Una negatividad frontal, replicando y ampliando resultados de una investigación previa (Leng et al, 2016) explicaba la emergencia de procesos empáticos vinculados a la emoción de culpa y con la posible partipación de la corteza medial prefrontal en la negatividad citada.

En el Estudio 2, nos centramos en el análisis de la vergüenza, intentando diferenciarla de la culpa, en un paradigma similar de estimación de puntos y manipulación de resultados pero cambiando los roles de la pareja participante (consejera y decisora. Áreas mediales estuvieron de nuevo implicadas en la emoción social buscada, con negatividad temprana centro parietal y positividad frontal posterior lateralizada a la derecha.

En el Estudio 3, estudiamos la vergüenza y el orgullo simultáneamente, con la tarea de estimación de puntos y una exposición social diferente a los anteriores estudios, al realizar el experimento junto con otros participantes de otras universidades online in vivo, aumentando el carácter de exposición social como inductor de las emociones buscadas. Las diferentes combinaciones del diseño experimental permiten graduar el nivel máximo y mínimo de vergüenza/orgullo elicitados obteniendo negatividades tempranas y positividades tardías, frontales y centrales para ambas emociones. Las áreas mediales del medial prefrontal, precuneus y cingulado posterior se vinculan como posibles fuentes de las negatividades y positividades obtenidas.

Los resultados obtenidos en los tres estudios demuestran que las emociones sociales pueden y deben ser investigadas en un contexto social en el laboratorio a través de escenarios más ecológicos y cercanos a la vida real, e implican un procesamiento cerebral en 2 etapas, temprana y tardía, con actividad vinculada fundamentalmente a áreas mediales.

PREFACE

Many years ago in a far, far away mexican desert a book called "El cerebro nos engaña" was by chance with me. Although it was a popular book it lead to a realization in my head since I realized then something which is very clear to me now: all human experiences are brain centered ones (but not brain-centric). And of course, those experiencies were indeed brain experiences, but not illusory. Years before, when my father lost his job, and somehow his health, he brought to me one only book book from the printing house where "Cuadernos para el diálogo" was published, specifically one called"Las enseñanzas de Don Juan". The author was a naughty anthropologist, compulsive liar, handler lady´s man but nevertheless a genius novelist and stalking master called Carlos Castaneda. Between algebra and thermodynamics, the teachings of Don Carlos and Don Genaro spread in my mind at times and paved the way to a Viaje a Itxlán in my youth as a true seeker, yes, amidst a strange industrial engineering and a devoted martial arts practise. In that desert, I was incidentally learning with an autor who has Castaneda's direct and smart oustanding pupil. He rooted these novels by applying a psychological and meaningfull approach, far away from an esoteric view. I searched and followed him for years after reading his book for him and followed years after reading his book. But, returning to the desert, all of this was clearly and without the shadow of a doubt, a brain experience. It was 2007 and a true interest in researching about where we came from, who we are and where we go arose in me, not only from a first person view but also from a third person's perspective, indeed a scientific one. It is clear that I never wanted to put out the fire of these kinds of answers with religious fire extinguishers. The path seemed exhausting and it truly has been this way, so why neglect it?: studying a second degree, three Masters and a PhD at an age where one perhaps should be doing other things. If the turns of the brain are complex, the turns of life gyri are not under our control. Naive in neuroscience, but I emailed Dr. Rubia, the autor of that desert book, when studying a Biotechnology degree in 4th year. I was looking for a PhD director long before enrolling in the Neuroscience Master. I was very fortunate, because in the that Master, there seemed to me to be only one option that could establish a link between life, dayly matters and my ectelctic past. That option turned to be the Center for Evolution and Human Behavior (neuroscience section) directed by Manuel Martín-Loeches. I am not trying to reorder the pieces of my life pieces from a retrospective view in order to highlight a serendipia, but rather, as we are the constructors of our own feelings and emotions, the meaning we believe can be the starting point and not the finish line.

Therefore I choose one that respects randomness, that embraces what we do not know yet and, as all human experiences are brain centered and most of them are social, this includes me as a part of something truly human, here and now, together.
So, since some years ago for a Master thesis dissertation and after that for a PhD, I started finally, jumping from one inspiring book to another one, in C.E.C.H for this new journey.

I admit I am twice lucky as emotions are a highly relevant subject for me either both personal or profesionally, although it is hard to extract something applicable from a “frontal negativity”. For better or worse, my best and worst decissions in life have been tainted by social emotions: an indescriptible shame for loosing, absolute desperation after a heartbreak or a planned revenge because of a sport match-fixing. Those with strange valence, being as ecological as well as social, have arosen my curiosity: unspeakable awe in nature, a strange serenity and calmness in the middle of the storm or a devotional feeling to people who live merging into an obliged present.

Social emotions dye life with colours and keep it far from the grays. They are so fast that they seem conductists but they truly tend to arise from our previous assumptions about what we must feel and how we have to respond to precise contexts in determined contexts/situations.

Researching emotions in third person is paradoxical, as any subjective human experience far from the hard sciences, but when talking about social emotions, even more.

We could somehow through this research come back to “Cuadernos para el diálogo”. I can not find a better example than remembering myself sitting down in a train station between a renowned psychologist in “emotional intelligence” and a medicine man without any academic credentials in third person, but with a conscious life in the first one. The train broke down and an announcement call told us to wait four hours more. The psychologist said: “Damn it! Four hours here waiting! Medicine man replied: “mmm, four hours to remain calm”.

Despite the obvious difficulties, in the three studies that form this PhD thesis we, somehow, have tried to stablish those encounters between real life and the laboratory, thinking, designing and setting up scenarios where participants felt at that moment the social emotion. In the name of science, we had to dupe them, as it is usual in psychology experiments to reach the most objective standards and therefore collect and measure what we were prompting to feel.

These experimental designs represent a way to study social emotions from an ecological view, a bridge between first and third person with the measure of a brain activity signal that lights up a step further in this complex affective neuroscience world, yet to be developed.

We have spared no efforts in this direction studying the neural correlates of social emotions in ecological lab settings. Guilt, shame and pride were the first chosen ones, and have been a central meeting point of thoughts, debates, analysis, discussions, corrections and publications. They surprisely elicited in us, hope, joy or frustration, and luckily less from themselves. There is still much to learn from them and other social emotions, but this PhD is a good summary of a teamwork and that took a significant number of years in this field that someone undoubtfully will follow latter taking the lead.

In this journey of social emotions several people have been involved in the researchs and as coauthors of the scientific papers. They are:

Study One*: Jose Sánchez-García, Javier Espuny, David Hernández-Gutiérrez, Pili Casado, Francisco Muñoz, Laura Jiménez-Ortega, Sabela Fondevila, Manuel Martín-Loeches.

Study Two: Jose Sánchez-García, David Hernández-Gutiérrez, Pili Casado, Sabela Fondevila, Francisco Muñoz, Laura Jiménez-Ortega, Manuel Martín-Loeches.

Study Three**: Jose Sánchez-García, Gema Esther Rodríguez, David Hernández-Gutiérrez, Pili Casado, Sabela Fondevila, , Laura Jiménez-Ortega, Francisco Muñoz Miguel Rubianes, Manuel Martín-Loeches.

Published: Sánchez-García, J., Espuny, J., Hernández-Gutiérrez, D., Casado, P., Muñoz, F., Jiménez-Ortega, L., ... & Martín-Loeches, M. (2019).Neurofunctional

characterization of early prefrontal processes contributing to interpersonal guilt. Cognitive, Affective, &BehavioralNeuroscience, 19(5), 1192-1202.

** Published in BrainStructure and Function.

PREFACIO

Hace muchos años en un desierto mexicano muy muy lejano me acompañaba casualmente un libro llamado *"El cerebro nos engaña"*. Aunque de corte divulgativo, provocó un estallido en mi cabeza porque me permitió darme cuenta de algo evidente ahora, y es que toda experiencia humana es cerebral (que no cerebrocéntrica). Y vaya si eran cerebrales aquellas experiencias, que no ilusorias. Años antes, cuando mi padre perdió su trabajo y parte de su salud, rescató un único libro de aquella imprenta donde se publicaban *"Cuadernos para el diálogo",* concretamente uno llamado *"Las enseñanzas de Don Juan"* cuyo autor era un pillo antropólogo, mentiroso compulsivo, mujeriego manipulador, pero genial novelista y maestro del acecho llamado Carlos Castaneda. Entre algebra y termodinámica se colaban a ratos esas enseñanzas de Don Juan y Don Genaro y fueron instalando un *Viaje a Itxlán* en mi juventud como buscador genuino, sí, en medio de una ingeniería extraña y una devoción a las artes marciales. En ese desierto, casualmente me encontraba aprendiendo con un autor, alumno aventajado de Castaneda que enraizaba esas novelas y les daba un sentido y dirección práctica psicológica muy concreta, lejos de una interpretación esotérica, y al que busqué y seguí durante años tras leer su libro. Pero volviendo al desierto, resulta que todo eso era, evidentemente, una experiencia cerebral. Estaba en el año 2007 y nació un interés muy fuerte por investigar no solo en primera, sino en tercera persona desde la ciencia, algo del quiénes somos, de dónde venimos y adónde vamos, respuestas que nunca quise apagar con extintores religiosos. El camino parecía agotador y así ha sido, para qué negarlo: estudiar una segunda carrera, 3 masters y una tesis doctoral a una edad en la que uno debería quizá estar haciendo otras cosas. Si los giros del cerebro son complicados los de la vida no son aptos para el control. Profano y ridículamente inocente en esto de la neurociencia, escribí al Dr. Rubia, el del libro del desierto, en cuarto de Biotecnología, buscando un director para una tesis cuando ni siquiera había empezado aún el Master. Acabé siendo muy afortunado, porque de todas las opciones que vi posteriormente en el Master de Neurociencia, solo una me parecía que podría tener un vínculo aplicable con la vida, lo cotidiano y mi ecléctico pasado. Esa opción resultaba ser la del grupo del Centro de Evolución y Comportamiento Humanos dirigido por Manuel Martín-Loeches, donde se investigaba todo ello. No trato de ordenar las piezas de mi vida en retrospectiva para destacar la posible serendipia, sino que, como constructores de emociones y sentimientos que somos, el sentido puede ser el punto de partida y no la meta. Y elijo uno que respete el azar, abrace lo que no sabemos aún y dado que toda experiencia es cerebral y la mayoría son además sociales, me incluya como parte de algo humano, aquí y ahora, juntos.

Así que desde hace unos años para un TFM y después para una tesis, empecé por fin, de libro inspirador en libro inspirador, en el C.E.C.H. para este nuevo viaje.

Confieso que he sido doblemente afortunado porque las emociones es un tema que me resulta profesional y personalmente relevante, aunque cuesta sacarle jugo profesional a una "frontal negativity". Al fin y al cabo, mis peores y mejores decisiones en la vida han estado teñidas de emociones sociales: que si una vergüenza indescriptible por una derrota, la desolación absoluta del desamor o la venganza planificada ante un amaño deportivo entre otras. Las de valencia rara, por ser tan ecológicas como sociales, han ocupado mi curiosidad: un sobrecogimiento inmensurable una y otra vez ante la naturaleza, una extrañísima serenidad en medio de las peores tormentas o una devoción intensa por los que viven mimetizados con un presente obligado.

Las emociones sociales tiñen la vida de colores y la alejan de los grises. Son tan rápidas que parecen conductistas y sin embargo suelen nacer de interpretaciones previas de qué debemos sentir y cómo debemos responder ante determinados contextos.

Investigar las emociones en tercera persona resulta paradójico, como toda experiencia subjetiva humana lejana a las ciencias duras, pero en el caso de las emociones aún más.

Quizá podamos aquí con estas investigaciones retomar de alguna manera esos *Cuadernos para el diálogo* y para ello no encuentro mejor ejemplo que uno de hace años en una estación de tren, sentado entre un reputado psicólogo experto en "inteligencia emocional" y un hombre medicina sin estudios en tercera persona pero toda una vida consciente en la primera. El tren se averió y por megafonía nos anunciaron un retraso de 4 horas. El psicólogo dijo: "Joder, cuatro horas aquí esperando". El hombre medicina replicó: "Mmm, cuatro horas para estar serenos".

De alguna manera, a pesar de las evidentes dificultades, en los tres estudios que conforman esta tesis hemos intentado establecer esos encuentros entre el laboratorio y la vida real, reflexionando, diseñando y poniendo a prueba escenarios en los que los participantes sintieran allí mismo in situ, ante el contexto creado, nada más y nada menos, que una emoción social. En nombre de la ciencia hemos tenido que engañarles, como es habitual en los experimentos de psicología para ser lo más objetivos posibles, y de este modo, medir objetivamente lo que inducíamos a sentir.

Estos diseños experimentales representan una vía para el estudio de las emociones sociales de forma ecológica y cercana, un intento de puente entre esa primera y tercera persona con una medida de la actividad cerebral que nos ilumine un poquito más en este complejísimo mundo de la neurociencia afectiva, aún en ciernes.

No hemos escatimado esfuerzos en esta dirección en el estudio de correlatos neurales de emociones sociales en entornos ecológicos de laboratorios. Culpa, vergüenza y orgullo han sido las primeras elegidas y fueron objeto de lecturas, reflexiones, debates, análisis, discusiones, correcciones y publicaciones. Nos elicitaron curiosamente esperanza, alegría o frustración y afortunadamente menos de ellas mismas. Seguimos aún con mucho por aprender de ellas y otras emociones sociales, pero esta tesis es un buen resumen de un trabajo común y colectivo en esta dirección que ya implica un considerable número de años y que sin duda alguien se encargará de ampliar y profundizar.

En este viaje de experimentos en emociones sociales varias personas han formado parte. Las que han participando en la realización de los experimentos que forman la tesis y son coautores de las publicaciones científicas correspondientes son:

Estudio Uno:* Jose Sánchez-García, Javier Espuny, David Hernández-Gutiérrez, Pili Casado, Francisco Muñoz, Laura Jiménez-Ortega, Sabela Fondevila, Manuel Martín-Loeches.

Estudio Dos: Jose Sánchez-García, David Hernández-Gutiérrez, Pili Casado, Sabela Fondevila, Francisco Muñoz, Laura Jiménez-Ortega, Manuel Martín-Loeches.

*Estudio Tres**:* Jose Sánchez-García, Gema Esther Rodríguez, David Hernández-Gutiérrez, Pili Casado, Sabela Fondevila, , Laura Jiménez-Ortega, Francisco Muñoz Miguel Rubianes, Manuel Martín-Loeches.

** Publicado:* Sánchez-García, J., Espuny, J., Hernández-Gutiérrez, D., Casado, P., Muñoz, F., Jiménez-Ortega, L., ... & Martín-Loeches, M. (2019). Neurofunctional characterization of early prefrontal processes contributing to interpersonal guilt. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(5), 1192-1202.

*** Publicado en Brain Structure and Function*

Chapter 1. Introduction

1.1. Emotional processing: Basic and Social emotions

To say it bluntly, emotions have been considered ‘bad press’, although they don´t deserve that adjective. Throughout our human history, reason represents the top knowledge in mind hierarchy. From Plato, in which reason rides in a horse-drawn carriage (horses, emotions; reason, charioteer) to Descartes (“I think, therefore I am”), emotions seemed to take a minor and disturbing role in human motives and goals. Thanks to psychological and neuroscience research advances, emotions appear now to represent a central role in understanding human mind. As Charles Darwin proposed, if emotions exist, they must accomplish a positive function to meet survival needs.

In *The Expression of the Emotions in Man and Animals* book, Darwin opened the field to ongoing investigations reinforcing the communicative aspect of emotions. Fear, joy, sadness or anger play an adaptive role informing others about our intentions and feelings. They also seem to share some commonalities with other species, stemming from common origins.

Several definitions and classifications of emotions have been proposed thoroughly by field experts reaching to agreement in considering emotions as multidimensional processes.

Emotions are thus defined as a multifactorial episode which includes subjective feelings, cognitions, motivational tendencies, physiological processes and behavioral expressions in response to conscious or unconscious stimuli (Scherer, 2009).

Emotions, therefore, are neural dispositions for adaptive reactions to meaningful biological and/or cultural events. An emotion as fear, at the sight of a snake, prepares us to flee or fight increasing heart rate, blood pressure, cortisol release and blood flow to muscles. Sadness, due to a loss or withdrawal, inhibits our sympathetic system, alters our breathing capacity, depletes activation levels and bends inward our posture.

This triad of subjective feelings, physiological reactions and behaviors can be triggered by either external or internal stimuli such as memories and thoughts, consciously or out of awareness, and serve to survival and social needs.

Emotions, as rapid reactions to stimuli, are brief. They fade away in seconds. Moods, on the other hand, last longer periods and may not need an eliciting event to startle (Verduyn, et al. 2015).

Two main groups of classifications are used to describe emotions: dimensional and categorical:

Dimensional theories set a specific emotion to deploy on a two-dimension scale: valence and arousal (Barrett, Lewis, & Haviland-Jones, 2016). Valence can be positive or negative regarding its relative pleasantness or unpleasantness, whereas arousal relates to the intensity of the emotion. Dimensional theories display emotions in vector or circumplex models. Several researchers (Larsen et al., 1992; Remington et al. 2000; Yik et al. 2002) consider this two-dimensional model to be scarce. Alternatively, Fontaine et al., (2007) propose four dimensions to apprehend emotion complexity: evaluation-pleasantness, potency-control, activation-arousal and unpredictability. This model considers emotion as a six-component process: (a) appraisals of events, (b) psychophysiological changes, (c) motor expressions, (d) action tendencies, (e) subjective experiences, and (f) emotion regulation.

Categorical classifications discriminate between basic and social (also called complex or self-conscious) emotions as discrete entities.

Basic emotions are described as universal (showing presumably an unique feature of expression), innate, heritable and first to develop (Tracy and Randles, 20011).

Social emotions are cross-cultural modulated, require high cognitive skills to be processed, emerge later in development, are influenced by language and imply a sense of self (Lewis, 1992). This discrete classification is currently considered more as a continuum with several degrees of interaction than a strict domain classification. Basic emotions are best studied and understood due to their simplicity compared with social emotions and even by assuming shared commonalities with nonhuman animals, who provide valuable research and theoretical perspectives.

There is not a clear consensus between authors in the number of basic emotions. Proposals vary from 4, to up to 10. Surprise, anger, sadness, fear, joy and disgust are the 6 main basic emotions in Ekman's model (Ekman, Levenson, & Friesen, 1983). Fear, anger, joy, sadness, confidence, disgust, anticipation and surprise are the 8 basic emotions for Plutchik (2001).

Many feelings, the subjective experience of emotion, can apply to social emotion definition: envy, shame, resignation, jealous, hope, remorse, longing, guilt, pride and others. Main social emotions studied so far are guilt, shame, pride and embarrassment (Tagney, 2003). As expected, social emotions are even more difficult to categorize because they are sensitive to cross-cultural differences, they do not rely easily in external features as face or body expression, and depend more on social context and self-implication than basic emotions. Considering these classifications, it is clear that emotions are not fully understood. In the last decades several theories have emerged to explain the affective human domain.

1.2. Overview of neurobiological theories of emotion

One of the first emotional theories proposals was the James-Lang theory (Cannon, 1987). For William James, for instance, when we are escaping from a predator, the acts of running, sweating, breathing and other physiological changes is what produces the subjective feeling of fear. The experience of an emotion depends on the prior feedbacks from the brain and body responses. Carl Lange emphasized the role of the heart and brainstem nuclei controlling cardiac function in emotion generation.

The Cannon-Bard theory argued that a specific autonomic nervous system response, as those stated by James–Lang theory, cannot discriminate the enormous variety of emotional states (Dror 2014). Cannon and Bard even demonstrated that and injection of hormones as noradrenaline does not elicit a particular emotional state, similar to Gregorio Marañón's experiments back in 1920´s (being this consideration – mainly based in James' theory- crucial to the constructed emotional theory, which will be exposed below). James – Lang and Cannon-Bard theories rely in the neurobiological basis of emotions. They suggested that the body´s fight or flight response is coordinated via the autonomic nervous system, preparing the individual to deal with an emotional situation. Through transection experiments in decorticated animals and absence or presence of emotional reaction they highlighted the role of the diencephalon (thalamus and hypothalamus) to direct the information to the neocortex, generating emotional feelings, and to the periphery, then generating emotional (physiological) reactions.

Some years later, James Papez proposed the limbic system theory by including the cingulare gyrus to mediate emotional feelings (Papez, 1937). In the Papez's circuit theory, cortical and subcortical inputs are integrated in the cingulate gyrus, linking emotional feelings and body reactions. The stream of thought, or cortical route, carries input from the thalamus or memories to the cingulate, while the stream of feeling, or subcortical route, carries input from the hypothalamus and thalamus to the cingulate, reinforcing the cortical route.

In an essay to integrate Cannon-Bard, Papez and even Freud ideas, Paul MacLean proposed an evolutionary 3-steps brain model in which emotions were linked to the limbic system, now including amygdala and orbitofrontal cortex, then broadening the scope of Papez´s circuit (Cory, 2002). Despite its conceptualization, the triune brain, as named in MacLean's model, was openly surpassed by Joseph Ledoux's findings (Ledoux, 2012). Ledoux stated that there are no real independent anatomical criteria to split the brain in 3 separate functional areas. For instance, an area fully recognized as the amygdala shows cognitive and categorization functions, while the most rational brain area, the dorsolateral prefrontal cortex, is recognized to serve different emotional regulation functions. Triune brain hypothesis is too simplistic: rarely a brain area serves for a single function and vice versa. Ledoux therefore surpasses the isolated triune brain theory to enhance the role of the connectivity between amygdala and prefrontal cortex in emotional processing. Amygdala lesions in human brain disrupt fear conditioning (Labar et al, 1996), and frontal lobe lesions impair fear processing, showing no signals of anxiety, stress or fear behavior. Prefrontal–amygdala connective pathways play a lead role in emotional processing and regulation. Deepening the understanding of neocortex in emotional processes, several researchers (Ross 2008, Harmon-Jones, 2010) have added valuable contributions by studying the hemispheric specialization and its influence in understanding emotions. Prosody and frontal asymmetry are two examples of this hemispheric discrimination. Affective prosody of language, based on functional and anatomical evidence in patients with focal ischemic strokes, is a dominant and lateralized function of the right hemisphere (Ross, 2008). Greater left than right frontal cortical activity is associated to positive affect and approach motivation, which can be positive (e.g., enthusiasm) or negative (e.g., anger) in valence.

Based on William James' previous emotional theories, Damasio´s somatic marker theory (Bechara and Damasio, 2005) focuses in brain and body change states in response to a stimulus. This theory emphasizes the role of emotions in decision making processes. Physiological changes, sensations and feelings, considered as pleasant, unpleasant or neutral, adjoin instincts, previous experiences and desires participate to form a final decision. They may even overcome rational thoughts: for instance, when jumping from a bridge to the void despite three security belts and a 100% success probability, body sensations reject jumping. In psyco- or sociopats, somatic markers do not function properly, avoiding guilt feelings from misbehaviors, infringement, non-compliance, trespass or harm to others. When brain injuries affect the frontal lobes, as in the famous case of Phineas Gage, decision making processes are severely impaired and social misconduct compulsorily appears. Damasio proposes that the neural control of emotions implies limbic structures, prefrontal cortex, somatosensory cortex, basal ganglia, forebrain and brainstem. Rational decision making is linked to emotional processing, being the somatic markers a tool to integrate and accomplish in an adaptive way.

1.2.1 The Theory of constructed emotions – Feldman Barret

All the theories reviewed above are based in neurobiological signatures of emotions. They all see emotion as a fixed reaction to internal or external stimuli, shaped by evolutionary needs and linked to precise neuroanatomical and physiological areas and functions. Universality emerges from this perspective and each emotion has a "fingerprint". They assume these are responses to external or internal stimuli, deeply rooted in evolutionary schemes and somehow universal, if not fully in the expression of emotions, at least in their unique neural feature. They even show a sort of ethno- and anthropocentrism, leading to explicitly assume that even animals, from chimpanzees to rats, experience feelings such as fear, joy or sadness in the same manner as we humans do, which is an extraordinarily controversial assumption.

Therefore, a constructionist view is very relevant to understand the dynamics and interchangeable aspects of social emotions, which are prone to cultural and context influences difficult to cope with fixed neurobiological standards. The brain, on the other hand, and despite the counterintuitive self-awareness of its function in ourselves, is far from a stimulus–response device.

If the brain were exclusively a reactive organ, it would be extraordinarily slow, unadaptive and ineffective (Frith, 2014; Friston 2010). The brain is more a predictive than a reactive system to cope and manage real-time processes. By predicting physiological resources and coordinating motor activity, the brain enables learning, lessening error predictions and reinforcing adaptive responses. Simulations and predictions are the main activity of the brain (De Ridder, 2013; Clark 2013). In every moment a huge amount of visual, auditory, vestibular, interoceptive, somatosensory and olfactory information arrives to the brain, not passively but predictably constructed. This simulation stems from our memories, context and expectations (Hoeman et al 2020). We are not passive spectators, rather active constructors of perceptions, sensations and… emotions, far from a linear cause – effect vision.

This construction is counterintuitive. We tend to feel reality is out there, unified and equally shared by others, and that the brain captures it as clear as it is. Nevertheless, only some of the wavelengths of light are perceived by our eyes and a small percentage of it is converted into seen objects. Not all the frequencies of sound are perceived by our auditory cortex, but in that limited human range only some of the air pressure changes are heard as words, sounds or music. By the same token only some bodily changes are felt as emotions (Barret, 2012).

Emotions, in the theory of constructed emotions (TCE) (Barret, 2017), emerge when people make meaning out of sensory body input and outside world, using memory of past experiences. Emotions arise from concepts adaptive to the immediate environment and preparing the person to respond to sensory inputs in a way suitable to the situation (Barrett 2006b). The process arises from an initial top down prediction about the meaning of the sensory information and context being hold (Bar 2007). The error between the prediction and the sensory activity is then minimized to produce a unified conscious field (Friston 2010).

Different to the prevailing idea that some specific networks or areas are responsible for the emotional activity, the TCE assumes that the functionality of individual brain depends largely on the connecting network regions firing in sync (McIntosh 2004).

Usually, each emotion is assumed to produce neural activity in a specific brain region or network along a coordinated set of physiological responses, facial expressions, and voluntary behaviors, like freezing, fleeing, and fighting, emerging from the stimulus (Ekman, 1972) or after the evaluation of the stimulus (causal appraisal models; e.g., Ekman & Cordaro, 2011; Lazarus, 1991; Roseman, 1991). Each emotion is thought to produce a pattern compulsorily (Ekman, 1992) or probabilistically (Roseman, 2011). If small variations appear in this observed emotional pattern, they are accepted, but significant deviations are considered as errors, or caused by processes far from the emotional response. Nevertheless, heterogeneity within each emotion as well as similarities across different categories are the norm, not the exception (Barrett, 2006a, 2009). Indeed, heterogeneity within emotion categories has been well studied and reported, both within individuals and across cultures (Ceulemans, Kuppens, & Van

Mechelen, 2012; Hortensius, Schutter, & Harmon-Jones, 2012; Kuppens, Van Mechelen & Rijmen, 2008; Kuppens, Van Mechelen, Smits, De Boeck & Ceulemans, 2007; Nezlek & Kuppens, 2008; Stemmler, Aue, & Wacker, 2007). Variation is observed even when comparing patterns with exactly the same method and stimuli (Kragel & LaBar, 2013, Stephens, Christie, & Friedman, 2010). In turn, similarities may emerge from distinct emotional categories, such as fear and anger where neural responses are similar when elicited in a similar context of social threat (Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011).

Overall, heterogeneity in emotional response challenges classical views of emotion (basic, appraisal or functionalist theories). Data contradicts the idea that each emotion has an essence or fingerprint that other emotions do not share such as a unique expression or a pattern of autonomous nervous system activity, or a set of specific appraisals. Consequently, comparing emotional categories of humans from different cultures with other non-human animals is not consistent with this lack of homogeneity (Russell, 1991; Gendron et al., 2014a,b, 2015; Wierzbicka, 2014; Crivelli et al., 2016). Heterogeneity may be explained by degeneracy (a concept derived from genetics), as many neural networks and areas respond to one emotional category (Lindquist et al., 2012; Kragel and LaBar, 2015).

Therefore, for the TCE, emotions are predicted by concepts, rooted in our memories and interoceptive networks responding to the external environment. These concepts are embodied (Oosterwijk et al., 2012, 2015) and emerge as multimodal summaries represented in the default mode network and the salience network (Peelen et al., 2010; Skerry and Saxe, 2015; Raz et al., 2016).

The TCE defines all sensory input from the body, such as raw somatic, visceral, vascular, and motor cues (James 1884), arousal (Duffy 1957; Mandler 1975; 1990;

Schachter & Singer 1962), or affect (Harlow & Stagner 1932; Hunt 1941; Wundt 1897/1998), as core affect, to outline mental representations of body changes and arousal (pleasant or unpleasant) (Barrett & Bliss-Moreau 2009; Russell 2003; Russell & Barrett 1999). In fact, bodily cues are essential for consciousness (Craig 2009; Damasio 1999).

This core affect is attached to an object; described in the TCE as categorization. In this process, we make meaning of our core affective state by engaging in a conceptual categorization linking to an object or event. Concept categorization is therefore the process by which stored past experiences are used to create meaning from sensations emerging in the moment (Wilson-Mendenhall et al. 2011).

The brain predictive primary purpose, regulating physiological resources, learning and meeting body needs before they arise is much more efficient than reacting after events. The TCE, by core affect and categorization, includes a top down view of constructing emotions that embraces heterogeneity, cultural differences and context interdependence.

1.3. The social emotions

As seen before in the TCE, heterogeneity and variety match with a non-fixed view of emotions as fingerprints and open up a social view of emotions. Most people seem to be daily involved in behaviors leading to get social approbation or to avoid social rejection. We engage daily in behaviors for losing or winning status, work hard to achieve specific goals and behave in such ways our relations are kept safe and stable (Baumeister, Stillwell and Heatherton, 1994). Social emotions, as opposed to basic ones, imply complex elicitors from social environment. In order to understand social emotions, special attention should be payed to the sense of self, moral values and social relationships (Tracy, Robins and Tangney, 2007) .

1.3.1. The Self aspect

Basic emotions may not necessarily imply self-awareness, but social emotions must involve this process (Lewis 1989, Tangney 2002). This is proposed by several theories, even if it is highly controversial. Sometimes we experience guilt or shame without a complete awareness of the process. It is true that sometimes we may feel fear or anger in a self-evaluative process, but social emotions usually require a sense of self awareness.

This self-identity allows self-representations and self-evaluation to occur. Social emotions require comparisons between other(s) and oneself, and an evaluative outcome of this comparison leading to promote or avoid behavior, or to repair the harm.

Tracy and Robins (2004) proposed a model for social emotions deeply rooted in the sense of self. For a social emotion to emerge, an identity goal must appear first. If the evaluation is relevant to survival, a basic emotion arises; if, in turn, it is relevant to our self-identity, a social one arises.

A second step to generate this sort of emotions is through attentional focus on the self. If this event is important for me, self-representations are activated stablishing comparison with the external event.

The third step relies on “Is this important for me or who I want to be?”. This consideration is crucial because a social emotion is experienced if the person´s identity is in danger or heightened when an eliciting event is relevant to the self. Negative evaluations from others do not produce embarrassment or shame if they are about non relevant personal domains. Positive evaluations may not heighten our personal status if they do not match our self-evaluative appraisals.

The fourth step to set up a social emotion is based in congruency. When the event is relevant to our identity, the question is: “is it congruent or incongruent with the identity goals?” The outcome is a positive emotion when feeling congruent with the event, and negative in the opposite case. Our current self-representations may collide with our long term identity goals, as a sportsman failing a shot ruining his chances to win the tournament or a student failing an exam confronting the self-representation of being a smart student.

Attribution of a causal locus of the actual event leads to the fifth step: “Am I the cause of the event?”. If the person feels the cause is outside, an external control locus, a basic but not social emotion may emerge, as sadness of anger, as a result of uncontrollable circumstances. In fact, this is a common way to elicit basic emotions, as threats to survival needs are not ubiquitous in our daily lives, but we usually feel anger, fear or sadness, due to external attacks to our identity which we consider truly relevant. If the cause attribution is internal, related to the sense of self, a social emotion may emerge, leading to the sixth step: “Is it about me or about something I can´t control?” When we assume stable, uncontrollable and global attributions causing the event, shame or embarrassment may appear. When causes are internal, unstable, controllable and specific guilt arises. “Did I study wrongly or am I a worse student?” represents the two approaches focusing in a concrete action or a self-characteristic. On the other hand, if the result of the event is a positive evaluation, either authentic or hubristic pride appears. “Was it good because of my effort or due to my special genes and talent?”. Authentic pride relates to merit, will and effort to achieve a desired outcome. On the other hand, hubristic pride relies in the sense of having special skills or natural talent to which we may attribute our success (Holbrook, 2014).

This sixth step integrates self and emotions to understand the different thoughts, behaviors and feelings in the process and to set up clear distinctions between basic and social emotions and furthermore disentangle the social ones more clearly. Social emotions therefore imply an awareness of discrepancy between a current state and an identity goal resulting in negative or positive affect. The result of this affect leads to the appearance of the distinctive shame, embarrassment, guilt or pride emotions.

1.3.2. The Moral side of social emotions

The discrepancy between a current state and an identity goal implies a comparison between one´s self and its evaluation according to ethics, moral and social norms. Social emotions give rise to moral behaviors as they produce immediate reward or punishment. We evaluate by ethic the consequence to others of our actions and assess our own self behavior in relation to moral standards. This evaluation is not only consequential. We may easily forecast and anticipate shame, pride or guilt in response to a full range of potential situations. Three components are studied to support moral behavior: standards, cognitions and emotions.

Moral standards are diverse and culturally oriented, although most people feel it is wrong to cheat, lie, do harm, hate or murder others. According to Shweder et al., (1997), ethics may be classified in three main groups: divinity, autonomy and community.

Divinity shows a high cultural variability and usually relates to purity (even physical). In fact, appearance of moral gods, usually by enhancing intrapersonal guilt feelings, are intertwined to social complexity (Whitehouse et al., 2019).Autonomy approaches to justice, equity, human rights and no harm banning. Community relates to social norms such as duty, loyalty, honesty and respect.

Moral cognitions are a core set of ideas believed to foster behavior according to our moral standards. This assumption is surpassed by researchers as people may use rationalization and cognitive dissonance (Bandura, 1990) to perpetuate a wrong or harming behavior by diminishing the consequences, show contempt to the victims or downplaying psychological harm to others.

Moral emotions are the motivational power to practise goodness and avoid bad (Kroll and Ega, 2004), linking moral cognitions to moral standards. According to Haidt (2000, 2003) moral emotions "are linked to the interests or welfare either of society as a whole or at least of person other than the judge or agent". Not only negative or unpleasant social emotions serve to moral functions. Compassion, awe, gratitude, empathy oriented actions or forgiveness may contribute to the common good by enhancing motivation to foster equity, rights and shared social values.

1.3.3. The social aspect. Attributional theories and social valuation theories

As we have exposed in the previous paragraphs social emotions activation depend mainly on how the individual assesses herself (Tangney and Dearing 2002, Tracy and Robbins 2004). Attributional theories set emphasis on the intrapersonal nature of social emotions as they emerge when events relevant to self-identity goals are attributed to internal causes. If this event does not match with our goal, guilt arises. If the event is attributed to an uncontrollable or stable aspect of the self, shame emerges. If the event is congruent with our goal-oriented action and produces a pleasant feeling, pride is elicited. If the result of the positive event is attributed to a stable aspect of ourselves, relatively independent of our actions, hubris pride appears. Therefore, for this attributional view, social emotions depend on how the individual evaluates successes and mistakes.

According to Sznycer et al., (2018), social valuation theories rely on the interpersonal, instead of intrapersonal, nature of these emotions, as they favour adaptive solutions to social valuation situations such as attend, associate, trade personal welfare, or reject others. These adaptive solutions provide the opportunity to promote others by a) increasing their valuations of the self (pride); b) devaluate them if needed (shame), and c) reassure social balance when one did not put effort or attention on the welfare of a valuable other (guilt).

The interpersonal nature of social emotions relies on human behavior shaped by natural selection (Barrett, 2014).

In a world full of scarcity, predators, diseases, uncertainty, fellow members of the group provide assistance to each other to guarantee survival, reproduction and thrive. In social groups, attend or forget a partner, choose the right fellows for collaboration and reward, selecting reliable mates and value, and help or punish others' reputation are cognitive mechanisms who serve to solve social adaptive problems. These social adaptive problems shaped decisions to confer benefits or inflict costs, valuating others socially based on the value they provide to the group (Lieberman et al. 2007).

The interpersonal nature of social emotions can be estimated when our own internal valuation, as in self-esteem, reflects directly the degree others accept and include us as part of a social group. Self-esteem seems to be a construct where social status evaluation from others guide our behavior to match group needs and standards (Eisenberg et al. 2011). On the other hand, ostracism is felt when, despite our efforts or successes, we are excluded and other´s impression of our self is devalued (Williams, K.D. and Nida, S.A. 2011).

Social valuation theories focus on other´s assessments of individual attitudes and action leading to value or disvalue the self. When others reflect upon ourselves' motivations, ideas and actions, they assess us up or downward, influencing our inner states, creating consequently social emotions to cope with this interpersonal nature.

1.4. A social brain

As exposed in previous paragraphs, consistently with the TCE and interpersonal theories, in order for these social emotions to emerge there must be a set of specialized brain functions allowing it. Before considering us as rational actors, we are social ones. Brain human evolution is deeply rooted in our social structure and function (Dunbar, 1998). We are hardwired to connect (and be vulnerable) to others, not being possible to easily disentangle an emotional brain from a social one, even in the so-called basic emotions, as neuroimaging supports (Adolphs, 2003). Many human adaptations emerge from a social need: white sclera in the eyes, face anatomy or face expressions serve to social functions. Inside our minds, mentalizing seems to be a default mode in the human brain (Raichle, 2001). We constantly infer intentions and solutions to other´s actions. The social brain allows us to interact fluently with people. By predicting what others are going to do next, we may set up the best decision: the better our predictions are, the more the success possibilities. This type of behaviour appears in short term emotional states, long term dispositions, desires or beliefs: they all need a mental state based on predictions about social interactions (Frith, 2007).

Several brain regions play a key role in the social functions. Amygdala has a main role in recognizing the intensity of emotional expressions. Even prejudice or untrustworthy faces activate this region (Winston et al, 2002; Phelps et al, 2000).

Amygdala is involved in social cognition by associating a value to an object, including social ones (Ledoux, 2000).

The temporal poles seem to be related to our ability for mentalizing, as they seem to serve as a convergent region for our knowledge of individuals, actions and social situations. Thoughts and feelings most likely to occur in a social interaction, based in previous situations and the possibility to modify them, assign to the temporal poles (Damasio et al 2004).

A third neural structure of interest is the mirror neurons system. It comprises the same brain areas when observing another person emotional experience (or movement) and ourselves. It also works for pain and touch: somatosensory regions are active when we see someone suffering or receiving pain (Singer et al, 2004). Although the mirror neurons system tracks emotional states and intentions from others, it does not inform about their attitudes (Mitchell et al 2006).

The posterior superior temporal sulcus (pSTS) and the temporo parietal junction (TPJ) are also involved in the social brain. The pSTS predicts movements' trajectories, allowing to be aware of prediction errors in our estimations. It is important to note that in social interactions, eye movements of others inform about their intentions and actions. The pSTS functions even for non- biological movements (Kawawaki et al. 2006). The TPJ is activated when we evoke other persons' view perspective. This ability helps us to infer that other´s people knowledge about the world might be different to ours (Saxe and Kanwisher, 2003).

Medial prefrontal cortex (MPFC) activity is observed when people engage in mentalizing situations (Amodio and Frith, 2006), mainly in low demands tasks, such as understanding the behavior of characters (Gallagher et al. 2000), observing social interactions (Iacoboni et al. 2004), answering questions about attitudes (Mitchell et al

2005) or about current oneself's feelings (Oschner et al, 2004). It is not surprising the extensive role of the MPFC in social interactions, being one of the most expanded areas in brain evolution compared to non-human primates (Dunbar and Schultz, 2007).

The complexity of social living drives cognition and emotion in our brains. We infer from eyes, face expression and body movements the personality of others, and we imagine other´s perspective to take an immediate action. We store stories from the people we know, to behave socially in our next encounter. We represent others in us as well as we are in others' brains to keep social living according to norms, values and standards. A social brain handles all these processes. Research through brain techniques might add extraordinary value to a better understanding of emotions, specifically of social emotions.

1.5. Electrophysiological correlates of emotions

Traditionally, several techniques have been used to study emotions in the brain. Since 1995, different neuroimaging techniques such as PET of fMRI opened the field to study dynamic and current processes in the brain. The advantage of neuroimaging with these techniques relies in its spatial resolution. The study of emotions with fMRI allows to understand networks, areas and regions involved, therefore adding knowledge to the structure and nature of emotions (Barrett, 2006).

Despite its popularity and advantages, the temporal resolution of PET or fMRI is very limited. In turn, electroencephalography (EEG) is one of the most valuable techniques to study the temporal processes in the brain. This signal reflects the summed activity of postsynaptic potentials of pyramidal neurons that fire synchronously while processing information (Peterson *et al*., 1995). By placing electrodes over the scalp, EEG recollects information in real time.

Back in 1875, Richard Caton, a British physiologist, wrote in "The electric Currents of the Brain" (Caton, 1875) about the different currents observed in rabbits and monkeys when placing electrodes on two different points of the external surface of the skull, opening the field to study brain electrical activity.

The birth of EEG, due to the technological advances, relies in Hans Berger (Berger, 1929). He was also the first to describe the different brain waves, such as alpha and the alpha blockage appearing when the subject opened the eyes. He is also recognized as a pioneer in describing EEG alterations in brain diseases as epilepsy. However, to study cognitive and emotional processes and behaviour along its underlying neural substrates, raw EEG cannot be used. Its activity includes random oscillations, and sensory, motor and cognitive brain fluctuations lacking the needed specificity of a precise cognitive process.

This limitation is surpassed by the Event Related Potentials (ERP) technique. ERP are very small voltages generated in the brain structures in response to specific events or stimuli (Blackwood and Muir, 1990). They are time-locked and phase-locked to sensory, motor or cognitive events and, by averaging the EEG activity of several single trials associated to an event, signal-to-noise ratio is increased (Luck, 2014). ERP waveforms consist of a series of positive and negative voltage deflections, which are related to a set of underlying components (Luck et al., 2014). ERP components in humans are classified in 2 categories. The early waves ('sensory' or 'exogenous'), components peaking roughly within the first 100 milliseconds after stimulus and depending mainly on the physical parameters of the stimulus. The late waves (cognitive' or 'endogenous'), occurring after 100 ms post stimulus, reflect the way the subject evaluates the stimulus. The waveforms are described according to latency and amplitude (Sur and Sinha, 2009), though they are usually displayed through a topographic map, a 2D or 3D figure representing the voltages across the electrode locations on a specific time (Hari & Puce, 2017). Therefore, each component can be defined by its amplitude, latency and topography, relative to a stage of cognitive or emotional process. ERP components may have positive or negative voltage at different latencies, so they may help to know the underlying cognitive processes related to specific events linked to experimental manipulation (Luck, 2014).

Several ERP components have been thoroughly studied (Sur and Sinha, 2009), as described in the following.

The N100 (also called "vertex potential") is a negative deflection peaking between 90 and 200 ms after stimulus onset. It is observed across sensory, motor, cognitive, and behavioral tasks, reflecting GABA-mediated inhibition processes.

The P200 or P2 is a positive deflection peaking around 100-250 ms after stimulus onset. Current evidence suggests that the P2 component reflects attentional processes.

The N200 is a negative deflection peaking at about 200 msec after presentation of the stimulus and can elicited by any discriminable change (Näätänen and Tiitinen, 1997).

The P300 dates back to Sutton et al. in 1965 and reflects processes involved in stimulus categorization, as it discriminates one event from another. Greater attention produces larger P3 waves. It is usually elicited using the oddball paradigm with a latency of 250 to 500 ms, and is measured mainly by the parietal lobe electrodes. The P300 signal is used as a way to assess cognitive functions in decision making.

The N400 is a negative wave appearing 300–600 ms post-stimulus (Kutas and Hillyard, 1980) and linked to semantic incongruity. The more incongruent a stimulus is, the larger the N400. Its activity peaks over centro-parietal electrodes. The N400 is multimodal, responding to visual, auditory, language and faces stimuli.

The P600 is a positive wave appearing 300–600 ms post-stimulus (Kutas and Hillyard, 1980) mainly used in language processing. Its effect occurs to sentences that contain some sort of syntactic violations (Osterhout and Holcomb, 1992).

As opposed to other techniques such as fMRI, ERP offers a high temporal resolution of milliseconds of the underlying cognitive and emotional operations occurring within the brain. By scientifically defining the experimental conditions, the ERP can discern the time course of the processing involved in each condition.
In addition to ERP, there are several methods (such as sLORETA, VARETA, S-MAP, ST-MAP, Backus-Gilbert, LAURA, SSLOFO. ALF, BESA, MUSIC or FINES) to localize electrical activity in different brain areas from EEG or ERP recording (Grech et al., 2008). They are based in inverse problem solution, a way to convert measurement into information about EEG observed signals. We selected Low-resolution brain electromagnetic tomography (LORETA) as EEG-based neuro-imaging technique in our experiments. LORETA computes a unique 3D electrical source distribution of 2394 voxels of 7x7x7 mm of neural activity in brain grey matter. It selects from all possible inverse solutions the smoothest one as neighboring neurons are active synchronously during a task. The solution space provided by LORETA is restricted to cortical gray matter and hippocampus. Subcortical structures are not accessible by this mean, being one of the main limitations of this technique. Several constraints are being exposed by authors (Kincses et al. 1999, Michel et al. 1999) along with its validation (Pascual-Marqui et al. 2002, Phillips et al. 2002, Yao et al. 2001). Despite its limitations, LORETA analysis may reinforce ERP temporal results and provide a source

localization estimation of regions associated to brain functionalities, based in Brodmann's areas mapping.

1.6. The neuroanatomical correlates of social emotions

Social emotions hold several shared characteristics despite individual differences. They all serve to keep behavior under the constraint of social norms (Beer and Keltner, 2004; Tangney and Fisher, 1995). Feelings of embarrasment, guilt or shame are painful enough, leading people to avoid experiencing them again. Pride, on the other hand, feels so pleasant, that an individual is drawn to repeat the behavior. In order to feel a social emotion, individuals need different levels of self-awareness: a) own self, b) other(s) judging that self, c) social norms right or wrong for that self-knowledge. These three levels point to specific brain structures involved in social emotions.

a. Self -encoding processes involve the activity of medial prefrontal cortex (Fossati et al, 2003, Gilligan and Farah, 2005). Different processes are included in self encoding, such as self vs others judgments, own face vs others faces, or positive vs negative self-judgments, being all linked to frontal lobes' activity comparing actual and self behavior to others or social behavior. Damage to the medial prefrontal cortex or the orbitofrontal cortex show some sort of anosognosia or lack of insight to our own social behavior (Beer et al, 2006) and impairment of social behavior (Clark et al, 2008; Anderson et al, 1999).

b. Self of others is a crucial process in social emotions to emerge as one needs the awareness of other(s) self to feel they are judging yourself. That is, by other means, mentalizing, the capacity to infer other´s people minds and emotions.

Emotional inference of others is associated with amygdala activity, as several studies report (Sato et al, 1994, Damasio and Damasio, 1994, Shaw et al. 2004, Stone et al. 2003)

Mental inferences about others that enables inferences to be made about goals, beliefs or moral issues are associated to medial prefrontal cortex, posterior cingulate/precuneus, and temporo-parietal junction (TPJ) activation (Denny et al.,2012; Van Overwalle and Baetens, 2009; Lombardo et al 2010).

Social emotions require that and individual is aware of other´s evaluating self behavior and positive or negative emotional reactions to that interaction.

c. Knowledge of social norms are a crucial part to elicit social emotions. Social norms are the rules others and self judge to ascribe or refrain from certain behaviors. Social brain networks involve the activation of several areas such as the amygdala, the somatosensory cortices, the temporal lobes and the frontal lobes (Frith, 2007; Johnson et al. 2005).

1.7. The neurophysiological study of social emotions in the lab

As we may infer from previous arguments the vast variety of emotions we experience opens up the field of scientific study of emotions so much that it is difficult to set its boundaries, definitions and define a scope or an appropriate approach. We do not even have yet an agreed definition of what an emotion is. Research often lies in either the evaluative process, the physiological changes, the facial or bodily expressions, the subjective experience, the mental processes, the behavioral dispositions or the brain correlates.

Neurophysiological studies of emotions are usually focused on few categories or dimensions (Ekman, 2016; Rotttengerg et. al, 2007, Russell, 2009). Koide-Majima et al.

(2020) selected up to 80 emotion categories to study emotions and broadening classic neuroimaging studies which are usually limited to a few emotional representations in the brain. 25 representative BOLD clusters of emotions, representing the 65% of the total variance, were founded. A subtantially large across-subject variability in the number of dimensions emerging from brain representation (from 18 to 36) was found, possibly reflecting the emotional granularity across subjects as proposed by the TCE. The emotions founded in this semantic space match a cortical gradient of representation, from sensory to abstract information, involving superior temporal area, inferior parietal area, precuneus and post central area. Future studies are needed to be performed to understand better these cortical representations of emotions. Furthermore, subcortical areas are known to play a key role in emotional processing (Damasio et al. 2000; Costafreda et al., 2008) and neuroelectrophysiological techniques advances may provide, in a future, a way to assess the complex study of social emotions.

In our research our aim is to study the social emotions of guilt, shame and pride under ecological conditions in the laboratory by concurrently measuring ERP and neural source analysis (LORETA). These are the most relevant, prominent and well known social emotions and share a crucial role in our interactions and behaviors. As we uncover in the coming chapters, our approach tries to resemble, as far as possible, a social emotion as it emerges in our daily environments.

Chapter 2. Objectives and hypothesis

From previous literature cited in chapter 1, the main objectives and hypothesis of this thesis are:

General objectives

- Highlight the importance of social contexts in social emotions.
- Explore the effects of social interactions on emotional processing.
- Study the temporal and spatial neural correlates of social emotions in ecological settings.
- Replicate and extent previous findings relative to the neural correlates of social emotions.
- Design ecological paradigms to study social emotions in the lab.

Specific objectives and hypotheses

The objectives and hypothesis are organized around three studies, as follows:

• Study One: investigates the characteristics of ERP fluctuations elicited by guilt emotions, broadening previous research (Leng et al., 2016) by: a) analyzing neuroanatomical sources; b) establishing relations between emotional and personality variables, and c) exploring a different cultural sample (other than the Asiatic in Leng et al., 2016). Participants performed a dot estimation task together with a partner, in which four experimental conditions occurred along 320 trials: partner wrong, self wrong, both wrong and both right situations. We hypothesize that interpersonal guilt could be clearly differenciated from the other (non guilt conditions) involving a first, fast and early brain process followed by a late one linked to self analysis and reappraisal process. We estimate there will not be substantial differences between our western sample and the eastern one (Leng et al., 2016). Self wrong condition is expected to elicit maximal feelings of guilt and the corresponding ERP fluctuations. We expect a frontal early negativity and a posterior late positivity, related to different stages in the analysis of the social situation.

• Study Two: investigates the neural correlates of shame extending a previous paradigm by Zhu et al (2017) by changing feedback outcome in an advice-decision game. Participants performed again a dot estimation task though, through different settings. Specifically, the participant plays the role of adviser giving place to the capacity to give advice and take responsibility. In this manner, shame may emerge as a function of the final outcome by the partner. We hypothesize that shame could be disentangle from guilt based in this particular setting of the experimental design and again involving a first, fast and early brain process in which a social shame situation is valued, followed by a late one linked to a self assessment process. An early distinction of shame compared to other conditions is expected in the ERP fluctuations, characterizing this emotion in this respect. We expect centro-parietal negativities reflecting the integration of emotional information in relation to social validation situations.

• Study Three: investigates the neural correlates of shame and pride at the same experiment, in a novel (original) setting in which the participant competes with 3 other online students and increasing social exposure. We hypothesize that shame and pride could be studied live, as opposed social emotions, ranging from the loss to gain in social status and again involving an early brain process in which shame or pride are valued, followed by a late one linked to a self assessment and self relevance processes. We expect early negativities and late positivities to the two emotions, linked to autobiographical (shame) and hedonic value (pride) areas such as the precuneus/PCC (shame) and the mPFC (pride).

Chapter 3. Study one. Interpersonal guilt.

3.1. Introduction

The study of interpersonal guilt has recently gained interest due to its social significance. Social emotions influence our daily lives and guilt plays a major role in bringing the group together and rejoining bonds. Guilt, along with shame, embarrassment and pride (Lewis, 2000) is one of the social and self-conscious emotions that appears when someone feels responsible for harming or negatively affecting another person when a different action might have avoided the situation. As a moral and prosocial emotion, it protects social relationships by punishing interpersonal wrongdoings, and promoting and rebalancing behavior (Haidt, 2003; Amodio et al., 2007). Moreover, it also acts as a marker for future partner behaviors in a clear prosocial focus as it prevents people from committing wrongful actions (Chang et al., 2011). Its absence is manifest in psychopaths, who behave abnormally and immorally and feel no remorse about those they hurt (Kiehl, 2006). Guilt is based on self-agency and, when our actions affect another individual, empathy is required by adopting the perspective of the affected individual(s) (Hoffman, 2000). Two main types of guilt have been studied: deontological or intrapsychic (Monteith, 1993; Wertheim & Schwartz, 1983) and altruistic or interpersonal (Baumeister et al., 1994; Tangney & Dearing, 2002), the latter being the aim of the present study.

Neuroimaging studies (namely, using functional magnetic resonance imaging, fMRI) have found several brain structures consistently involved in interpersonal guilt. These areas relate to highly studied networks involving social cognition, such as the medial prefrontal cortex –including both the ventromedial and dorsomedial, and the anterior cingulate cortex-, the orbitofrontal cortex, temporal poles, temporal-parietal junction, precuneus, posterior cingulate, posterior superior temporal sulcus, and insular

cortex (Shin et al., 2000; Takahashi et al., 2004; Kédia et al., 2008; Morey et al., 2012; Zahn et al., 2009; Basile et al., 2011; Mclatchie et al., 2016; Bastin et al., 2016).
The information provided by most of these studies, however, exhibits two main limitations. On the one hand, the most extensively used paradigm has been the elicitation of feelings of guilt by means of instructing participants to imagine or remember personal experiences in which they felt guilty, or by presenting participants with hypothetical scenarios of guilt. This approach is far from reminiscent of a situation of natural and real moral conflict, guilt being mediated by episodic memory and imagination and, hence, arguably lacking validity and intensity. As a result, several of the reported brain regions might not be primarily related to the feeling of guilt.

This limitation was largely overcome in Yu et al. (2014), who used a paradigm to elicit interpersonal guilt in the neuroimaging setup, largely used in behavioral studies (e.g., Fliessbach et al., 2007). In this paradigm, two participants perform a dot estimation task with monetary rewards. In Yu et al. (2014), only one of the participants was recorded by the MRI scanner, and there were four experimental conditions: when both participants were right in the dot estimation task (BR), the partner of the recorded person did not receive a pain stimulus, the opposite being the case in the other 3 conditions. These were: only the recorded participant was wrong (self-wrong, SW), only the partner was wrong (PW), or both were wrong (BW). After the dot estimation, a feedback screen indicated the outcome and responsibility of their action. In these paradigms, interpersonal guilt is more intense in the SW condition, as a participant's mistake is the direct cause of the partner's pain (De Hooge et al., 2011; Nelissen, 2009). The main finding of Yu et al. (2014) was that the brain area mainly involved in interpersonal guilt is the anterior cingulate cortex (aCC), with a secondary implication of bilateral anterior insula.

A second limitation of these studies is that the neuroimaging techniques employed do not provide information on the timing and dynamics of the processes involved, but a still picture of possibly involved brain regions. This shortcoming may be overcome by using event-related brain electrical potentials (ERP). Although relatively limited in their spatial resolution, ERP provide a resolution in milliseconds. To our knowledge, only two studies have addressed interpersonal guilt with ERP, using the dot estimation task paradigm described above. In the study by Leng et al. (2016), instead of pain stimulation to the partner, the authors used the variant that both participants could earn money in the BR condition, losing money instead in the other three situations. Their results indicated that the effect of guilt was initiated at about 350 ms and peaked round 500-600 ms after the feedback onset, with larger amplitudes for a frontal negativity on the high-guilt condition (SW) compared with the low-guilt condition (BW). The SW–BW contrast also showed a Late Positive Complex (LPC) 500-800 ms after feedback onset with a centroparietal distribution, probably related to sustained processes of reanalysis, evaluation and memory encoding (Kissler et al., 2007). The LPC was nevertheless not replicated in the second of the two experiments in Leng et al. (2016), while frontal negativity appeared to be a robust finding (in their words, "an index of interpersonal guilt"), proposed to reflect interactions of self-reflection, condemnation, and negative emotion. The study by Zhu et al. (2018) used instead a variant in which the economic outcome for the participant was fixed, while that of the partner was contingent exclusively on the rightness or wrongness of her own performance, with the recorded participant playing a limited role as advisor. With this paradigm, only a distinctive ERP pattern for shame could be reported, while guilt was hard to discriminate either from happiness or from shame –depending on the time interval considered.

Overall, the number of studies on the neural foundations of interpersonal guilt is still insufficient, particularly those using ERP. The main aim of the current study is to contribute doubly to this research by replicating and further extending the ERP study of Leng et al. 2016, which needs replication to settle the issue of the actual robustness of a possible electrophysiological index of guilt, given the potential value of such a measure. Furthermore, our study will also extend this data in several meaningful manners.

First, the study of Leng et al. was conducted on a group of Chinese participants, while the present study works with a European (Spanish) population. This is particularly interesting, given that cultural differences have been proven to exist in the experience of emotions, at least when comparing eastern and western societies (e.g., Lim et al., 2016), and this is particularly significant for complex social emotions such as guilt (Walbott & Scherer, 1995; Anolli & Pascucci, 2005). It could be the case that the frontal negativity reported by Leng et al. might show a different latency, amplitude, or topography in the present study, or it may even be absent. Cultural differences of the samples would be a main underlying reason if replication were not achieved.

Second, a neural source analysis on frontal negativity will be performed, in case it emerges. In addition to the high temporal resolution, ERP can provide an estimation of the current distribution from the scalp's electrical potentials. In the present study, Low Resolution Electromagnetic Tomography Analysis (LORETA) will be used, one of the most established and widely used reconstruction algorithms (Pascual-Marqui et al., 2002). It enables an estimation to be made of distributed activity throughout the brain by decomposing the overlapping EEG voltage patterns into their underlying sources and positioning them within the brain. Leng et al did not perform this kind of assessment. In view of the fMRI study by Yu et al. (2014) with an analogous paradigm, it appears plausible that guilt-related frontal negativity originates in the aCC. A LORETA estimation in this regard would strongly support this possibility.

Third, whether the frontal negativity actually reflects feelings of interpersonal guilt and its relationship to emotional variables must be demonstrated. This approach is absent in Leng et al. (2016). The authors discarded that frontal negativity was reflecting fluctuations related to conflict or error monitoring because it emerged solely in the condition in which interpersonal guilt would be more intensive, SW, and was absent in the other cases of conflict or self-committed errors, i.e., BW and PW. These last two appeared very similar to each other. This is a valid but possibly partial argument with regard to the specificity of the processes actually reflected by frontal negativity. Guilt is a complex emotion, conceivably involving several brain areas and networks, as the neuroimaging studies reviewed above would suggest, with each node arguably contributing differently to the eventual emotional feeling. In this study, we wanted to better specify the significance of frontal negativity or other ERP fluctuations to interpersonal guilt by relating them to a series of variables linked to emotions and social cognition.

A first important variable in this regard is the degree of interpersonal guilt actually reported by the participants throughout the experiment. Following procedures similar to those used in Leng et al. (2016) or Yu et al. (2014), participants will be asked whether they felt guilt after each trial. The number of trials in which a positive response is delivered is expected to be highest in the SW condition. The number of responses in this condition will be correlated to the amplitude of the frontal negativity or other ERP fluctuations found to be specific of the SW condition. Second, feeling bad for personal actions that harm others requires empathy (Morey et al, 2012), and for that reason we also wanted to see whether frontal negativity (or other ERP fluctuations associated with guilt) can be related to differences in empathy. This was measured with TECA (a test of affective and cognitive empathy adapted to the Spanish population, Fernández-Pinto, et al., 2008). The test assesses four different empathy factors; two of them are cognitive (perspective taking and emotional understanding) and two emotional (empathic stress and empathic joy). Third, increased anxiety or arousal is linked to distressing feelings of guilt (Etxebarria & Apodaca, 2008). In this regard, skin conductance response (SCR) is widely used for measuring stress, anxiety and emotional reactions (Liu et al, 2014). Therefore, we measured SCR throughout the experiment in each participant in order to estimate whether SW trials increased arousal as compared to the other conditions, and its possible relationship to ERP fluctuations linked to guilt. Furthermore, and along these same lines, we used the STAI test (Skapinakis, 2014) to assess state- and trait-anxiety levels of our participants.

Finally, our sample was composed exclusively of female participants. The purpose is to homogenize the sample and reduce sex-related variability in psychophysiological measures, as in Amodio et al. (2007). Guilt and empathy seem stronger in females than in males (Etxebarria et al. 2009; Silfver & Helkama, 2007), while they are functionally equivalent for both women and men (Monteith, 1993; Monteith et al., 2002). Working with females could increase the power of possible ERP fluctuations related to interpersonal guilt while assessing our results, which would also be applicable to males.

3.2. Material and methods

3.2.1. Participants

Twenty-four pairs of best friends (all undergraduate females) participated in the experiment. Only one member of each pair (n=24) was randomly selected to be EEG recorded and the results presented below will refer exclusively to these selected participants. Their ages varied between 18 and 25 years, (mean age = 19.9; SD=1.54). All were right handed, ranging from 50 to 100 (M=84,44%) according to the Edinburgh Handedness Inventory (Oldfield, 1971). Participants reported no history of psychiatric or neurological illnesses and provided written informed consent before the experiment. The study was performed in accordance with the Declaration of Helsinki and approved by the ethics committee of the Hospital Clínico San Carlos (Madrid). Participants were reimbursed for taking part in the experiment.

3.2.2. Procedure

Before the ERP experiment, the participating couple individually completed two different questionnaires: STAI (Spielberger et al., 1983) and TECA (Fernández-Pinto, et al., 2008). Thereafter, one of the participants was invited to enter the EEG shielded chamber room while her partner sat in an adjacent room. Both received the same instructions and completed the task on a computer connected to two LCD screens. Stimuli were presented on a screen placed 65 cm away from their eyes.

Each trial (Fig. 1) began with a fixation cross appearing at the center of the screen for 500 ms, followed by a black screen with twenty white dots (dot size = 3x3 pixels), randomly displayed, in a 300x300 pixels frame around the center of the monitor (x=0, y=0) to minimize ocular movements. Participants were instructed to estimate the number of dots on the screen. The dot stimuli were displayed for up to 1500 ms. After the dots disappeared, a number (19, 20 or 21, randomly chosen) appeared on the screen for 3000 ms. Meanwhile, participants had to press one of two buttons as soon as possible using the index or middle finger, respectively (counterbalanced across sessions and participants), to indicate whether the number of dots estimated was larger or smaller than the number shown. The response hand was also counterbalanced across participants, while the other hand was used to measure the SCR.

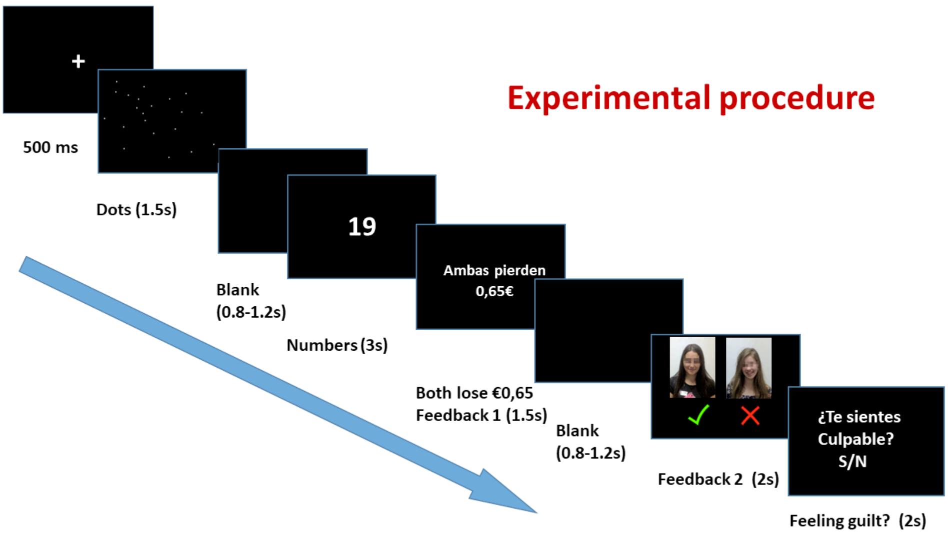


Figure 3.1. Experimental procedure. A fixed cross appears for 500 ms at the center of the screen before the onset of random dots for 1500 ms, where participants estimate the number. After a blank screen, 19, 20 or 21 appears for 3 s while participants have to choose whether the actual number was greater or lesser. Feedback 1 screen appears for 1.5 s assessing money loss or reward. After another blank screen, Feedback 2 with portraits of the participants appears for 2 s with a tick or a cross for individual performance. A "do you feel guilty?" question (yes/no) ends a trial.

After the presentation of a blank screen for 800, 1000 or 1200 ms depending on the trial (randomly chosen to avoid habituation), a number was displayed on the screen to indicate how much money the participants earned (or lost) for their performance (feedback 1). This feedback 1 screen was displayed for 1500 ms and followed by a blank screen. Then the feedback of the performance of each player was presented for 2000 ms with either a green tick or a red cross below each participant's picture, indicating a right or wrong outcome (feedback 2). After completing a trial, a yes-no question appeared on the screen for 2000 ms, and participants were asked to indicate whether they felt guilty upon seeing the outcome of their cooperation performance. The position of the response alternatives (‘yes’, ‘no’) at the left/right of the screen was counterbalanced across participants. Importantly, participants received the instruction that they could leave the question unanswered if they were uncertain about their feelings of guilt. Accordingly, this is not a forced choice task. We only asked our participants whether they actually felt guilt, following the procedure by Leng et al. (2016). Other studies (e.g., Zhu et al., 2017) have included a choice between alternative emotions that could play some role in the corresponding condition. In our case, shame might also be present in the SW condition because of bad performance of the participant, and therefore shame and guilt would be ideal choice alternatives. However, when we performed a pre-study (with n=15) exploring this procedure, most of the trials were left unanswered. Shame and guilt seem highly overlapping and difficult to unravel at the subjective level (Giner-Sorolla et al., 2011). Although it is admissible that shame is present to a similar degree in both SW and BW conditions, both conditions would noticeably differ in the degree of guilt experienced. Asking only for guilt therefore avoided participants’ conflict and indecision (i.e., non-responded trials), while the ratings relative to feelings of guilt in each condition (cf. the Results section) were largely consistent with the expected outcomes (see also below).

All performance feedbacks were predetermined by the experimenter as follows: there were a total of 320 trials consisting of 80 PW (partner wrong), 80 SW (self wrong), 80 BW (both wrong) and 80 BR (both right) trials, so earning trials were one out of every four. The participants were unaware of this manipulation. Indeed, the number of dots remained constant at 20 on every trial although their spatial distribution varied randomly. Participants were told to perform a cooperative task with their partner;

receiving additional payment only if both performed successfully. Both participants lose €0.65 when one or both responded incorrectly. Only when both participants answered correctly did both earn €0.65 each. The last situation occurred in only ¼ of the trials, so during the complete task they would inevitably lose money. Accordingly, participants were told that they started with an initial payment of €30 for participation, and that their outcome would depend on their performance (i.e. 65 cents would be discounted from this amount in a wrong trial and added in a correct one). As the task was highly demanding, we assumed that they could not be mentally updating on the exact amount of money lost while doing it. They probably felt that they were losing some money, but not exactly how much. At the end of the study, participants were told that considering their incorrect and correct responses, together with compensation for participating in the study, they would receive a final total amount of 20 EUR, but they were never informed of the exact amount of money that they had lost. Overall, the whole experiment lasted about 55 min.

We wanted to explore whether the difference between SW and BW conditions involves other emotions apart from guilt by conducting a behavioral study with a sample of 18 participants (other than those in the ERP main study). Their ages also varied between 18 and 25 years (mean age = 22.5; SD=2.21), and all were right handed, ranging from 50 to 100 (M=89,47%) according to the Edinburgh Handedness Inventory (Oldfield, 1971). They performed 60 trials of the same task as in the ERP study but rated their feelings of happiness, guilt, shame, frustration and pride in a 0-7 Likert scale after each trial. Guilt, shame and frustration were considered possible candidates to differ between SW and BW conditions, while happiness and pride were used as controls as well as possibly involved in PW and BR conditions. On the other hand, Yu et al. (2014) already proved that the conditions of interest do not differ in the negative feelings of fear or anger. The participants were told that in order to avoid confusing guilt and shame (see above), guilt was defined as the unpleasant feeling subsequent to one's own action harming others, while shame refers to the unpleasant feeling due to a negative evaluation of oneself by others. The main results appear in Table 1. An ANOVA revealed that the main effects of Emotion [$F(4,68) = 14.38$, $p<.001$], Condition [$F(3,51) = 8.94$, $p<.001$] and interaction "Emotion x Condition" [$F(12,204) = 37.6$, $p<.001$] were significant. The results exhibited the following pattern (all ps Bonferroni corrected). Frustration appeared the dominant feeling in the two conditions of interest (BW and SW), with a small but significant increase from BW to SW [$T(17)=3.7$, $p=.007$]. In this comparison, there was also a significant increase of the feeling of guilt [$T(17)=7.4$, $p<.001$], this increment being of larger magnitude than that

	Happiness	Guilt	Shame	Frustration	Pride
BR	4.6 - 1.6	0.3 - 0.59	0.2 - 0.3	1.3 - 1.4	3.8 - 2
PW	2.4 - 1.9	0.5 - 0.7	0.3 - 0.5	2.6 - 1.8	2.1 - 1.9

of frustration (1.1 vs. 0.5 points, respectively). In turn, shame displayed lower values than guilt or frustration in either condition, and although it appeared to increase from BW to SW, this was not significant [$T(17)=3$, $p=.06$]. Importantly, the difference between reported frustration and guilt in the SW condition was not significant [$T(17)=1.6$, $p=.8$], which suggests that both were equally dominant in this condition.

SW	0.4 - 0.5	3.7 - 1.8	2.2 - 1.72	4.2 - 1.9	0.2 - 0.5
BW	0.4 - 0.6	2.6 - 1.7	1.6 - 1.5	3.7 - 1.9	0.2 - 0.6

Table 1. Mean and standard deviation values of reported feelings for each condition in the behavioral study

3.2.3. Electrophysiological recording and analysis

EEG was recorded from 59 scalp electrodes mounted on an electrode cap (EasyCap), following the 10/20 International System. Bipolar vertical and horizontal EOGs were recorded to monitor blinks and horizontal eye movements. During recording, all scalp electrodes as well as one electrode at the left mastoid, were originally referenced to one electrode at the right mastoid; offline, they were re-referenced to the average of the right and left mastoids. The impedance of all electrodes was kept below 5 kΩ. EEG data were analyzed with Brain Vision Analyzer® software. Raw data were filtered online with a band-pass from 0.01 to 100 Hz and sampled at 250 Hz; they were digitally filtered offline from 0.1 to 30 Hz.

The continuous EEG was segmented into 1200-ms epochs, starting 200 ms before the feedback 2 screen onset. Eye movements were corrected using Independent Component Analysis (ICA, Makeig et al., 2000) as implemented in BrainVision Analyzer®. Remaining artefacts were further removed by a semi-automatic rejection procedure, eliminating epochs exceeding ± 100 μV in any of the channels. The mean rejection rate of epochs was 4.8%, and there were no significant differences between conditions [$F(3,69) = 1.98$; $p = 0.142$; $\eta p2 = 0.079$]. Mean rejected epochs (and SDs) for conditions were: BR (Mean = 3.69; SD = 4.768), PW (Mean = 5.29; SD = 6.931), SW (Mean = 5.50; SD = 8.119) and BW (Mean = 4.75; SD = 6.732).

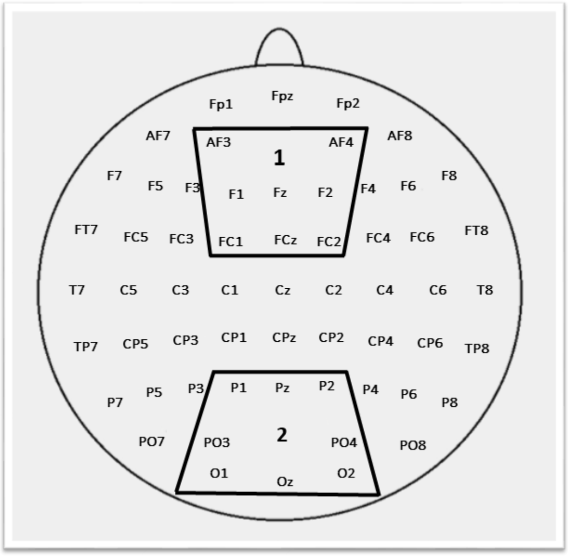


Figure 3.2. Location of the electrodes covered by the two main regions of interest (ROIs); a frontal ROI (ROI1) and a parieto-occipital one (ROI2).

Statistical analyses were computed with IBM SPSS Statistics (Version 22). An overall repeated measures ANOVA was first performed including two factors: Electrode (59 levels) and Condition (4 levels: BR, PW, SW, BW). Amplitude was measured as the mean amplitude within a particular time interval (from 200 to 1000 ms in 100ms-wide windows). To avoid a loss of statistical power, based on ERPs and components found, second ANOVAs were performed in two regions of interest (ROIs) (Fig. 2): One frontal ROI that included the electrodes AF3, AF4, F1, Fz, F2, Fc1, Fcz and Fc2 (ROI1), and one parieto-occipital ROI, that included P1, Pz, P2, PO3, PO4, O1, Oz and O2 (ROI2). For the aforementioned analyses, specific time windows (350-450 and 750-950 ms) were selected as based upon visual inspection of the main ERP components. The Geisser–Greenhouse correction for non-sphericity was applied when necessary (Greenhouse and Geisser, 1959). Post-hoc tests were corrected with the Bonferroni procedure.

3.2.4. SCR recording and analyses

The skin conductance response (SCR) was monitored throughout the experiment in each participant. An 8-channel Multibox polygraph (Brain Products, Munich) was connected to the index and middle fingers of the non-responding hand in the experimental task. SCR results were analyzed with Ledalab software (Benedek & Kaernbach, 2010a). Downsampling was applied to the original data to 125 Hz, and pre-processed by adaptive smoothing and filter selection. Following these procedures, a composition decomposition analysis (CDA) was performed to separate from the raw SCR both the tonic and the phasic signals (Benedek & Kaernbach, 2010b). Phasic data were considered with a minimum SCR amplitude of 0.01 µS; the examining window covered from 1 to 5 s after the onset of feedback 2. Individual measures of SCR in µS were separately averaged for each participant. Six out of 24 participants were discarded due to artefacts in SCR signals. Phasic mean activity data were standardized within subjects to facilitate the comparison of SCR means across the conditions: SW, BW, PW, BR.

3.2.5. Source generator analyses

LORETA software, an algorithm based on an inverse problem solution that offers brain-activated areas from EEG neural activity, was used to measure the sources of the ERP components in selected time windows for the difference between SW and BW conditions. This difference presumably best reflecting the isolation of the maximum values of feelings of interpersonal guilt (Leng et al., 2016), as it is the result of subtracting two degrees of relatively comparable emotions (see performance results) without the contamination of other confounding emotional responses.

3.3. Results

3.3.1. Performance

Throughout the experiment the rating scores for the feeling of guilt (yes/no question in the screen after feedback 2) for participants in the EEG chamber under the different conditions were calculated as percentage scores by dividing the total number of "yes" responses by the overall sum of trials (80) in each condition. The mean and SD for the individual ratings obtained were: SW (M=63.4%, SD=0.3), BW (M=33.6%, SD=0.3), BR (M=5.4%, SD=0.06) and PW (M=6.8%, SD=0.07). As expected, rating scores revealed the highest feelings of interpersonal guilt in the SW condition, a minor presence of this feeling was found BW, and for BR and PW it was negligible. ANOVA yielded significant differences between the four conditions [$F(3,69) = 41.174$, $p<0.001$]. Post hoc analyses showed significant differences between all conditions ($ps<0.001$) with the exception of BR vs PW ($p=0.054$).

3.3.2. SCR

Feedback 2 did not generate significant differences in the phasic activity between conditions [$F(3,51) = 0.137$, $p=0.86$; $\eta_p^2 = 0.008$].

3.3.3 ERPs and LORETA analyses

A visual inspection of ERP overall averages (Figure 3) suggested possible differences between the four conditions starting at around 300 ms and up to the end of the epoch. BR shows a clear difference from the other three conditions at all the electrodes, as is the case of SW, which was more negative than BW and PW in frontal sites. In the main conditions of interest, SW and BW required further detailed comparisons (Figures 4 and 5), showing a clear negative difference wave between about 300 and 500 ms after the onset of accuracy feedback 2, as well as minor positivity in parieto-occipital regions from about 700 ms up to the end of the epoch.

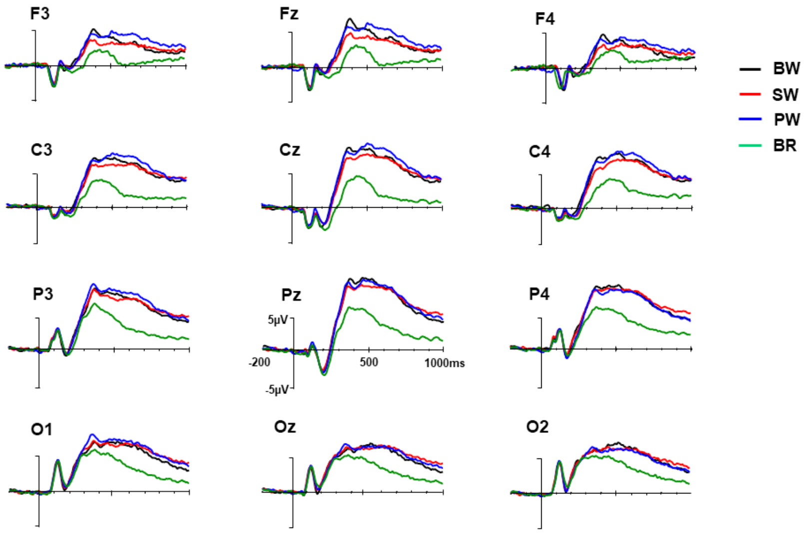


Figure 3.3. ERP grand means for BW (both wrong), SW (self wrong) BR (both right), and PW (partner wrong) conditions at selected electrodes.

An Electrode (59) by Condition (BR, PW, SW, BW) overall ANOVA showed significant effects of Condition between 200 and 1000 ms with all Fs $(3,69) > 2.36$ and all $ps<0.05$, endorsing subsequent ANOVAs at selected ROIs and time windows. An ANOVA at ROI1 for the 350-450 ms window revealed significant effects of Condition [$F(3,69) = 17.373$; $p<0.001$; $\eta p^2 = 0.430$] and "Electrode x Condition" [$F(21,483) = 6.808$; $p<0.001$; $\eta p^2 = 0.228$]. Post hoc analyses exhibited significant differences between all conditions BR vs. PW, BR vs. SW, BR vs. BW, PW vs. SW and SW vs. BW (all $ps<0.05$), with the exception of PW vs. BW ($p=0.1$).

An ANOVA at ROI2 within the 750-950 ms window showed significant effects of Condition [$F(3,69) = 16.86$; $p<0.001$; $\eta p^2 =0.423$] and "Electrode x Condition" [$F(21,483) = 7.04$; $p<0.001$; $\eta p^2 = 0.235$]. Post hoc analyses in ROI2, however, revealed that these effects were the consequence of the differences between the BR condition and the other conditions (all ps<0.001), while the latter did not differ significantly (ps>0.1).

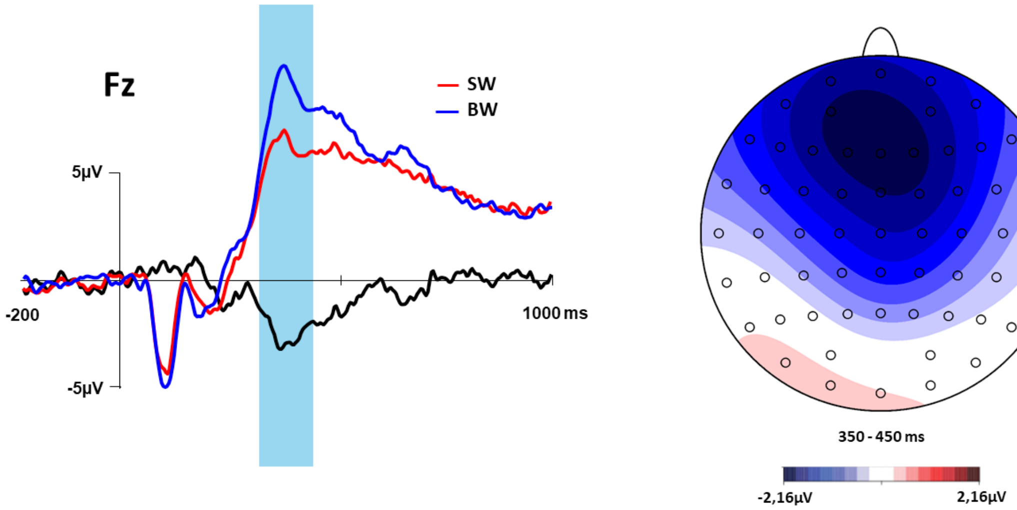


Figure 3.4. Guilt effect as the difference wave between SW (self wrong) and BW (both wrong) in electrode FZ, and a map representing the 350 – 450 ms interval.

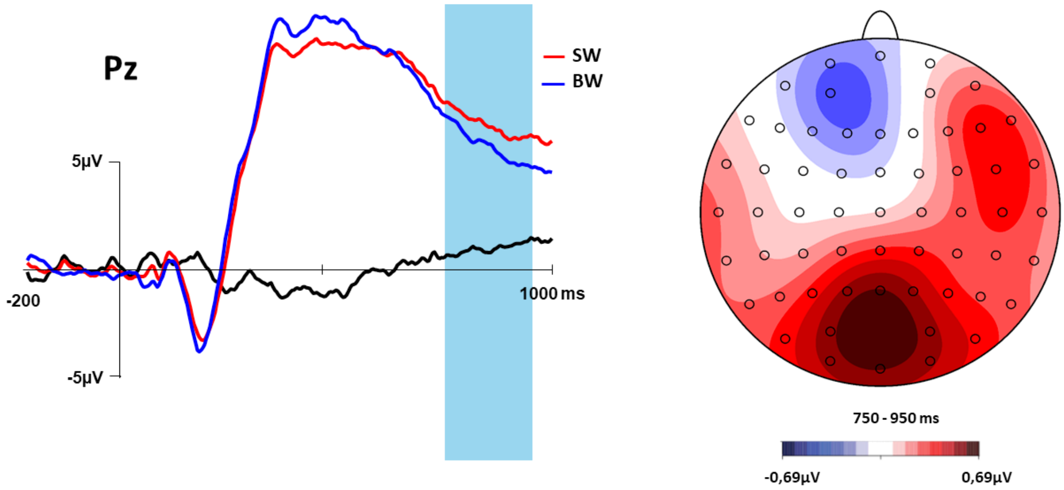


Figure 3.5: Guilt effect as the difference wave between SW (self wrong) and BW (both wrong) in electrode Pz, and a map representing the 750 – 950 ms interval (LPC).

In order to estimate the origin of frontal negativity (FN) related to interpersonal guilt, a LORETA analysis was performed for the SW difference minus BW conditions around a peak at about 400 ms (namely, 380-416 ms window) to increase the specificity of the solution. This consisted of an involvement of frontal regions with the highest value (0.95 mA/m2) at MNI coordinates (X=-10, y=45, z=50), corresponding to the dorsal medial prefrontal cortex (mPFC) (Fig. 6). A second minor contribution by the superior medial and lateral parietal cortex could also be mentioned, though this appeared to be an overlap of later, non-significant effects.

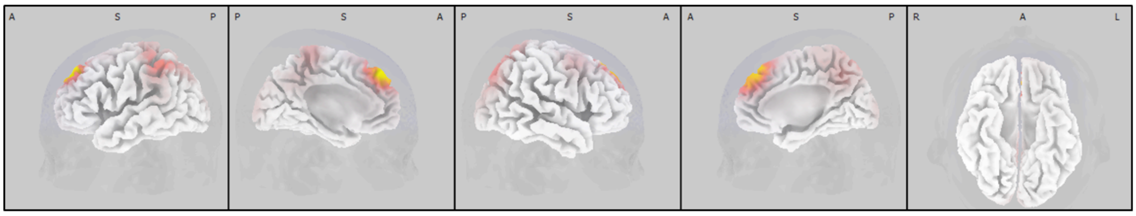


Figure 3.6. Neural generators (LORETA analysis) of frontal negativity, performed for the difference SW minus BW (self wrong vs both wrong) conditions, around its peak of about 400 ms (380-416 ms window) to increase the specificity of the solution.

3.3.4. Correlations between frontal negativity and variables linked to emotions and social cognition

Remarkably, the feeling of guilt reported during the experimental sessions and the amplitude of frontal negativity (FN) did not significantly correlate to each other, neither State-and trait-anxiety (STAI) scores with FN or reported guilt feelings. Reported guilt significantly correlated with empathic overall scores in TECA (Spearman's rho = 0.5, p=.012). Of main interest, the amplitude of the FN correlated highly and significantly with the TECA subscale Emotional understanding, related to cognitive empathy (Spearman's rho = -0.7, p<.001 after Bonferroni correction).

3.4. Discussion

Our study aimed to explore the behavioral, electrophysiological and anatomical correlates of interpersonal guilt, by concurrently measuring ERP, neural source analysis (LORETA), SCR, state- and trait-anxiety (STAI), and empathy (TECA). Our main aim was to replicate and expand the study conducted by Leng et al. (2016) on a Chinese sample, using a group of western participants. The goal was to further define the processes plausibly reflected by the main ERP modulations in an interpersonal guilt paradigm based on a dot estimation task game, with economic incentives, that presumably elicits interpersonal guilt.

The ERP fluctuations obtained in the present study largely replicate those of Leng et al. The components found here and in that study, as well as their latency and topography, are almost identical, with the only difference being relatively larger amplitudes in the present study –probably due to differences in the technical setup. These fluctuations consisted of a frontal long-lasting negativity (FN) in the SW condition, starting at about 300 ms and peaking around 400 ms, and a Late Positive Complex (LPC) with parieto-occipital distribution, from 500 ms onwards. The latter did not hold after statistical analyses, also largely paralleling the results in Leng et al. (2016), while the FN resulted in a robust finding. Of even greater interest, the FN appeared in our sample of Western (European) participants. This implies that at least the processes reflected by the FN and possibly connected with guilt appear to be universal and common across cultures under similar circumstances, despite reported cultural dissimilarities in experiencing complex social emotions (e.g. Walbott & Scherer, 1995; Anolli & Pascucci, 2005).

Leng et al. (2016) claimed that the FN was related to interpersonal guilt, as it emerged in the SW condition, which yielded the noticeably highest levels of self-reported guilt, while BW and PW did not differ from each other. This pattern is replicated here and the argument seems consistent. The rating of the feeling of guilt induced in participants in the SW (self wrong, partner right) condition was noticeably and significantly higher than that in the BW (both self and partner wrong) condition, while this feeling was essentially non-existent in the other conditions. This suggest that SW and BW conditions involve different levels of interpersonal guilt, and that the comparison SW minus BW might enable this variable to be isolated experimentally. However, to our surprise, we could not find a significant correlation between the amplitude of the FN and the reported feelings of guilt in the SW condition. There are two possibilities to explain this result. One is that the difference between SW and BW conditions yielding the FN involves other emotions apart from guilt. Indeed, shame might also be present because of the poor performance of the participant. Overall, the task appeared frustrating, and this feeling might be present in all conditions and even increase in the SW condition. The second possibility is that the FN is not specifically reflecting the feeling of guilt, but intermediate processes contributing to its final achievement.

We explored the first of these possibilities in deep and investigated the extent to which other emotions might account for the differences between SW and BW by conducting a behavioral study with a different sample of participants, as detailed in the Procedure subsection of the Materials and Methods section.

Overall, it appeared confirmed that other emotions can be present in the conditions of interest. Importantly, however, the results of this behavioral study also suggested that the differences between SW and BW are predominantly contributed by differences in the degree of guilt, even if smaller differences in frustration are also present. In sum, the paradigm still appears an appropriate experimental approach to study guilt, though with some limitations that must be pondered. This also applies to the studies by Leng et al. (2017) and Yu et al. (2014) employing it.

Accordingly, guilt seems importantly contributing to the differences between SW and BW, but the FN did not correlate with the reported feelings of guilt. This suggests that other processes, unobserved with the present paradigm, would also be required to finally yield this emotion, the FN reflecting only part of the operations necessary for achieving it. Overall, guilt seems a complex emotion, and it would be surprising if it were reflected by a single ERP modulation. The involvement of a number of areas in fMRI studies of guilt, as reviewed in the introduction, further reinforces this assertion. It remains to be elucidated what exactly the FN is reflecting in relation to guilt. The variables measured in the present study might provide clues in this respect.

The FN does not seem to be related to anxiety. Although the feeling of guilt is linked to personal distress (Etxebarria, & Apodaca, 2008), this did not seem to differ across conditions, as reflected in SCR. Similarly, overall levels of state- or trait- anxiety as measured with STAI did not correlate with the FN. Accordingly, the FN does not appear related to anxiety or arousal.

The most revealing results corresponded to scores in the measurements of empathy (TECA). The amplitude of the FN highly and significantly correlated with emotional understanding, a cognitive factor of empathy. Also of interest, the degree of reported guilt in the SW condition correlated with overall scores in TECA, which conveys an indirect link between the FN and self-reported guilt in our study, apparently mediated through empathy. Previous literature suggests that empathy is a core element for guilt to emerge (Baumeisteret al., 1994; Hoffman, 2000; Eisenberg, 2000; Morey et al, 2012). In view of our results, it can be suggested that the FN relates to cognitively understanding that the outcome of one's own action has caused distress to the partner, since people with higher scores in this factor exhibited higher FN values. Importantly, it is not the mere comprehension of the other's distress that would be reflected in the FN, since it did not emerge in other conditions that were possibly also upsetting for the partner (BW and, particularly, PW). It would rather reflect an empathic component specifically related to situations of self-caused harm to others, necessary for ultimately achieving feelings of interpersonal guilt. The result also reinforces the idea that the FN seems primarily related to processes involved in interpersonal guilt, disregarding the possible contribution of frustration to this fluctuation, even if frustration differed somehow between SW and BW conditions.

The processes reflected in the FN start as early as about 300 ms after the appearance of the information that permits participants to understand their own and other people's achievements in a joint task, which seems to originate in frontal brain regions. In this respect, the LORETA analysis showed that dorsal medial prefrontal (mPFC) areas appear critical in the generation of the FN. This is consistent with findings that relate this region with social cognition networks (e.g., Kédia et al., 2008; Morey et al., 2012), and, indeed, it emerges as one of the main areas underlying the feelings of guilt in a recent meta-analysis (Bastin et al., 2016). What is more important, dorsal mPFC is also a core area underlying empathy (Seitz et al., 2006). Overall, the evidence seems consistent with the processes suggested here, underlying the FN and their contribution to the feelings of guilt through empathic assessment. However, our result is not in line with the fMRI study by Yu et al (2014) employing a relatively similar paradigm to elicit guilt in the experimental contexts, as these authors reported a main implication of the aCC while our results were distributed more dorsally. Although this discrepancy might relate to the inherent risk of inaccuracies in the solutions for ERP generators, it appears to us that it is rather the result of important dissimilarities in the experimental procedures. These include the fact that the partner would receive physical pain stimulation in the wrong conditions. The aCC has been closely related with feelings of physical pain, both one's own and observed in others (Singer et al., 2004), as is the case of the anterior insula, a secondary finding in the Yu et al. (2014) study.

3.5. Conclusion

To conclude, we have been able to validate a frontal ERP modulation that seems consistent across cultures and that plausibly reflects empathic processes necessary for a final outcome of feelings of interpersonal guilt, though not the feelings of guilt themselves. Those processes, presumably occurring in dorsal mPFC areas, start around 300 ms after performance feedback is presented, and peak around 400 ms. The present data contribute to a better understanding of the dynamics of the neural mechanisms underlying such a complex social emotion as guilt.

Chapter 4. Study Two. Shame

4.1. Introduction

As previously cited in the introduction chapter emotions have been commonly classified into basic and social (Tangney & Fisher, 1995; Lewis, M. 2000). Basic emotions are the first to develop, are heritable and seem characterized by universal facial expressions, although this assumption is lately being discussed (Barret et al, 2011; Crivelli et al 2016; Fernandez-Dols & Ruiz-Belda, 1995). On the other hand, most social emotions require superior cognitive abilities that involve the self (Lewis, 2000; Tangney & Dearing, 2002). These imply self-evaluation, self-consciousness and the consideration about how the self is being evaluated by the others as we have broadly exposed in Chapter 1. Among them, guilt and shame are considered as highly relevant for human sociality, because of their relationship with moral rules, and their tendency to ensure the adherence to social norms through their internalization. The present study investigates on shame.

Shame requires self-evaluation and self-consciousness and usually appears in social evaluative situations as feelings of incompetence, when failing a task, or when we have as a consequence of an inappropriate social behavior (Menesini & Camodeca, 2008). Shame implies a damaged self and the devaluation of self-esteem, and could as well derive in aggressive behaviors such as anger, hostility or resentment (Tagney, Wagner, Hill-Barlow, Marschall & Gramzow, 1996), as responses to protect self-esteem (Thomaes, Stegge, Olthof, Bushman & Nezlek, 2011) and seems to emerge in order to avoid the loss of the individual or group status (Tracy & Robins, 2007).

Most of the functional neuroimaging studies have found that shame activates neural networks related with self-reflection, inner thoughts and mentalization (e.g., Basile et al., 2011; Gilead et al., 2016; Mitchell et al., 2006; Takahashi et al., 2004; Whittle et al., 2016). Indeed, two of the most consistent findings involve the dorsomedial prefrontal cortex (dmPFC) as well as the precuneus/posterior cingulate cortex (PCC) and the temporal poles, core areas of the so-called 'default mode network' (DMN) implicated in decoding the mental states of social agents (Andrews-Hanna, 2012). As a negative social emotion, it causes negative feelings and psychological pain (Tangney et al., 1996). Accordingly, the feeling of shame implies greater activation of the anterior insula (AI) and amygdala (Basile et al., 2011, Roth et al, 2014).

Although in social situations shame and guilt feelings usually overlap (Giner-Sorolla et al., 2011), despite several similarities, differences between shame and guilt in terms of involved brain regions seem neither conspicuous nor consistent, this being the case even when these social emotions are contrasted with some basic emotions, such as sadness or anger (Bastin et al., 2016).

The few studies performed so far have mostly employed paradigms in which these emotions are simply evoked by recalling or imagining situations. This might convey the primary involvement of circuits related to memory or imagination, rather than shame. This limitation has been recently surpassed by Sánchez-García et al. (2019), as we exposed in Study 1. In Zhu et al. (2017, 2019) they induced both guilt and shame in the same study, in the frame of an interpersonal context, by means of an economic game. In brief, in this paradigm shows some differences from study 1. The recorded participant advises a confederate (decider) on the number of dots that had just appeared on a screen; the advisor could see the dots for the double the time than the decider, the latter finally making a decision on the number of dots, earning money whenever the outcome was right. There were four possibilities, from the point of view of the feelings of the recorded participant: a) the decider follows the advice and both are wrong (guilt+shame); b) the advice is wrong, the decider ignores the advice yielding a right outcome (shame); c) the decider follows the advice and both are right (joy); d) the advice is right, the decider ignores the advice yielding a wrong outcome (uncertainty). However, in spite of the good ecological validity of this approach, the results of this study reproduced again the outstanding similarities between shame and guilt using functional neuroimaging although the authors could nevertheless disentangle both social emotions within the anterior cingulate and dmPFC when a multivariate pattern analysis was applied (Zhu et al., 2019).

It is possible that a more noticeable pattern of shame can emerge by studying the neural temporal dynamics of these emotions by means of Event-Related brain Potentials (ERP). ERPs permit the study of cognitive processes as these unfold over time, with a millisecond temporal resolution. This approach was used by Zhu et al. (2017) in order to disentangle shame and guilt, also using the interpersonal context of an economic game as described above. The authors were able to distinguish shame and guilt as soon as about 200 ms after the presentation of the decider's outcome to the adviser, as a significant amplitude increase of the P2 ERP component occurred in the shame condition. As the P2 relates to early selective attention (it is actually originated in visual associative areas; see Mangun & Hillyard, 1991), the finding was interpreted as an index of the self-relevance of the information involved in the shame condition.

The only studies in which both emotions have been jointly investigated and in the frame of interpersonal context are those by Zhu et al. (2017, 2019). The present study aims to contribute to this poorly explored field by studying shame in an interpersonal context by means of ERP, with a number of relevant novelties relative to Zhu et al. (2017, 2019).

In our study, a main change with regard to Zhu et al. (2017, 2019) relates to the stimuli presentation procedures. In Zhu et al., the adviser was informed about her failed advice preceding (by 1s in Zhu et al. 2017; 2s in Zhu et al., 2019) the critical point at which she was informed on the choice of the decider. In our view, this procedure might notably minimize the strength of shame feelings at the critical point at which it was registered in the ERP, that is, the outcome of the decision. As a result of own wrong outcome, shame would start earlier and, then, be already present along the baseline used to prove its effects. In the present study, based on previous research on guilt approached with ERP in an interpersonal context (Leng et al., 2016; Sánchez-García et al., 2019 / Study 1), the outcome of both adviser and decider were shown simultaneously. Another outstanding difference between the present and Zhu et al. studies is that we analyze the activity linked to the 'uncertainty' condition, in which the outcome of the adviser is right while that of the decider is wrong. Indeed, it might convey in some degree the feeling of emotions such as, e.g., pride or others- and then could serve to further contrast shame and guilt with emotional conditions other than joy (see Study 3). This could be done by increasing the total number of trials in the session, yielding enough linked to this condition –which was set to be less than other conditions to enhance participant's feelings of shame, as in Zhu et al. (2017, 2019).

In the present study, we have also used a different strategy to explore ERP fluctuations as used in Zhu et al. (2017). In the latter, the authors focused on a-priori defined ERP fluctuations linked to overall cognitive functions such as attention, cognitive control or working memory (i.e., P2, N2 and P3 components). Here we will not make hypothesis on specific ERP components, in line with previous studies (e.g., Yu et al., 2014; Leng et al., 2016, Sánchez-García et al., 2019 / Study 1). We directly contrasted the ERP fluctuations to the conditions of interest without a-priori restrictions, in search of possible effects specifically linked to the feelings of interest. Further, the so-obtained effects will be analyzed with Low Resolution Electromagnetic Tomography Analysis (LORETA; Pascual-Marqui et al., 2002). This algorithm decomposes the overlapping EEG voltage patterns into their underlying sources, localizing them within the brain. This way, a direct contrast of the results with previous neuroimaging studies is feasible.

With the purpose of better characterizing the functional significance of ERP fluctuations related to the social emotions conditions, we explored their possible relationship with a number of variables of potential interest that were measured in our participants. First, trait levels of guilt and shame were assessed with the Test of Self Conscious Affect (TOSCA; Tangney et al., 1996). Finally, we also recorded skin conductance response (SCR) along the experimental session as a measure of distress feelings, anxiety and emotional reactions (Liu et al, 2016) conceivably caused by shame and guilt.

The present study has two peculiarities that should be highlighted. First, after each trial participants only had to report whether they felt shame. Several reasons justify this procedure. Zhu et al. (2017) asked their participants to choose between guilt and shame after the trials in which either of these could be felt (the guilt and shame conditions), letting the participants to choose not to respond if they were uncertain about the currently felt emotion. However, when we performed a pre-study with this procedure, most of the trials were left unresponded. This was at variance with the results of Zhu et al. (2017), who nevertheless found a high number of equivocal responses (i.e., guilt responses in trials in which shame was expected, and vice versa). In spite of their conceptual differences, shame and guilt seem highly overlapping and difficult to unravel at the subjective level consistent with the neuroimaging literature, as discussed above. We also think that trials in which guilt is expected in the present paradigm (both adviser and decider are wrong) may also contain some degree of shame, even if guilt is the dominant feeling (as supported in Zhu et al., 2017, 2019), since the adviser is wrong despite having double the time than the decider to inspect the dots. Asking only for shame is intended to avoid participants' conflict and indecision (i.e., non-responded trials) while boosting participants' attention to this emotion. This in turn conveys the advantage of possibly enhancing the differences between shame and guilt in the ERP fluctuations, favoring the former, which on the other hand has been less studied than guilt with this technique.

The second particularity of the present study is that our sample comprises only female western participants, so the sample is more homogeneous while reducing sex-related variability in the psychophysiological and neurophysiological measures, as in Amodio et al. (2007) or Sanchez–García et al. (2019 / Study 1). As an outcome, the power of the ERP fluctuations related to shame may be enhanced and the results could be extendable to males.

4.2. Material and methods

4.2.1. Participants

Twenty-six female students (all undergraduate) participated in the experiment. Their age varied between 18 and 28 years, (mean age =20.5; SD=2.33). All but one were right handed, with scores ranging from 50 to 100 (M=76.91%) according to the Edinburgh Handedness Inventory (Oldfield, 1971). Participants reported no history of psychiatric or neurological illness and provided written informed consent before the experiment. The study was performed in accordance with the Declaration of Helsinki and approved by the ethics committee of the Hospital Clínico San Carlos. Participants received course credits for taking part in the experiment.

4.2.2. Procedure

Before the ERP experiment, the participating student met another student in the lab (actually, a confederate chosen by experimenters) and both completed two questionnaires: TECA (Lopez-Pérez et al., 2008) and TOSCA (Tangney et al., 1996). Thereafter, participants were told that they would play an advice-decision game. The participant was assigned (apparently for her, by chance) with the role of advisor and was invited to enter in the EEG shielded chamber while the confederate played as the decider and sat in an adjacent room. Participants would always receive the same amount of course credits, but they were told that the amount received by the decider depends on their successful decisions. Participants were also informed that the decider might not follow the advice. As mentioned in the Introduction section, and following Zhu et al. (2017, 2019), there were 4 possible conditions considering the feelings of the participant. When the decider follows the advice and both are wrong, this is assumed to cause feelings of *guilt* (as the dominant feeling –Zhu et al. (2017, 2019)-, even if, as explained, shame might also be present). When the advice is wrong and the decider ignores it yielding a right outcome, the feelings would correspond to *shame*. If the decider follows the advice and both are right, this would elicit *joy*. Finally, when the advice is right but the decider ignores the advice and yields a wrong outcome, this was categorized as *uncertainty* by Zhu et al. (2017, 2019), though in our opinion some degree of other emotions could not be discarded.

Both the adviser and the decider saw the stimuli on two LCD screens located at 65 cm distance from their eyes. Each trial (Fig. 1) began with a fixation cross appearing at the centre of the screen for 500 ms, followed by a black screen with twenty white dots (dot size = 3x3 pixels), randomly displayed, in a 300x300 pixels frame around the center of the monitor (x=0, y=0) to minimize ocular movements. Participants were instructed to estimate the number of dots on the screen. The dot stimuli were displayed for up to 1500 ms for the student and 750 ms for the confederate. After the dots disappeared, a number (19, 20 or 21, randomly chosen) appeared on the screen for 1500 ms, during which the advisor pressed a button with her response relative to whether the number of just-appearing dots was higher of lower than the appearing number. The response hand was counterbalanced across participants. Then, the decider made her decision within the following 3 s. After the presentation of a blank screen for 800, 1000 or 1200 ms depending on the trial (randomly chosen to avoid habituation) the feedback of the performance of the players was presented for 2 s with either a green tick (right response) or a red cross (wrong) below each participant's picture (feedback stimulus). To finish the trial, a yes-no question appeared on the screen for 1500 ms, and participants were asked to indicate if they felt shame upon seeing the outcome of advisor-decider performance. The position of the response alternatives ('yes', 'no') at the left/right of the screen was also counterbalanced across participants.

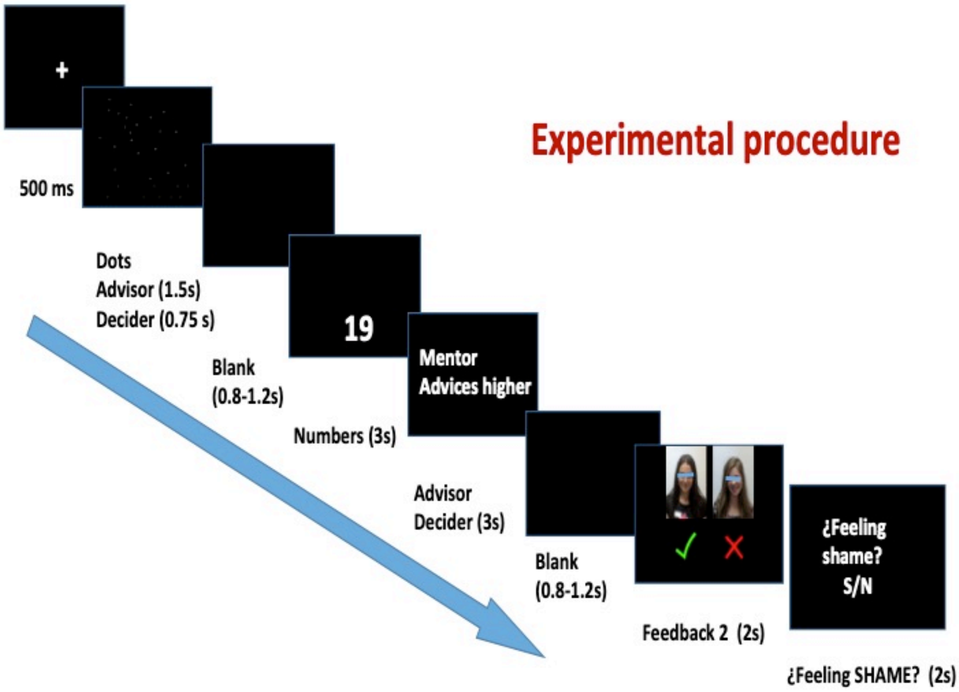


Figure 4.1. Experimental procedure. A fixed cross appears for 500 ms at the center of the screen before the onset of random dots for 1500 ms, where participants estimate the number, advisor for 1500 ms and decider for 750 ms. After a blank screen, 19, 20 or 21 appears for 3 s while advisor has to choose whether the actual number was greater or lesser. Feedback 1 screen appears for 3 s displaying the advice. After another blank screen, Feedback 2 with portraits of the participants appears for 2 s with a tick or a cross for individual performance. A "do you feel shame?" question (yes/no) ends a trial.

Importantly, all performance feedbacks were predetermined by the experimenters as follows: of the total of 360 trials (divided into two blocks with a brief pause in between), 30 were of the uncertainty condition, 110 of the shame condition, 110 of the guilt condition, and 110 of the happiness condition. The number of trials in the uncertainty condition was notably lower than in the other conditions, (Zhu et al.

2017, 2019), to avoid a feeling of disapproval in the advisor against a high negative decider response. The number of dots remained constant to 20 on every trial, though their spatial distribution varied randomly. The participants were unaware of these manipulations. A brief training of 15 trials was set to the adviser but acting as the decider, in order to mimic and experience their partners' role. Then, another training of 15 trials as advisor was set prior to the experiment. Overall the whole experiment lasted about 70 min.

4.2.3. Electrophysiological recording and analysis

EEG was recorded from 59 scalp electrodes mounted in an electrode cap (EasyCap), following the 10/20 International System. Bipolar vertical and horizontal EOGs were recorded to monitor blinks and horizontal eye movements. During recording, all scalp electrodes as well as one electrode at the left mastoid, were originally referenced to one electrode at the right mastoid; offline, they were re-referenced to the average of the right and left mastoids. The impedance of all electrodes was kept below 5 kΩ. The EEG data was analyzed with Brain Vision Analyzer® software. Raw data were filtered on-line with a band-pass from 0.01 to 100 Hz and sampled at 250 Hz; they were digitally filtered offline to a 0.1-30 Hz band-pass.

The continuous EEG was segmented into 1200-ms epochs, starting baseline 200 ms before the feedback screen onset. Eye-movements were corrected using Independent Component Analysis (ICA, Makeig et al., 2000) as implemented in BrainVision Analyzer®. Remaining artifacts were further removed by a semi-automatic rejection, eliminating epochs exceeding ± 100 μV in any of the channels and manually removing drifts or muscular artifacts. The mean rate of accepted epochs was 86.47%, and there were no significant differences between conditions in this regard ($F(3,66) = 1.645$; $p = 0.19$; $\eta p2 = 0.07$).

Statistical analyses were computed with IBM SPSS Statistics (Version 22). A first, overall repeated-measures ANOVA was performed including two factors: Electrode (59 levels) and Emotion (4 levels: shame, guilt, happiness, uncertainty). ERP voltage was measured as the mean amplitude within a particular time interval (from 200 to 1000 ms, in 100ms-wide windows). Second order post-hoc ANOVAs were thereafter performed in two regions of interest (ROIs) (Fig. 2) and at selected specific time intervals in order to increase statistical power: one central ROI, that included the electrodes, FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2, P1, Pz and P2 (ROI1), and one right-frontal ROI, that included AF4, AF8, F2, F4 and F6 (ROI2). The Geisser–Greenhouse correction for non-sphericity was applied when necessary (Greenhouse and Geisser, 1959). Post-hoc tests were corrected with the Bonferroni procedure. Activity considered as specific either of shame or of guilt was that in which the fluctuations linked to the corresponding condition exhibited a differential pattern when compared with the other social emotion as well as with the other two conditions (happiness, uncertainty). Specific results and topographies were established as based on the differences between shame and guilt or vice versa, as these are the result of subtracting two relatively comparable emotions without the influence of other confounding emotional responses of different typology and/or valence.

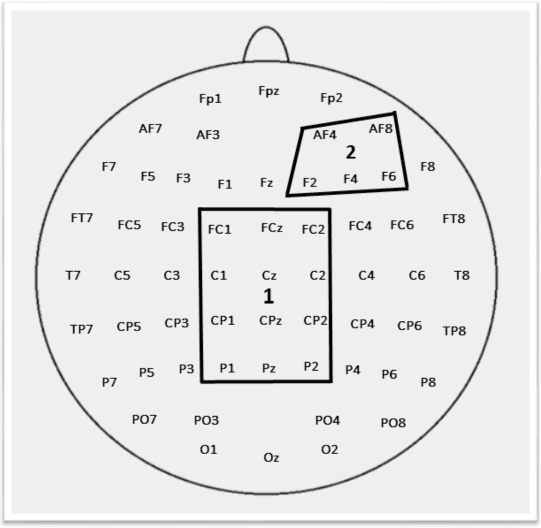


Figure 4.2. Location of the electrodes covered by the two main regions of interest (ROIs); a central ROI (ROI1) and a right frontal one (ROI2).

4.2.4. Source generator analyses

LORETA software, an algorithm based on inverse problem solution that offers brain activated areas from EEG neural activity (Pascual-Marqui et al., 2002), was used to locate the sources of the ERP fluctuations that could be estimated as specific of one of the social emotions of interest, i.e., shame and guilt. Only activity that could be considered as specific of one of these emotions –in consonance with ERP analyses- was used in these computations. These were made at selected time windows and based on the contrast values between shame and guilt and vice versa.

4.2.5. SCR recording and analyses

Skin conductance response (SCR) was monitored in each participant throughout the experiment. An 8-channel Multibox polygraph (Brain Products) placed to the index and middle fingers of the non-responding hand in the experimental task. SCR results were analysed with Ledalab software (Benedek & Kaernbach, 2010a). A downsampling process was applied to the original data to 125 Hz, and pre-processed by adaptive smoothing. Following these procedures, a composition-decomposition analysis (CDA) was performed to separate both the tonic and the phasic signals from the raw SCR (Benedek & Kaernbach, 2010b). Only phasic data with a minimum SCR amplitude of 0.1 µS were considered. Individual measures of phasic SCR were then averaged separately for each participant and condition. Phasic mean activity was standardized within subjects to facilitate the comparison of SCR means across conditions.

4.3. Results

4.3.1. Performance

The rating scores of the feeling of shame (yes/no question in the screen after the end of the feedback) under the different conditions were: joy (M=3,12%, SD=3,1), uncertainty (M=7,68%, SD=10,2), shame (M=60,24%, SD=30,39) and guilt (M=66,60%, SD=31,32). An ANOVA yielded significant differences in condition as main factor [F (3,66) = 72.5, $p<0.05$, $\eta p^2 = 0.76$]. Post-hoc analyses showed significant differences between all conditions [Ts(22) between 2.2 and 9.9, p always <0.05] with the exception of the comparison joy vs. shame [T(22) = 1.1, $p>0.1$].

4.3.2. SCR

There were no significant differences in the phasic SCR activity between conditions [F(3,51)= 1.04; $p>.05$; $\eta p^2 = 0.058$], possibly indicating constant high levels of arousal without trial-to-trial noticeable discrepancies, given the difficulty of the task, which is in consonance with the results of Sánchez-García et al. (2019) relative to guilt.

4.3.3 ERP and LORETA analyses

A visual inspection of the ERP waveforms (Figure 3) suggested differences between the four conditions starting around 200 ms after the feedback screen onset up to end of the epoch. Happiness and uncertainty revealed the most noticeable differences relative to the emotions of interest-shame and guilt-. The former presented smaller but apparent and specific differences when compared to the other conditions at two main time points. First, between about 200 to 500 ms after the onset of the accuracy feedback, shame displayed a more negative ongoing activity when compared to the other three conditions, a negativity with a mainly central distribution when compared to the guilt condition (Fig. 4.4.). Thereafter, guilt seems to display a specific difference relative to the other conditions, consisting in a–mainly right- frontal positivity when compared to the shame condition, and covering from about 400 ms to 800 ms (Fig. 4.5).

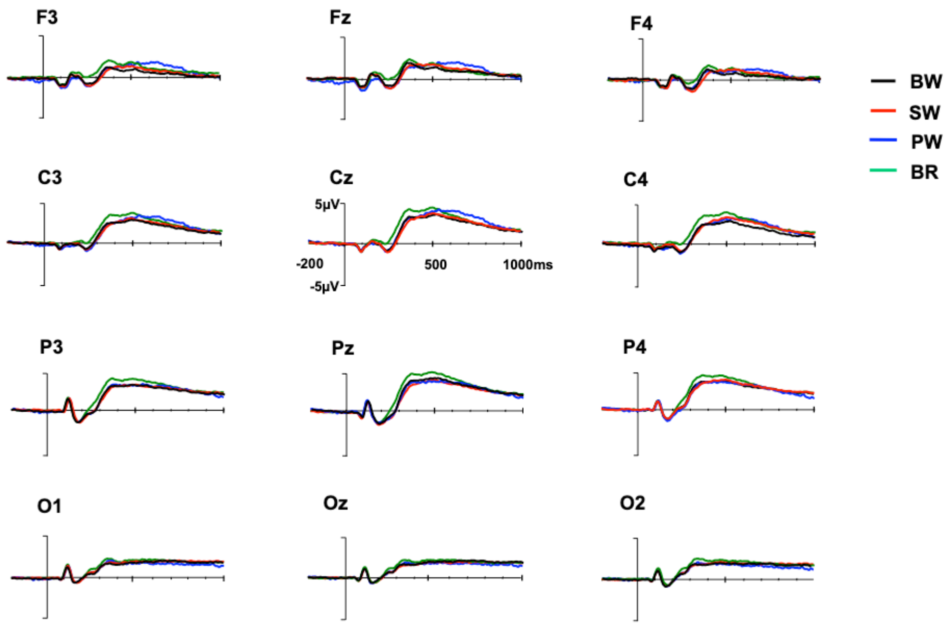


Figure 4.3. ERP grand means for BW (both wrong - guilt), SW (self wrong - shame) BR (both right - joy), and PW (partner wrong - uncertainty) conditions at selected

electrodes.

Statistical analyses confirmed these observations. First, overall ANOVAs showed significant effects of Electrode by Condition at all the 100-ms wide windows between 200 and 800 ms, with all F´s (174, 3828) > 4.801 and all p´s<0.05, endorsing subsequent ANOVAs at selected ROIs and time windows. An ANOVA at ROI1 for the window 300-400 ms, when the central negativity specific to shame exhibited its largest values, revealed significant effects of Condition [F(3,66) = 15.4; p<0,05; $\eta p^2 = 0.412$]. Post hoc analyses in this ROI and window exhibited significant differences between shame vs. guilt [T(22) = 3.7, p<0.001] and shame vs. joy [T(22) = 6.4, p<0.001]. The remaining comparisons were also significant (joy vs. guilt [T(22) = 4.6, p<0.001], joy vs, uncertainty [T(22) = 4.2, p<0.001]), with the exception of uncertainty vs. guilt [T(22) = 0.3, p>0. 1] and shame vs. uncertainty [T(22) = 1.7, p>0.1. An ANOVA at ROI2 for the window 450-550 ms, when the frontal negativity specific to guilt exhibited its largest values, revealed significant effects of Condition [F(3,66) = 4.170; p<0,05; $\eta p^2 = 0.159$]. Post hoc analyses in this ROI and window exhibited significant differences between all conditions, that is, joy vs. shame [T(22) = 2,8, p<0.001], joy vs. guilt [T(22) = 5.6, p<0.001], uncertainty vs. guilt [T(22) = 3.7, p>0.001], uncertainty vs. shame [T(22) = 2.6, p<0,05], and shame vs. guilt [T(22) = 2.1, p<0.001], with the exception of uncertainty vs. joy [T(22) = 0.7, p>0.1],

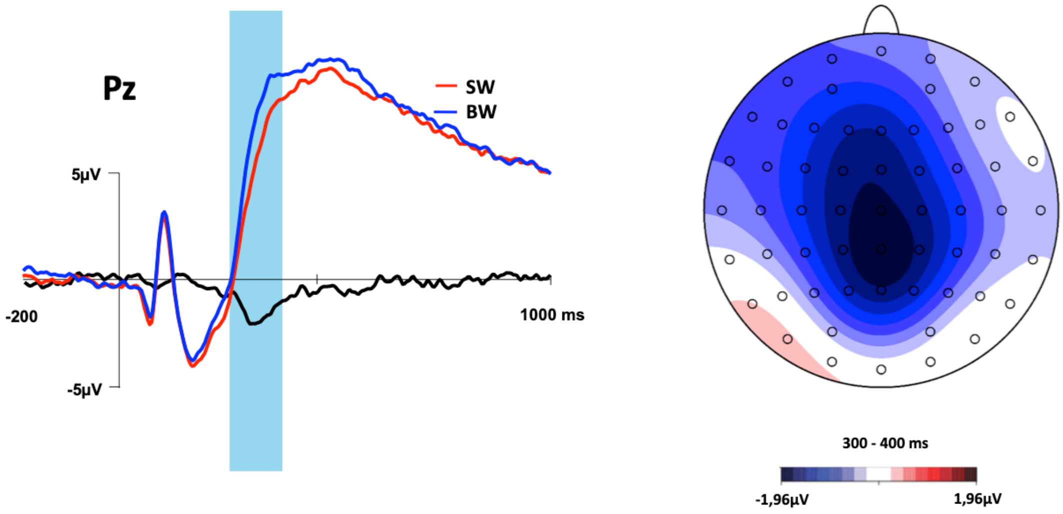


Figure 4.4. Shame effect as the difference wave between SW (self wrong) and BW (both wrong) in electrode Pz, and a map representing the 300 – 400 ms interval at ROI1.

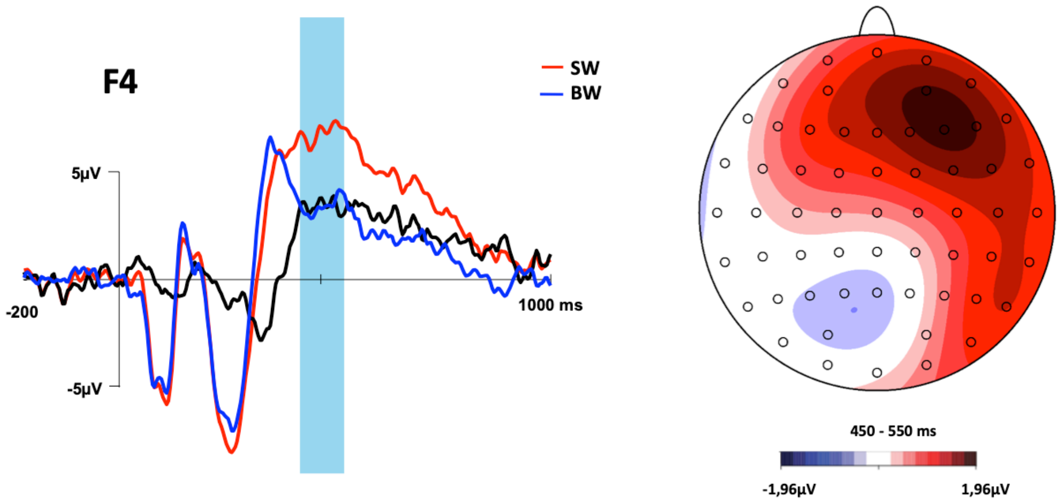


Figure 4.5. Shame effect as the difference wave between SW (self wrong) and BW (both wrong) in electrode F4, and a map representing the 450 – 550 ms interval (LPC) at ROI2.

To estimate the origin of the central negativity related to shame, a LORETA analysis was performed for the difference shame vs guilt around its peak; namely, 300-400 ms window. The solution consisted of an involvement of mid-parietal regions at MNI coordinates (X=5, y=-65, z=65), corresponding to precuneus/posterior cingulate (Fig. 5A). The same analyses were performed for the 450-550 ms window, in search of the neural origins of the frontal negativity related to guilt. The solution consisted of an involvement of mid-parietal regions at MNI coordinates (X=5, y=-65, z=65), corresponding to precuneus/posterior cingulate (Fig. 5B).

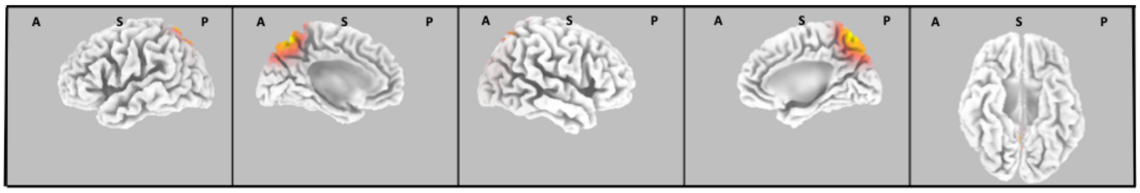


Figure 4.6. Neural generators (LORETA analysis) of shame condition, performed for the difference SW minus BW (self wrong vs both wrong) conditions, around its peak of about 350 ms (300-400 ms window) and 500 ms(450-550 ms window) to increase the specificity of the solution.

4.3.4. Correlations between ERP results and variables linked to emotions

Trait levels of guilt and shame as measured by TOSCA did not significantly correlate with any of the main ERP fluctuations that appeared to be specific of the corresponding conditions, neither with the reported responses about the current feelings of shame. Accordingly, traits did not relate with states in this paradigm, at least with the here uncovered fluctuations and reported outcomes. The same was the case with cognitive and affective empathy measures of the TECA, which is at variance with previous studies in which guilt has been approached with a different paradigm in interpersonal context (Sánchez-García et al., 2019).

Finally, the amplitude of the main ERP fluctuations that seemed specific of either shame or guilt did not correlate significantly with the feelings of shame reported after each corresponding condition (in both cases, Spearman's Rho = 0.29, p>.05).

4.4. Discussion

Our study explored the behavioral, electrophysiological and anatomical correlates of shame in an interpersonal context, by concurrently measuring ERP, neural source analysis (LORETA), SCR, shame and guilt variables (TOSCA-2), and empathy (TECA). Our main aim was to replicate and expand the study by Zhu et al. (2017) in a Chinese sample, by using a group of western participants and further defining the processes plausibly reflected by the main ERP modulations in dot estimation task economic game arguably eliciting these social emotions.

The rating score of shame feeling induced in participants in the shame and guilt conditions were significantly higher than in the other conditions. This is in line with Leng et al. (2016) and Zhu et al. (2017), noticing a higher rating score in our study and supports the achievement of a valid procedure to induce this social emotion. Nevertheless, shame and guilt scores were noticeably similar suggesting an overlapping and semantic confusion between participants to discriminate between the feeling and the appropriate word to express similar social emotions.

The percentage of responses reporting feeling shame in the guilt condition was similar to that in the shame condition does not necessarily convey that the intensity of shame felt was comparable in both shame and guilt conditions, since the measure is not scoring intensity. The result is simply indicating that some degree of shame was felt in the guilt condition, as predicted. Also as expected, happiness and uncertainty yielded very low values of reported shame.

Beside that our results are opposite to the idea that in non individualistic cultures, the feelings of guilt or shame could be higher. These differences are very remarkable for complex social emotions such as guilt (Walbott & Scherer, 1995; Anolli & Pascucci, 2005) and shame (Camras & Fatani, 2004; Tsai, 2006; Tracy & Robins, 2006). For instance, in Chinese language there are more than 100 terms related to shame (Li et al., 2004). In our case, the early shame results found might reflect latencies, amplitudes and topography that could not be attributed to cultural differences of the samples suggesting an equal feeling and expression of these social emotions between, at least, the two samples studied (Chinese Zhu et al. 2018 and our sample).

On the other hand, the ERP fluctuations obtained in the present study are quite different to those of Zhu et al., 2017. The components found, as well as their latency and topography, are similar since we find a centro parietal negativity (CPN) in the SW minus BW condition, starting at about 200 ms and peaking around 400 ms, CPN was significance solely in 200-300 ms window for Zhu et al. (2017). We also found a late right frontal positivity, from 450 ms onwards. The latter did hold statistical significance in the 450-550 ms time window, further differenciating the results in Zhu et al. (2017) and linked to a two step processing (early and late) of social emotions as in Sánchez-García et al 2019.

The present work therefore adds knowledge to the neuroscientific study of emotions by verifying that a centro-parietal negative fluctuation starting about 200 ms and a right-frontal positivity at 450 ms concurs with conditions in which people feel shame. Of the highest interest, these fluctuations have appeared in our sample of

Western (European) participants. This implies several consequences as the shame processes reflected by the centro-parietal negativity and right-frontal positivity seem to extend beyond cultures under similar circumstances eliciting these feelings, despite well reported cultural differences in experiencing complex social emotions (e.g., Walbott & Scherer, 1995; Anolli & Pascucci, 2005).

Zhu et al., (2017) claimed that an early P2 was related to shame, emerging from the SW condition. This pattern is replicated here, confirming P2 CPN as an early, and longer (up to 400 ms) component of shame linked to early attentional processing of self-relevant information (Chen et al., 2011). The argument seems consistent as shame, as a moral emotion, relates to self reference processes. This early component might reflect quick information that leads to social devaluative situations, therefore providing a way to protect ourself image and status (De Hooge, et al., 2010). The right frontal positivity emerging at 450 ms, not found in Zhu et al. 2018, could reflect reanalysis and self-assesment processes linked to social comparison. Several studies have reported LPC components during social evaluative encounters (Qiu et al., 2010; Wu et al., 2012).

Furthermore, right-prefontal cortex plays a major role when moral decisions are taken, as perceived unfairness (Knoch et al. 2006), negative emotional judgment (Grimm et al. 2008) and inhibitory regulation in cognitive and emotional processes (Depue et al., 2015) in line with the inhibitory control and re-assesment of shame feelings.

We could not find a significant correlation between the amplitude of the centro parietal negativity and right frontal positivity and the reported feelings of shame in the shame (SW) condition, which might indicate that both fluctuations do not seem to relate directly to the final outcome of shame feelings. Centro parietal negativity (CPN) and right frontal positivity (RFP) are probably reflecting a very important part of the processes involved in these feelings, but other processes might also be in play. Our interpretation hypothesis relies on the common semantic overlapping when expressing in conscious thoughts previously felt. Since shame is a complex emotion, processes of late evaluation may indeed involve further reanalysis.

Although shame feelings may even relate to associations between hypothalamic-pituittary- adrenal (HPA) and proinflamatory immune cytokines release and shame experience related to threaten of social self (Dickerson & Kemeny, 2004; Gruenewald et al., 2004). Therefore, linked to personal distress, this did not seem to be different across conditions, as reflected in SCR. The presence of other activating emotions in the other conditions, and even similar (guilt - BW condition) or high arousal levels throughout the session due to the demanding dot estimation task, might explain this result. In fact, along the experiment, participants never reached to control the task or actually count the total number of dots at ease. The ERP shame results therefore do not seem related to distress or arousal.

The underlying sources of the EEG voltage patterns point to mPFC (second best match) and Precuneus/PCC (first match) as source regions. Medial prefrontal (mPFC) area appears to be crucial in the generation of the RFP being consistent with findings that relate this region with social cognition networks reflecting what might others think about our performance and social status (e.g., Kédia et al., 2008; Morey et al., 2012). Precuneus and posterior cingulate cortex (PCC) is also involved being linked to monitoring environmental changes and memory re-encoding (Andrews-Hanna et al. 2014).

As expected, scores in the cognitive and affective measurements of empathy (TECA) do not relate to shame feelings. Empathy is a core element for guilt, not shame, to emerge (Baumeister, Stillwell and Heatherton, 1994; Hoffman, 2000; Eisenberg, 2000; Morey et al, 2012) as the comprehension, cognitive and emotionally of other´s distress is crucial to elicit guilt feelings and amending and approaching consequent behavior. Empathy indeed is a core component of watching others pain (Singer et al. 2004) a key issue for feeling interpersonal guilt. On the other hand, shame behavior leads to self appeasement, hiding and avoiding, not empathizing, actions.

This shame process starts as early as about 200 ms after the feedback in which participant realizes own and other's achievements in the advisor task. In this regard, the LORETA analysis showed that precuneus area seem to be critical in the generation of this CPN and RFP. This is consistent with findings that relate this area to self evaluative processes as proposed by Takahashi et al. (2004) comparing fMRI brain activation of guilt and shame, showing shame to guilt more activation in the anterior and posterior cingulate cortex.

Overall the evidence seems highly consistent with the suggested dynamic processes of centro parietal negativity and right frontal positivity and their contribution to the feelings of real shame through a social comparison task.

4.5. Conclusion

In conclusion, we have been able to validate an early ERP modulation that seems consistent across cultures and that extends to late reanalysis processes of emotional content. Nevertheless these processes do not relate to the feelings of shame themselves. Those processes, presumably occurring at precuneus areas, start as soon as about 200 ms after performance feedback is presented, peak around 400 ms and keep on with a latter phase at 500 ms. The present data contributes to better understand the dynamics and complexity of the neural mechanisms underlying the social emotion of shame.

Chapter 5. Study Three. Pride and shame

5.1. Introduction

As we have understood in previous chapters social emotions play a critical role to establish social behavior and human moral, being considered of the highest relevance for people´s adherence to social norms (Beer & Keltner, 2004; Tangney et al., 2007). They regulate and motivate human thoughts, feelings, and behavior, encouraging us to achieve actions in order to attain a valued social position (Tracy & Robins, 2007). On the other hand, social emotions are cognitively complex. They require self-awareness (Abe & Izard, 1999; Lewis, 2000), as well as the ability of mentalizing, as they imply concerns about others (Haidt, 2003; Tangney and Dearing, 2003). Despite their centrality to understand the distinctiveness of human mind and its evolution (Flinn et al., 2005), the neural underpinnings of social emotions have been scarcely studied. The present experiment aims at contributing to this field by focusing at the same time on two of these emotions: pride and shame.

Pride is a pleasant emotion which is experienced when an individual improves in social status by achievement and effort (Tracy and Robins, 2007). This is called *authentic pride*, and should be discerned from *hubristic pride*, a non-adaptive emotion arising from lack of empathy, narcissism and self-importance (Carver et al. 2010; Tracy and Robins, 2007). Shame, by contrast, is an unpleasant emotion that evokes pain and aversive feelings arising from situations driven by failure to one or others' ideal standards (Tangney et al., 2007) or from an inappropriate social behavior (Menesini & Camodeca, 2008). It implies a damaged global self, leading to anger, hostility, or resentment (Tangney et al., 1996), and even anxiety or depression (Gilbert, 2000; Tangney et al., 1992). Though the identification of a distinct neurobiological basis is critical for the characterization and definition of any basic emotion (Tracy & Randles, 2011), this has been highly elusive regarding social emotions. Overall, these emotions seem to involve basic emotional processing brain structures, typically located subcortically -such as the amygdala-, together with others related to self-reference and mentalizing and mostly placed in cortical areas (Bastin et al., 2016). Indeed, the latter seem to better characterize the neural underpinnings of social emotions. Yet, the depiction of distinct patterns singularizing social emotions like shame and pride within these self-reference and mentalizing regions is far from available.

Recent neuroimaging evidence reveals the special relevance of medial prefrontal and posterior regions belonging to self-referential and mentalizing circuits during both shame (Michl et al., 2014; Gilead et al., 2016; Roth et al., 2014; Zhu et al., 2019) and pride (Zahn et al., 2009, 2014; Simon-Thomas et al., 2012; Roth et al., 2014). Besides, other studies have also reported the involvement of the anterior temporal lobes in shame (Bastin et al., 2016), and the ventrolateral and dorsolateral prefrontal cortex or the superior, middle and inferior temporal regions in pride (Hong et al., 2019; Takahasi et al., 2008). Remarkably, however, Roth et al. (2014) studied pride and shame concurrently and found them indistinguishable in terms of brain activated areas: both emotions engaged the same basic emotion-related (amygdala, insula and ventral striatum) and self-referential (dorsomedial prefrontal and posterior cingulate cortex) regions. Only a larger activation of the basic emotion-processing areas by pride could distinguish these emotions, interpreted as due to an intrinsic hedonic value of pride as a positive emotion. If we pursue the identification of distinct underpinnings to better characterize and dissociate pride and shame at the neural level, the current situation appears restraining. The reduced number of studies is one determining factor. The present study aimed to expand current knowledge on the neural mechanisms by dissociating pride and shame in two remarkable ways that we explain below.

On the one hand, Event-Related brain electrical Potentials (ERP), at variance with other neuroimaging techniques, permit to study the neural temporal dynamics in milliseconds, which could, by itself, unveil noticeable differences between shame and pride. Moreover, ERP can also provide an approach to structural information by means of source-estimating algorithms (Awan et al., 2009). Zhu et al. (2017) used ERP to disentangle shame and guilt: both emotions were distinguished about 200 ms after the presentation of the stimulus prompting either emotion, since a significant increase of the P2 component occurred in the shame condition. As the P2 relates to early selective attention (it is actually originated in visual association areas; see Luck & Hillyard, 1984), the finding was interpreted as an index of the self-relevance of the visual information eliciting shame. Our second experiment, went a step further to study shame finding a centro parietal negativity emerging at 300 ms after presentation of the stimulus, linked to early attentional processing of self-relevant information in social comparative situations (Chen et al., 2011, De Hooge, et al., 2010) and a right frontal positivity emerging at 450 ms linked to and self-assesment in social comparisons.

To our knowledge, while no other studies has approached shame by means of ERP, none has been done in relation to pride. The present study records ERP fluctuations to stimuli presumably eliciting pride and shame in the same experimental session.

A second way in which the present work contributes to the literature is through the development of an experimental paradigm seemingly prompting genuine feelings of pride and shame, in controlled social context in the lab. With the exception of the Zhu et al., 2019 (studying shame), most of the neuroimaging literature on pride and shame elicited these emotions through cue recalling or visual imagination (e.g., reading a sentence or viewing a related picture). With these procedures, memory and imagination neural processes may interfere with the target emotions processes, while social context and evaluations, which are critical factors in shame and pride, might be absent. (Robertson et al., 2018). This caveat can be overcome by implementing interpersonal games in the lab. This approach has been used in studies on shame (using ERP -Leng et al. 2016; Zhu et al., 2017; Sánchez-García et al. 2019- and fMRI -Zhu et al., 2019-), but it has never been applied to study pride.

The present study used an innovative paradigm, endorsing the direct study of both emotions at the neural level in the same experiment. In the paradigm developed here, the participant played online against three other people in a dot-estimation task. After every trial, the participant received feedback on proper and others' correctness or failure. Eight different conditions were created, depending on the feedback pattern. In 4 of them, the participant was right, and the performance of the other three participants could be: a) they all were right, b) two of the others were right, c) one of the others was right, or d) none was right. This was repeated in the 4 conditions in which the participant was wrong. What is distinctive about this paradigm is that emotions of pride and shame can be elicited in the same session, as a function of the comparison between proper and others' performance. *Pride* would be maximal when the participant is right and all the other players are wrong. The corresponding control trial for this condition would be when both the participant and the other three players are right. In turn, *shame* would be maximal when the participant is wrong and all the other players are right, and the corresponding control trial would be when both the participant and all the other three players are wrong.

According to previous findings on ERP with social emotions (namely, guilt and shame) (Leng et al. 2016; Zhu et al, 2017; Sánchez-García et al. 2019) we expect a dual-time processing pattern, i.e., early and late components related to pride and shame, possibly reflecting different processing steps contributing to achieve these social emotions. Although the ERP technique is relatively unsuitable to explore activity from subcortical emotion-processing areas such as the amygdala, it is in turn highly sensible to cortical activations, including the medial areas (Luck, 2014). By means of source-estimating techniques, we expect to find the involvement of self-reference and mentalizing areas in both the medial and lateral cortices, presumably differing in their dynamic involvement when comparing pride and shame.

5.2. Materials and methods

5.2.1. Participants

Thirty-six undergraduate students participated in the experiment. Four were discarded due to excessive artifacts in their recordings. The remaining sample was composed of 16 females and 16 males. Their age varied between 18 and 28 years (M = 19.7 years; SD = 1.87). All participants were right-handed, with scores ranging from 50 to 100 (M=86.66%) according to the Edinburgh Handedness Inventory (Oldfield, 1971). Participants reported no history of psychiatric or neurological illness and provided written informed consent before the experiment. The study was performed in accordance with the Declaration of Helsinki and approved by the ethics committee of the Hospital Clínico San Carlos (Madrid). Participants received 15€ for taking part in the experiment.

5.2.2. Procedure

Before the ERP experiment, the participant was instructed in the experimental procedures. The goal was to play a dot-estimation task online with three other students from different universities, other than the host institution. Actually, before the task, experimenters showed a simulated conference video call where the three other students were being set up their EEG caps in their respective labs to enhance a social context setting (Figure 1). The participant was unaware of this pretending manipulation, since conference video call was pre-recorded, in a way that appeared realistic to the ongoing situation. The aim of the video was to make feasible and reliable the social setting of the experiment. In order to balance the gender composition in the social context, the video call was different depending on the gender of the participant: for female participants, the other 3 participants were two males and one female; for male participants, two females and one male. The participant was told that she/he was participating in a competition between different universities in Spain on the visuospatial abilities of their students, therefore promoting a social comparison situation.

Figure 5.1. Simulated pre-experiment ‘conference video call’, where the three other pretended participants are shown to the actual participant

The task consisted in a dot-estimation task, in which the participant received feedback on proper and others' correctness or failure after every trial. Each feedback consisted in a composition with the pictures of the four persons presumably performing the task, the one of the participant located at the bottom center of the image, together with a hit or an error mark below each person's picture (Figure 2). As depicted in Figure 2, there were 16 possible feedbacks, as a function of the combination of the correctness of the answer of the participant and of each of the 3 opponents.

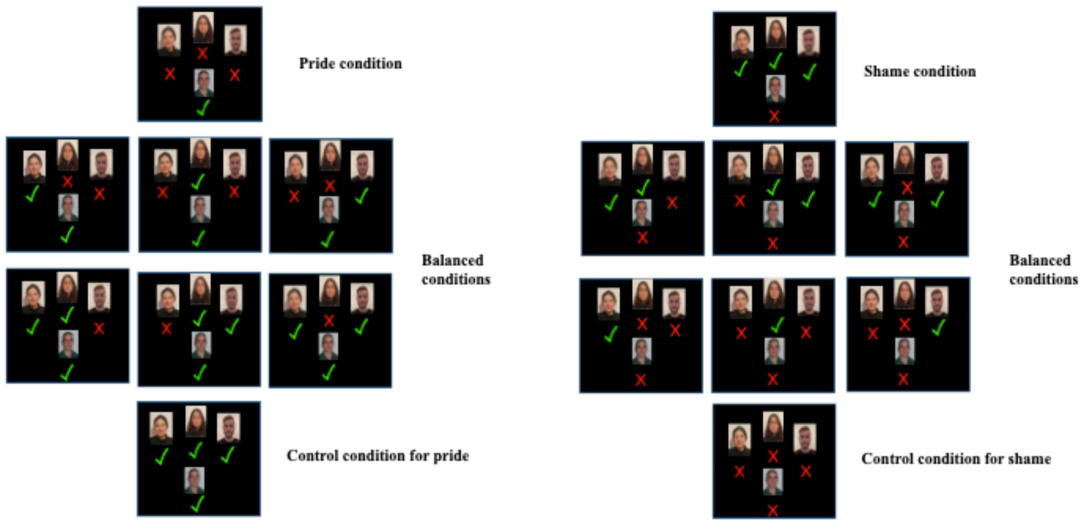


Figure 5.2. The 16 possible feedbacks displayed after every trial in the dot-estimation task.

The participant was seated on a comfortable chair while the stimuli were presented on an LCD screen located at 65 cm distance from her/his eyes. Each trial began with a fixation cross appearing at the center of the screen for 500 ms (Figure 3), followed by a black screen with twenty white dots (dot size = 3x3 pixels), randomly displayed, in a 300x300 pixels frame around the center of the monitor (x=0, y=0) to minimize ocular movements. The participant was instructed to estimate the number of dots on the screen. The dot stimuli were displayed for up to 1500 ms. After the dots disappeared, a blank screen appeared for 800, 1000 or 1200 ms depending on the trial (randomly chosen to avoid habituation), followed by a number (19, 20 or 21, randomly chosen) for 1500 ms, during which the participant pressed a button, the response being relative to whether the number of just-appearing dots was higher of lower than the preceding number. The response hand was counterbalanced across participants. After the presentation of a blank screen for 800, 1000 or 1200 ms, depending on the trial, the feedback about the performance of the 4 players was presented for 2 s, followed by a 1s black screen.

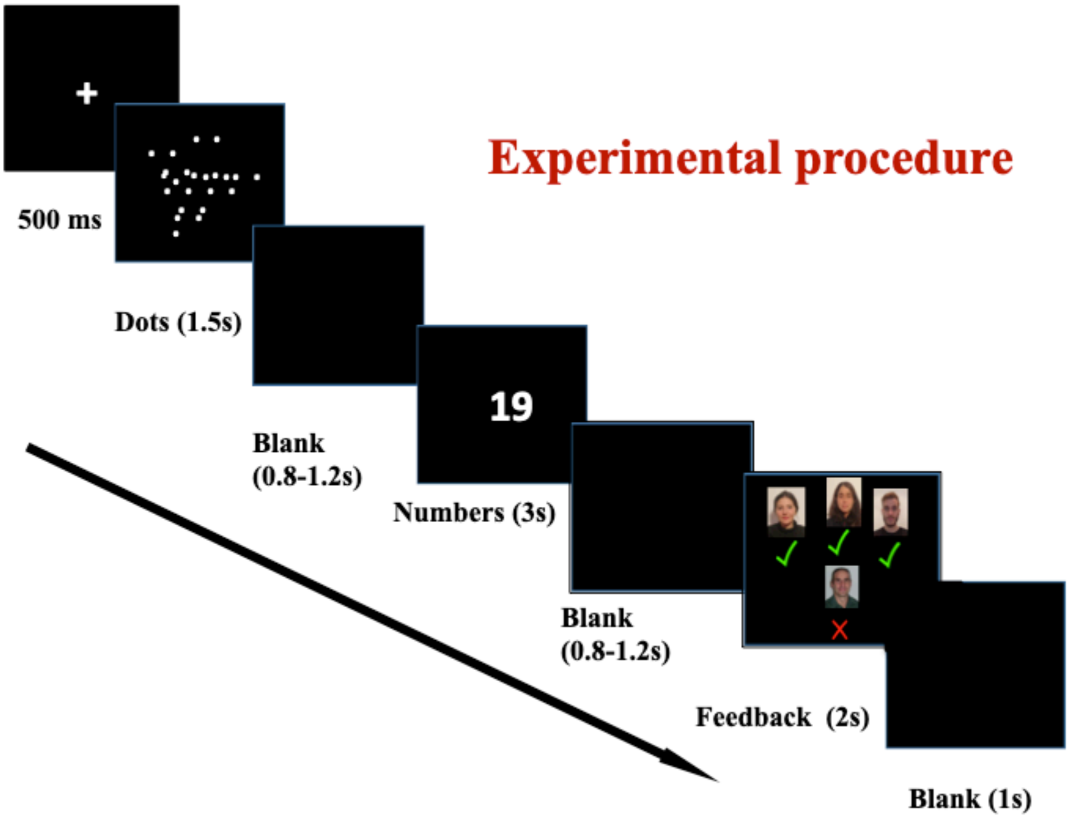


Figure 5.3. Schematic representation of the experimental procedure.

All performance feedbacks were predetermined by the experimenters, the participant being unaware of this circumstance, and they were randomly distributed in a realistic manner, as follows. There were a total of 480 trials. In half of them, the participant was right, and wrong in the other half. Within the 240 right trials, the distribution was: in 35, the other three participants were wrong (pride condition); in 55, all the other three participants were also right (control for pride condition); in 75, two of the other participants were also right, and in 75, one of the other participants was also right. Correspondingly, within the 240 wrong trials, the distribution was: in 35, the other three participants were right (shame condition); in 55, all the other three participants were also wrong (control for shame condition); in 75, two of the other participants were also wrong, and in 75, one of the other participants was also wrong. A brief training of 12 trials preceded the experiment. Overall, the whole experiment lasted about 90 min, divided into two blocks with a brief pause in between.

At the end of the recording session, the participant completed a survey on self-reported emotions to each different feedback, in order to check the validity of our assumptions relative to the emotions elicited. To each of the 16 different possible feedbacks a 7-point Likert scale was used to assess five different feelings: pride, joy, shame, anger and sadness.

5.2.3. Electrophysiological recording and analysis

EEG was recorded from 59 scalp electrodes mounted in an electrode cap (EasyCap), following the 10/20 International System. Bipolar vertical and horizontal EOGs were recorded to monitor blinks and horizontal eye movements. All scalp electrodes as well as one electrode at the left mastoid were originally referenced to one electrode at the right mastoid during recording, and offline re-referenced to the average of the right and left mastoids. The impedance of all electrodes was kept below 5 kΩ. The EEG data was analyzed with Brain Vision Analyzer® software. Raw data were filtered on-line with a band-pass from 0.01 to 100 Hz and sampled at 250 Hz; they were digitally filtered offline to a 0.1-30 Hz band-pass.

The continuous EEG was segmented into 1200-ms epochs, starting with a baseline of 200 ms before the feedback screen onset. Eye-movements were corrected using Independent Component Analysis (ICA, Makeig et al., 2000) as implemented in BrainVision Analyzer®. Remaining artifacts were further removed by a semi-automatic rejection, eliminating epochs exceeding ± 100 µV in any of the channels and manually removing drifts or muscular artifacts. The minimum percentage of accepted epochs per condition was always 80%, with a mean of 85,55%.

Statistical analyses were computed separately for pride and shame, as a) these emotions elicited activations with dissimilar time windows and b) their corresponding control conditions were not comparable. The comparisons comprised the ERP time-locked to pride and shame conditions feedbacks and those to their corresponding control condition. Considering the advantages and disadvantages of different types of analyses, two approaches for data analyses were applied: a) factorial cluster-based permutation analyses (Fields and Kuperberg, 2019; Groppe et al., 2011a,b), and b) time windows analyses based on visual inspection and guided by consideration of the results of the cluster analyses.

Cluster-based permutation analyses were calculated using the Factorial Mass Univariate Matlab® Toolbox. In agreement with the dual-time processing pattern observed in the ERP, and guided by previous literature (e.g., Brusini et al., 2017; Jiménez-Ortega et al. 2020), we considered early (0-500 ms) and late (500-900 ms) time segments for these analyses. Subsequently, two (one per time segment) factorial cluster-based permutation analyses were calculated separately for each emotion (pride, shame), each with 10,000 iterations and alpha level of .05 involving the factor Emotion

(Emotion –i.e., pride or shame- vs the corresponding Control condition). For the time windows analyses, repeated-measures analyses of variance (ANOVAs) with the SPSS 22® software were performed. Guided by both visual inspections and cluster analyses, two scalp regions of interest (ROIs) and time-windows were used per emotion. For pride, ROI1 comprised the electrodes F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2, CP1, CPz and CP2, and was analyzed in the 250-350 ms window. ROI2 for pride comprised the electrodes F1, Fz, F2, FC1, FCz and FC2, and was analyzed in the 600-800 ms window (Fig.4). For shame, ROI1 comprised the electrodes AF3, AF4, F1, Fz, F2, FC1, FCz, and FC2, analyzed in the 330-410 ms window. ROI2 for shame comprised the electrodes FC1, FCz, FC2, C1, Cz, C2, CP1, CPz and CP2, and was analyzed in the 715-875 ms window (Fig.5). The main ANOVA used the mean voltage values of the electrodes grouped within each ROI, and involved the factor Emotion, in the same terms as in the cluster-based analyses. Violations of the sphericity assumption were corrected when necessary by the Greenhouse-Geisser method.

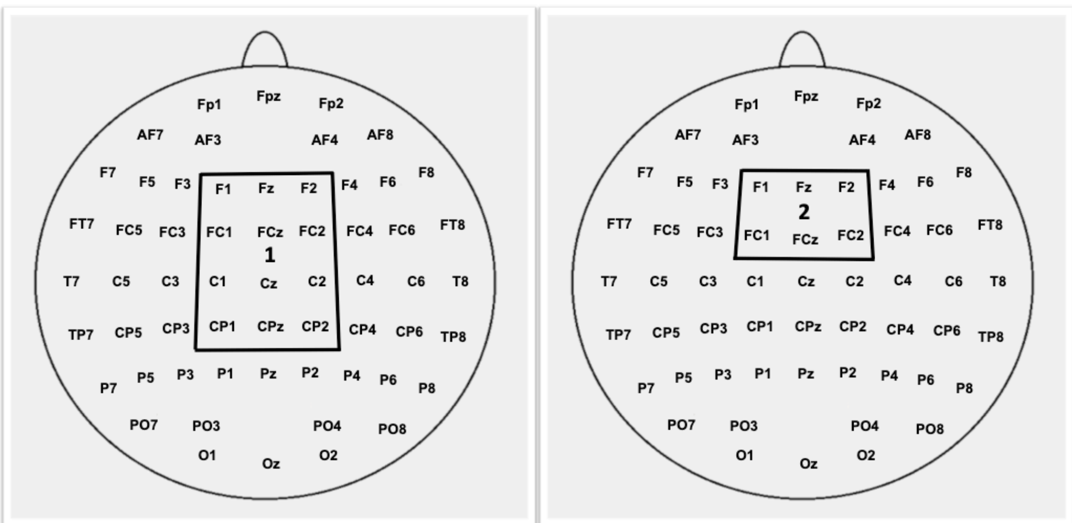


Figure 5.4. Location of the electrodes covered by the two main regions of interest for pride (ROIs); a fronto central ROI (ROI1) and a frontal one (ROI2).

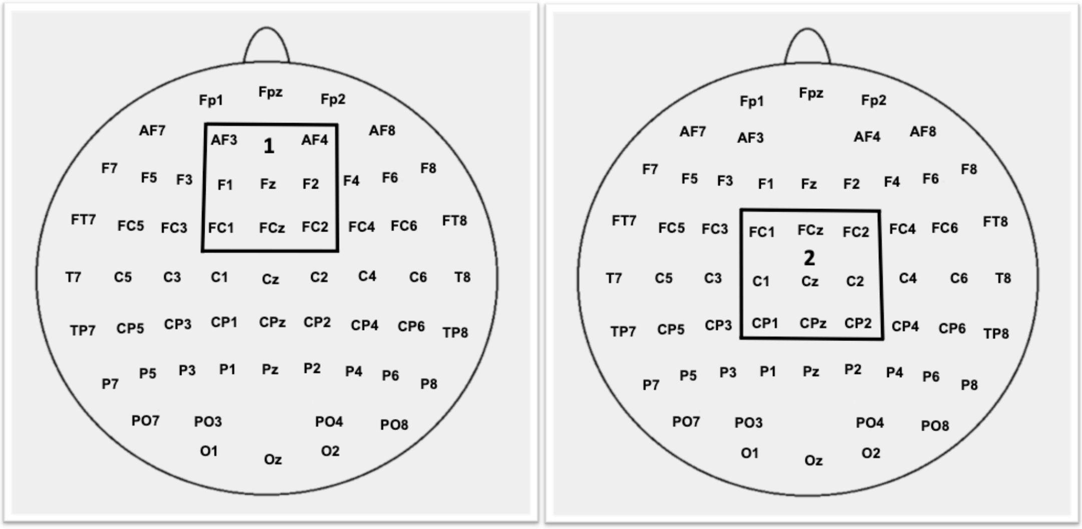


Figure 5.5. Location of the electrodes covered by the two main regions of interest for shame (ROIs); a frontal ROI (ROI1) and other frontal one (ROI2).

LORETA software, an algorithm based on inverse problem solution that offers brain activated areas from EEG neural activity (Pascual-Marqui et al., 2002), was used to locate the sources of the ERP fluctuations that could be estimated as specific of pride and shame. Only the activity that could be considered as specific of these emotions, in consonance with ERP analyses, was used in these computations. These were made at selected time windows and based on the contrast values between pride or shame and the corresponding control condition.

5.3. Results

5.3.1. Self-reported emotions ratings

The ratings obtained immediately after the experiment (Table 1) showed the largest emotions of pride in the condition in which the participant was right and the 3 opponents were wrong (pride condition), compared to the corresponding control condition (both the participant and the opponents were right). This difference was significant (t (31) = 6.45, p < .001) [all *ps* are Bonferroni-corrected]. Emotions of shame, sadness and anger were similarly low in both conditions, with no significant difference between them (1.74 < ts (31) < 2.41, all ps > .1). Joy, in turn, was similarly high in both conditions, but it was not statistically significant (t(31)= 1.76, p>0.1).

	PRIDE	SHAME	JOY	SADNESS	ANGER
Pride condition	6.22 (1.34)	1.41 (0.89)	5.84 (1.6)	1.34 (0.81)	1.089 (0.52)
Control for Pride	4.69 (1.74)	1.03 (0.17)	5.34 (1.42)	1.12 (0.33)	1.25 (0.61)
Shame condition	1.28 (0.94)	4.15 (2.08)	1.19 (0.55)	3.22 (1.71)	3.94 (1.95)
Control for shame	1.78 (1.24)	2.22 (1.47)	2.15 (1.54)	2.28 (1.44)	3.09 (1.84)

Table 1: Ratings in a 7-point Likert scale (Mean, SD) regarding the feelings in different emotions as a function of feedback received

Regarding the wrong trials, the ratings showed larger feelings of shame in the condition in which the participant was wrong and the 3 opponents right (shame condition), compared with the corresponding control condition (both the participant and the opponents were wrong). This contrast was significant (t(31)= 7.42, p<.001). Feelings of pride and joy were similarly low in both conditions; however, even if slightly, they were significantly larger in the control condition (t(31)= 3, p=.05 for pride; t(31)= 9.87, p<.001 for joy). The values for sadness were relatively moderate in the two conditions, but significantly larger in the shame condition (t(31)= 3.11, p<0.05). Both conditions also exhibited moderate values in anger, not differing significantly (t(31)= 2.46, p>.1).

5.3.2. Electrophysiological data: Pride

Main results corresponding to pride are summarized in Figures 6-8. As can be seen in Fig. 6, the pride condition exhibited a fronto-central negativity around 300 ms and a long-lasting frontal positivity from around 500 to 900 ms. It can be appreciated that the earlier negativity exhibited a progressive amplitude increase, from the control to the pride condition, as an inverse function of the number of opponents being right. The late positivity, in turn, appeared as an all-or-none process, only present in the pride condition. The cluster-based permutation analyses (Fig. 7) in the early time window (0–500 ms) revealed a significant cluster between 250 and 350 ms, involving up to 44 electrodes, showing the largest F values around frontocentral and centroparietal electrodes. In a later time window (500–900 ms), the cluster-based permutation analysis revealed a significant cluster between around 600 and 900 ms involving virtually all electrodes**,** showing the largest F values around the frontal and frontocentral electrodes

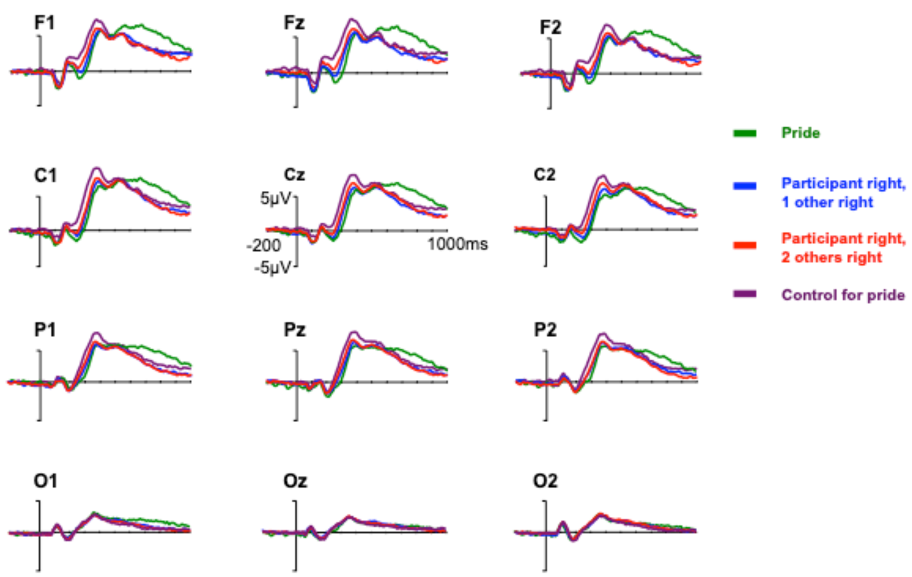


Figure 5.6. Mean average ERP fluctuations for pride conditions in a sample of electrodes.

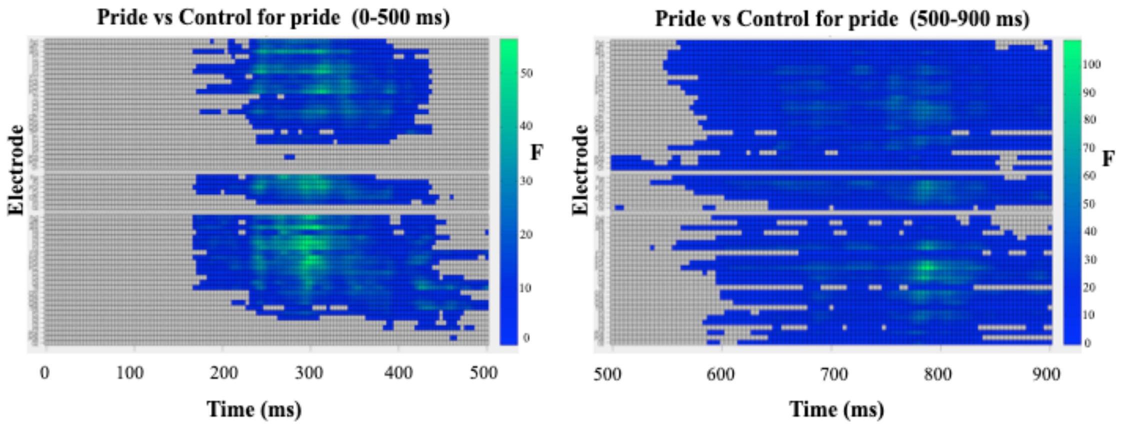


Figure 5.7. Results of cluster analyses for pride in the 0-500 (left) and 500-900 (right) time intervals for the comparison between the pride condition and its corresponding control condition. From top to bottom, electrodes are displayed orderly for the left hemisphere, midline and the right-hemisphere.

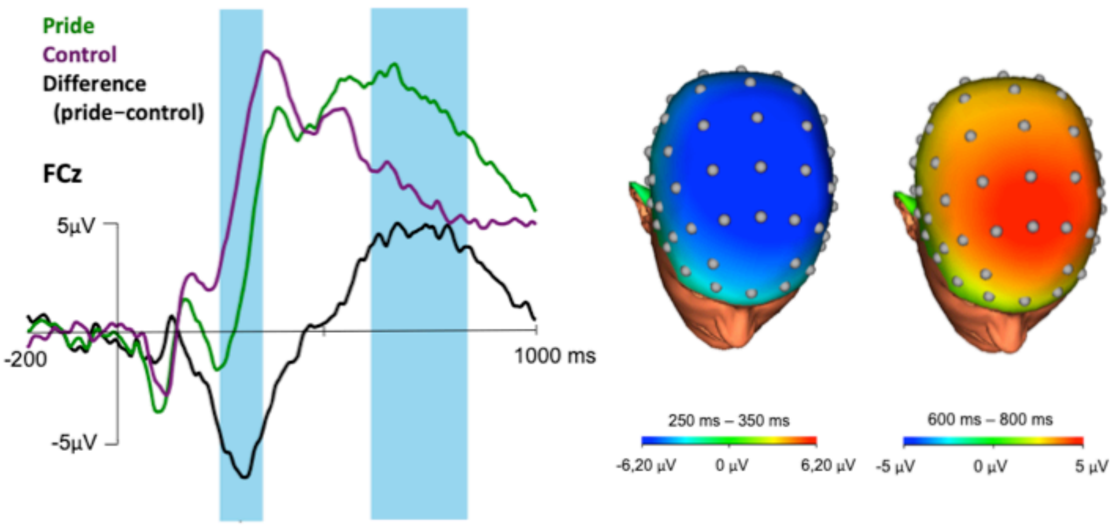


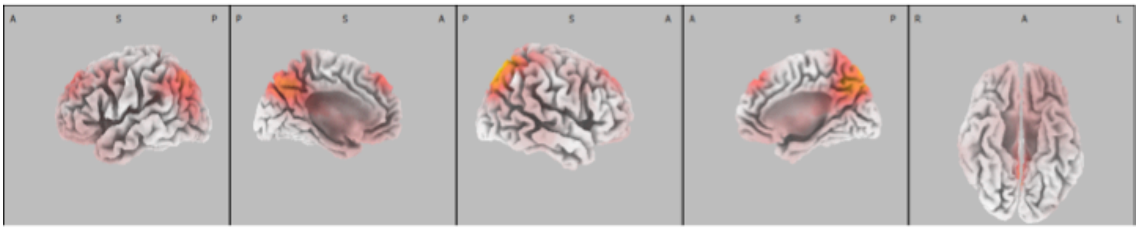
Figure 5.8. Main results of the comparison between the pride condition and its corresponding control condition. Left, ERP and difference wave resulting from this comparison at FCz electrode. Center and Right, topographic maps of the two main

fluctuations related to pride.

The cluster-based permutation analyses complement the data in Fig. 7, where the topographies of the early negativity and the late positivity are represented, substantiating the ROIs and time-windows employed in the ANOVA analyses. These revealed significant differences between the pride and its corresponding control condition, both in the early negativity [$F(1,31) = 62.9$; $p<.0001$; $\eta p^2 = .67$] as well as in the late positivity [$F(1,31) = 53.4$; $p<.0001$; $\eta p^2 = .633$]. The progressive amplitude increasing of the early negativity across the four conditions in which the participant was right seemed to justify an ANOVA analysis of this component comprising all these conditions. The results revealed main significant differences between conditions [$F(3,93) = 43.819$; $p<0.001$; $\eta p^2 = 0.586$], all post-hoc pair-wise Bonferroni-corrected comparisons being significant ($-7.93 < ts (31) < 8.18$, all $ps < .001$). Finally, neither the early negativity nor the late positivity correlated with the self-reported degree of pride felt in this condition (Spearman's R=0.14 and -0.13, respectively; p>0.1).

To estimate the neural source of the early negativity and late positivity related to pride, LORETA analyses were performed for the difference pride vs control around their corresponding peaks (250-350 and 600-800 ms time windows, respectively). The solution for the early negativity consisted of an involvement of the precuneus and the superior parietal regions and, with the best match at MNI coordinates X=20, y=-70, z=55 (Fig. 9, top). The solution for the late positivity involved the superior frontal and cingulate gyri, the best match being at MNI coordinates X=20, y=0, z=70 (Fig. 9, bottom). The solution also seemed to visibly involve right temporal regions.

Pride: Early component



Pride: Late component

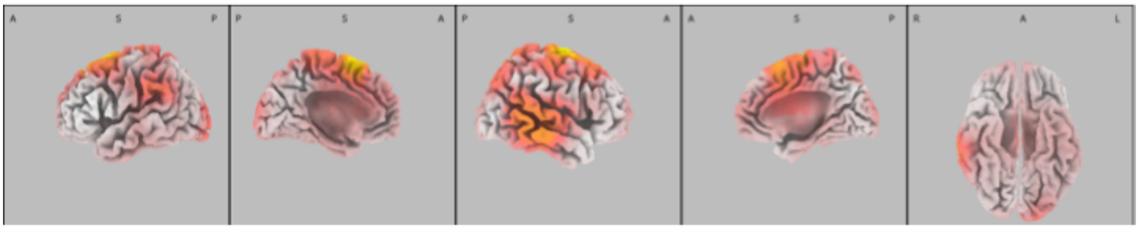


Fig. 5.9. Neural generators proposed by LORETA for the early negativity (top) and the late positivity (bottom) obtained in the pride condition.

5.3.3. Electrophysiological data: Shame

Main results corresponding to shame are summarized in Figures 10-13. Fig. 10 shows that the shame condition exhibited, when compared to the corresponding control condition, a frontal negativity around 370 ms and a long-lasing central positivity from around 600 to 1000 ms. The earlier negativity did not exhibit a progressive amplitude increase from the control to the pride condition across other conditions. Although not so remarkably as in pride, the late positivity appeared again as an all-or-none process, mainly present in the shame condition. The cluster-based permutation analyses during the early time segment (0–500 ms) revealed no significant cluster. Possibly, this originates in the tendency of this method to miss narrowly distributed effects occurring across a limited number of time points and electrodes (Groppe et al., 2011a); indeed, as will be noticed below, the ANOVA analysis was able to find significant effects within this period. During the late time segment (500–900 ms), the cluster-based permutation analyses (Fig. 12) revealed a significant cluster apparent mainly between around 700 and 900 ms involving up to 42 electrodes, with largest F values around frontocentral and centroparietal electrodes in both hemispheres, with a slight right tendency.

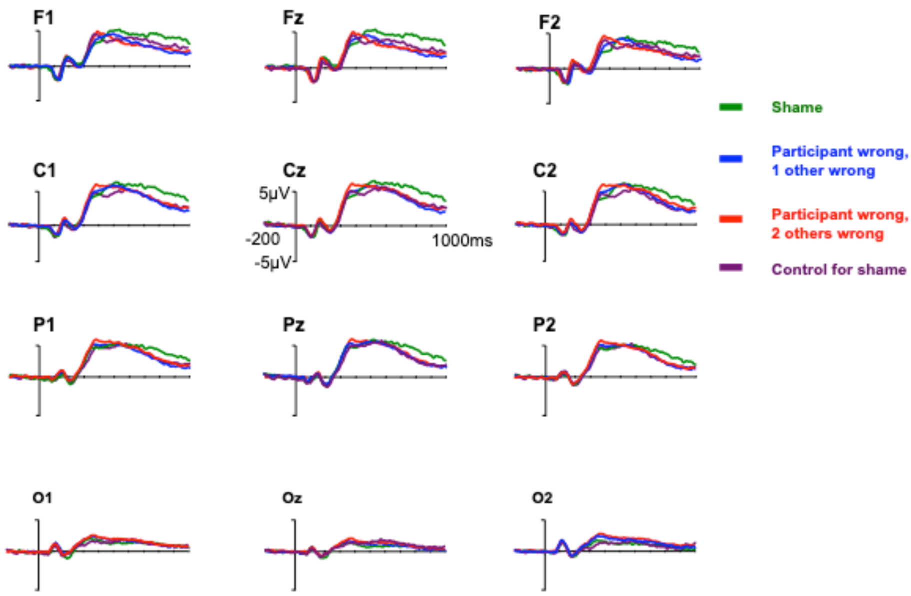


Figure 5.10. Mean average ERP fluctuations for shame conditions in a sample of electrodes.

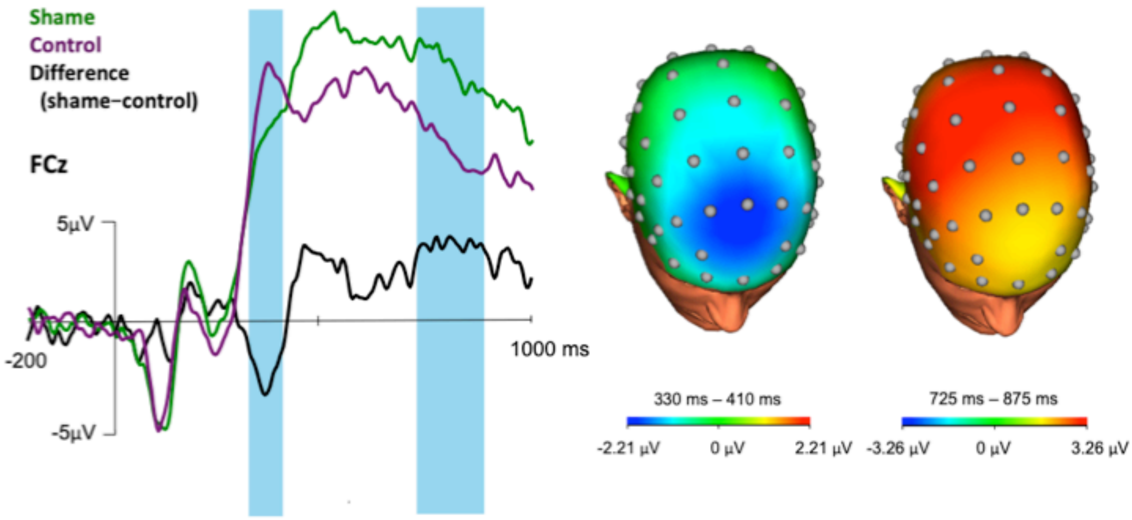


Figure 5.11. Main results of the comparison between the shame condition and its corresponding control condition. Left, ERP and difference wave resulting from this comparison at FCz electrode. Center and Right, topographic maps of the two main fluctuations related to shame.

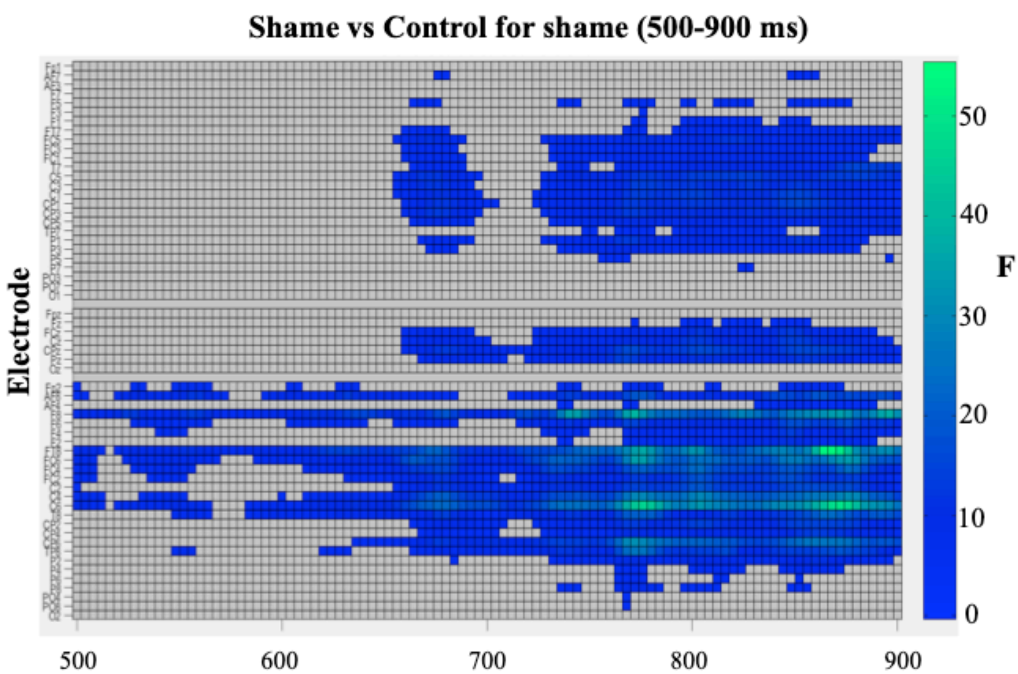
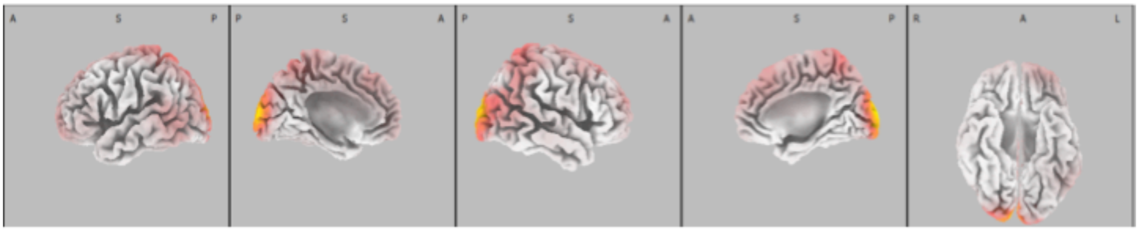


Figure 5.12. Results of cluster analyses for shame in the 0-500 (left) and 500-900 (right) time intervals for the comparison between the shame condition and its corresponding control condition. From top to bottom, electrodes are displayed orderly for the left hemisphere, midline and the right-hemisphere.

The data in Fig. 11, where the topographies of the early negativity and the late positivity are represented, are complemented by the cluster-based permutation analyses only for the late positivity, verifying the ROI and time-window employed in the ANOVA analyses for this component. The ROI and time-window used for the early negativity, in turn, was established solely by visual observation of the data. These analyses revealed significant differences between the shame and its corresponding control condition, both in the early negativity [$F(1,31) = 6.74$; $p<.014$; $\eta p^2 = .179$] as well as in the late positivity [$F(1,31) = 8.29$; $p<.007$; $\eta p^2 = .211$]. Neither the early negativity nor the late positivity correlated with the self-reported degree of shame (Spearman's R=-0.03 and -0.23, respectively; p>0.1).

The neural sources of the early negativity and the late positivity related to shame were estimated using LORETA analysis for the difference shame vs control around its peaks -i.e., 330-410 and 715-875 ms windows, respectively-. The solution for the early negativity involved the superior cuneus, with the best match at MNI coordinates X=15, y=-100, z=15 (Fig. 13, top). The solution for the late positivity involved the superior parietal lobe and precuneus, with the best match at MNI coordinates X=5, y=-70, z=55 (Fig. 13, bottom), though it seemed to extend to the cuneus. A possible secondary solution for this positivity involved left inferior frontal regions.

Shame: Early component



Shame: Late component

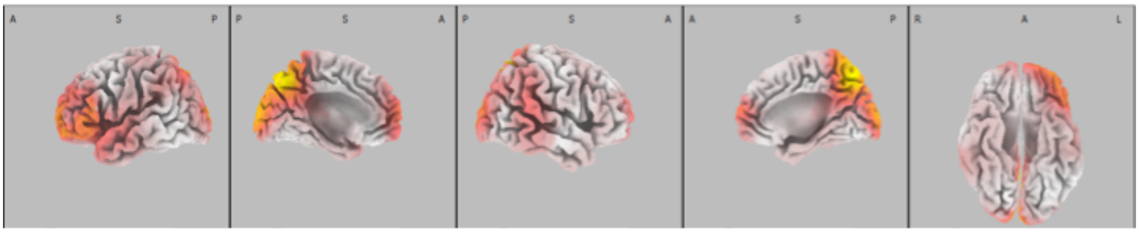


Fig. 13. Neural generators proposed by LORETA for the early negativity (top) and the late positivity (bottom) obtained in the shame condition.

5.4. Discussion

Brain electrical activity was recorded while participants performed a task in social context, with situations seemingly prompting feelings of pride and shame. The ERP technique used here permits a dynamic approach to the neural processes as these unfold over time, contributing therefore to better describe the neural underpinnings of these scarcely studied social emotions.

5.4.1. Prompting pride and shame in experimental environments

A main contribution of the present study is the setting of a procedure that permits studying pride and shame simultaneously and in experimental environments. The adaptation of a task previously used to prompt social emotions such as guilt or shame

(Leng et al. 2016; Zhu et al, 2017; Sánchez-García et al. 2019), has allowed the elicitation of pride and shame within the same participants and session, endorsing direct study and comparison of these emotions at the neural level in the same experiment. The feelings reported by the participants seem to indicate that the procedure was successful. The largest feelings of pride were reported in the pride condition, significantly larger than in the corresponding control condition. Other emotions (shame, sadness and anger) were low in either the pride or its control condition, with the exception of joy, which was equivalent in both conditions. In turn, the feelings exhibiting the largest values in the shame condition corresponded to shame, and although anger was also relatively high, the latter did not differ with the control condition. Sadness, which appeared also moderately present in the shame condition, was less rated than shame (this comparison was tested: $t(31)= 3.1$, $p<0.05$ after Bonferroni correction). Finally, even if shame and the corresponding control condition differed in pride and joy, all these values were always low. Overall, the comparison between shame and its control condition could yield relatively valid results, even if some cautions could be needed.

It is noticeable that the amplitude of none of the brain fluctuations obtained under these manipulations correlated with the self-reported feelings of either pride or shame. This would suggest that although our electrophysiological findings probably relate to processes essential to endorse these emotions, they are seemingly not reflecting the neural underpinnings of their hedonic component. This feature is of relevance for the interpretation of our results.

5.4.2. Neural Mechanisms involved in pride

The brain activity related to pride exhibited the expected dual-time pattern, consisting of an early negativity and a late positivity. The former, occurring around 300 ms after the presentation of the feedback, appeared to originate in medial parietal regions, mainly the precuneus. The late positivity emerged along a larger time interval, being visible between 600 and 900 ms and seemingly originated in medial frontal areas and the anterior cingulate. There is also a possible contribution of right temporal regions to this activity. These results harmonize well with previous research using neuroimaging (fMRI), in the sense of underscoring the involvement of medial prefrontal and posterior regions in situations of pride (Zahn et al., 2009; Simon-Thomas et al., 2012; Roth et al., 2014). The present findings also provide evidence on the differential involvement of anterior and posterior medial areas underlying pride across time. The first and early involvement of the medial parietal cortex appeared relevant to bear evaluative processes on the degree of individual success, as compared to the other 3 opponents. The amplitude of the early negativity to trials in which the participant was right increased stepwise as a function of the number of opponents being wrong. Among other processes, the medial parietal cortex and the precuneus have been traditionally linked to self-referential processing, particularly in situations in which one is compared to other people (e.g., Northoff et al., 2006; Qin & Northoff, 2011; Davey et al., 2016; Lou et al., 2004). It appears plausible, therefore, that this fluctuation is reflecting social comparison processes determining the degree of exceptionality of own performance in successful trials.

As a consequence of the computations reflected in the early negativity, the late positivity would emerge uniquely in conditions in which the outcome implied that oneself is the only right in the group. This situation was self-reported as maximally granting feelings of pride. The areas presumably contributing to this effect -medial frontal, anterior cingulate and right temporal regions- have all been involved in pride (Roth et al., 2014; Hong et al., 2019) and appear related to appraisals of social meaning and theory of mind (Takahasi, 2008; Schurz et al., 2014). This is compatible with the possibility that the late positivity reflects at least part of the processes underlying the emotion of pride. This interpretation, however, seems not straightforward, considering that a) pride was also appreciable in the control condition, even if with lower values than in the pride condition, while the late positivity was an all-or-none process, and b) its amplitude did not correlate with the reported feelings of pride, as commented above. It might be, accordingly, that the later modulation is rather reflecting the verification of the singularity of a successful trial, and this in social terms, in view of its apparent origins in brain regions associated with theory of mind and social meaning. Finally, it appears plausible that thereafter, and as an outcome of these processes, other, more basic emotion-related subcortical regions -such as the amygdala or the ventral striatum- become involved, this underlying the feelings of pride. However, these regions are not accessible through the ERP technique used here.

5.4.3. Neural mechanisms involved in shame

A dual-time pattern consisting of an early negativity and a late positivity was also found for shame. The early negativity was not supported by the cluster-based permutation analysis, though standard ANOVA analyses yielded significant results. Replication is pending. This fluctuation occurred around 370 ms after feedback presentation, and seemed originated in the cuneus. Interestingly, Zahn et al. (2014) reported that individuals with reduced volumes of the cuneus and the precuneus are more pride-prone, probably in consonance with the fact that shame and pride are categorically opposed emotions. The cuneus has been involved in mental imagery in social situations, as compared to when the same situations are imagined as occurring alone (Mochizuki et al., 2014). Overall, our results suggest that mental imagery of the social situation might be a critical process emerging soon in conditions engendering shame.

The late positivity related to shame was again a long-lasting activity, starting about 700 ms and visible until the end of the epoch. This fluctuation was apparently originated in the superior parietal lobe and the precuneus, also involving the cuneus and, secondarily, left inferior frontal regions. Interestingly, the main involved regions were those seemingly originating the early positivity linked to pride. It is tempting to reason, given its opposite electrical polarity, that it might be reflecting contrary processes to those reflected in the early negativity for pride, i.e., operations determining the degree of exceptionality of own performance, in this case in wrong trials, necessary to achieve feelings of shame. Though conceivable, this interpretation is nevertheless not clear-cut, since an ERP positivity is not necessarily the opposite of a negativity (Luck, 2005; Lopes da Silva & Niedermeyer, 2005). Moreover, this late positivity for shame and the early negativity for pride also differ in timing, duration and voltage topography. Actually, the late positivity for shame also seemed to engage the cuneus, already involved alongside the earlier component, while a secondary contribution appeared in the left inferior frontal regions, which have been related to emotion recognition in social perception (Keuken et al., 2011). Considering that its amplitude did not correlate with the reported feelings of shame, it can be speculated that this fluctuation might reflect late assessment processes of the social episode presumably necessary for triggering feelings of shame, the latter probably involving subcortical neural structures not accessible through the ERP technique.

5.4.4. Limitations of the present study

The main limitations of the present study are intrinsic to the ERP technique. On the one hand, some portions of the brain, particularly those subcortically situated and possibly underlying primary basic emotional areas, are not accessible to this methodology. Our results, therefore, outline only part of the brain structures involved in pride and shame, and must be considered as complementing data obtained with other neuroimaging technologies. On the other hand, the areas defined here have to be approached as most plausible candidates, i.e., the result of computations based on algorithms that attempt to overcome the inverse problem (Pascual-Marqui et al., 2002). Nonetheless, the areas defined here largely harmonize with extant literature on pride and shame using methodologies with better spatial resolution, while the present data provide outstanding temporal resolution.

Other limitation relates to the data concerning shame. According to self-reports, the comparison between shame and its corresponding control condition implies outstanding differences in shame, but other emotions can also be present, even if to a reduced degree. This contrasts with the more straightforward results for pride. Moreover, the early negativity for shame was not supported by the more conservative cluster-based permutation analysis. Further research is needed to elucidate the validity of our inferences and conclusions on this emotion.

5.4.5. Concluding remarks

Our results exhibit similarities and differences with previous literature. Some of the discrepancies are the consequence of using the ERP technique, while some others probably derive from prompting genuine feelings in laboratory conditions, at variance with less realistic procedures used previously. Overall, the available literature is notably scarce, and more research is granted. The brain regions seemingly involved in our data appear concerned in computations related to social cognition and, consequently, would be essential for the emergence of emotions that are directly defined by the position of the self in relation to others. In line with Roth et al. (2014) it appears plausible that the hedonic facets of pride and shame actually depend on primary basic emotional areas (e.g., amygdala, ventral striatum), the joint work of areas involved in social cognition being yet essential to establish and define these emotions. The present work suggests that the dynamic description of structures underlying social cognition can help to clarify their differential involvement in pride and shame, occasionally indistinguishable when more static neuroimaging techniques have been used. In conclusion, in pursuit of more accurate neural models of social emotions, more research is needed, and this should proceed integrating different available techniques.

Chapter 6. Discussion

The primary objective of our studies was to study the neurophysiological correlates of guilt, shame and pride in a social interaction in the lab. Our emphasis was placed on how these emotions unfold over time and to address similarities and differences among them. To achieve that objective, we conducted our three experiments.

In our *first study* we explored interpersonal guilt, by concurrently measuring ERP, neural source analysis (LORETA), skin conductance resistance (SCR), state- and trait-anxiety (STAI), and empathy (TECA) replicating Leng et al. (2016) study on a Spanish sample.

Our results confirmed the frontal long-lasting negativity (FN) in the self wrong, partner right condition (SW), starting at about 300 ms and peaking around 400 ms, and a Late Positive Complex (LPC) with parieto-occipital distribution, from 500 ms onwards being the former the one who hold statistical analyses. As the FN appeared in our Spanish sample we concluded that the processes reflected by the FN related to guilt and mediated by empathy seem similar across cultures under these circumstances, despite cultural dissimilarities found when experiencing social emotions (Walbott & Scherer, 1995; Anolli & Pascucci, 2005).

The rating of the feeling of guilt induced in participants in the SW condition was noticeably and significantly higher than that in the BW (both self and partner wrong) condition, confirming a behavioral study we previously run to assess the emotions elicited by the different conditions. As the FN did not correlate with the reported feelings of guilt, which would be extremely surprising if a complex emotion were reflected by a single and early ERP modulation, we addressed whether other processes may be involved. Based in our research with personal distress, overall levels of state – or trait – anxiety and measurements of empathy we conclude that the latter, which correlates with FN significantly, may reflect processes endorsing guilt. Indeed, empathy is a prerequisite for guilt to emerge as it relates to situations in which we may cause harm to others (Baumeister et al., 1994; Hoffman, 2000; Eisenberg, 2000; Morey et al, 2012). Our previous behavioral study and correlations with TECA scores reinforce the view of the FN as a fluctuation linked to interpersonal guilt and not to frustration due to the inherently difficult task. The LORETA analysis showed that dorsal medial prefrontal (mPFC) areas may be the source of the FN. This finding matches previous literature research of mPFC as a region for social cognition and empathy processes (Kédia et al., 2008; Morey et al., 2012; Bastin et al., 2016; Seitz et al., 2006).

Our *second study* explored shame in an interpersonal context, by concurrently measuring ERP, neural source analysis (LORETA), skin conductance resistance (SCR), shame and guilt variables (TOSCA-2), and empathy (TECA). As with the guilt experiment in Study 1, our main aim was to replicate and expand the study done by Zhu et al. (2017) with a Chinese sample. The dot estimation task economic game, with some adaptations to create the role of advicer-decider, was again used to elicit the social emotions of shame and guilt and reflect the main ERP modulations.

The rating score of shame feeling in the shame and guilt conditions were significantly higher than in the other conditions, supporting the procedure to induce this social emotion. Nevertheless, shame and guilt scores were noticeably similar. We presumed an overlapping and semantic confusion between participants to discriminate between the feeling and the appropriate word to express similar social emotions, somehow reflecting the daily mix-up with these feelings (Tangney and Dearing, 2002).

Two fluctuations were remarkable. The early shame latency appeared to be a centro parietal negativity (CPN) starting at about 200 ms and peaking around 400 ms, possibly reflecing a social devaluative pre-analysis, to protect ourself image and status

(De Hooge, et al., 2010). The late right frontal positivity, from 450 ms onwards, was different to Zhu et al. (2017), but nevertheless similar to the one reflected in our first study (Sánchez-García et al 2019), yielding a two-step processing (early and late) of social emotions. This right frontal positivity emerging at 450 ms may reflect reanalysis and self-assessment processes linked to social comparison (Qiu et al., 2010; Wu et al., 2012).

As happened with study 1 we could not find a significant correlation between the amplitude of either the centro parietal negativity nor the right frontal positivity and the reported feelings of shame in the shame (SW) condition, suggesting other processes might also be in play.

Although shame feelings may even relate to personal distress, this did not seem to be different across conditions, as reflected in SCR. The high arousal levels along the demanding dot estimation task might explain this result. The ERP shame results therefore do not seem related to distress or arousal.

The underlying sources of the EEG voltage patterns point to the medial prefrontal cortex (mPFC), part of the social cognition network reflecting what others might think about our performance and status, and the precuneus and the posterior cingulate cortex (PCC), which are highly involved in monitoring environmental changes and memory re-encoding processes (Andrews-Hanna et al. 2014; Kédia et al., 2008; Morey et al., 2012. The PCC, as proposed by Takahashi et al. (2004) participates in self-evaluative processes, showing more activation in the anterior and posterior cingulate cortex fMRI brain activation in shame minus guilt condition.

In sum, evidence found in study 2 seems highly consistent with the suggested dynamic processes of a centro-parietal negativity and a right frontal positivity and their contribution to the feelings of real shame through a social comparison task.

In our *third study* the participants performed the dot-estimation task used in previous studies (Leng et al. 2016; Zhu et al, 2017; Sánchez-García et al. 2019) but modified to a new social context along with other 3 online students and therefore increasing social exposure to prompt feelings of pride and shame. Our goal here is the elicitation of pride and shame within the same participants and session, and compare these emotions at the neural level in the same experiment. A behavioral study confirmed that the feelings reported by the participants along the experimental conditions were as expected.

As in the first two studies, the amplitude of the brain fluctuations did not correlate with the self-reported feelings, neither of pride nor of shame. ERP fluctuations for these emotions probably relate to essential processes in generating these emotions, but far from a causal-effect attribution.

The brain activity related to pride exhibited, as expected, a dual-time pattern consisting of an early negativity and a late positivity. The former, occurring around 300 ms after the presentation of the feedback, appeared to originate in the precuneus (Qin & Northoff, 2011; Davey et al., 2016; Lou et al., 2004). The late positivity was visible between 600 and 900 ms, with medial frontal areas and the anterior cingulate as possible sources. These results fit well with previous neuroimaging research (fMRI), involving medial prefrontal regions and the precuneus in situations of pride, both traditionally linked to self-referential processing in social comparison, among other functions (Zahn et al., 2009; Simon-Thomas et al., 2012; Roth et al., 2014; Takahasi, 2008; Schurz et al., 2014). This fluctuation may reflect social comparison processes of self successful trials.

The late positivity for feelings of pride presumably stems from medial frontal regions, anterior cingulate and right temporal regions, all involved being in appraisals of social meaning and theory of mind. This later modulation may reflect the verification of the singularity of a successful trial, giving a social meaning to this result through mentalizing.

Shame also exhibited the dual-time pattern of an early negativity and a late positivity. The former was not supported by the cluster-based permutation analysis, though standard ANOVA analyses yielded significant results. This early negativity occurred around 370 ms after feedback presentation, and seemed originated in the cuneus, which is usually involved in mental imagery in social situations (Mochizuki et al., 2014). The latter started about 700 ms until the end of the epoch, being apparently originated in the superior parietal lobe, the precuneus, the cuneus and, secondarily, the left inferior frontal regions. We observed that the late positivity related to shame implied regions also involved in originating the early negativity linked to pride, which therefore in the present case may be reflecting contrary processes, such as the degree of exceptionality of own performance, in wrong trials, necessary to achieve feelings of shame. Nevertheless, we assume this cautiously as they also differ in timing, duration and voltage topography.

Our secondary objective was to highlight the importance of social contexts in social emotions, explore the effects of social interactions on emotional processing, and design ecological paradigms to study social emotions in the lab. As we have proved, three different social contexts for the same dot estimation task lead the emergence of different social emotions. In order to elicit guilt we created an experimental design similar to Leng et al. (2016), based on a cooperative situation where, mediated by empathy, the participants may felt interpersonal guilt as a result of an economic loss attributed to one member of the couple. To try to disentangle shame and guilt, although semantically intertwined, we adapted and enhanced Zhu et al. (2017) paradigm, creating a situation in which the role of the advisor, a social one, results in shame or guilt, depending on the decision adopted by the decider. Based in these two scenarios, keeping in mind the importance of social evaluation, we addressed it in the study of pride and shame by creating a whole new setting in which participants' emotions could vary along a continuum from a shame and devaluative feeling to a pride and successful achievement. We, therefore, manifest and emphasize that social contexts and interactions are key elements on emotional processing. These social contexts and interactions are present, to some degree and extent, in the lab situation. This intended similarity to real life situations provides more valid results, and the conclusions obtained are more rooted than previous designs based on imagining or remembering emotions.

Altogether, the results of this thesis demonstrate that the neurophysiological dynamics of the social emotions of guilt, shame and pride imply a dual time process, early and late. This two-step process reflects their complex nature, linked to mentalizing, social cognitions networks and involvement of the self in emotional processing. The social emotions, as opposed to the basic ones, have no universal and external elicitor. Although we may correlate some situations to the occurrence of a particular social emotion, it is always the way we think, assess and view ourselves the cognitive startup of guilt, shame and pride. They, therefore, require the mental representation of the self along with the knowledge and responsibility about standards, rules and goals of our social environments.

Our results exhibit some similarities and differences with previous literature. Being the available literature really scarce, the results may be considered under the need to do more research. The regions that in which these social emotions may stem in our studies, are reinforcing the results of previous studies that show medial cortical areas, such as the posterior and the anterior cingulate, the medial prefrontal cortex and the precuneus, to be the major structures involved in these social emotions.

This two-step neurophysiological dynamic process does not straightforwardly reflect the studied social emotion in each experiment. As reported feelings of guilt, shame and pride do not correlate with the amplitudes of the ERP´s fluctuations found, other processes are necessarily involved and further research is needed.

Our results agree with a conception of social emotions in which the predictive, instead of reactive, nature of our social brain reflects an emotion constructivism view (Barret, 2017) and a broad semantic space for them to emerge (Koide-Majima, 2020).

Limitations of our studies

One of the main limitations of our studies is intrinsic to the ERP technique. Despite its optimal time-course resolution, subcortical regions are not accessible to this methodology. Areas such as the amygdala, the hypothalamus, the basal ganglia, the ventral tegmental area or the insular cortex play a key role in emotional processing (Roth et al. 2014). Our results, therefore, highlight only part some of the brain structures involved in social emotions.

Our spatial approach to study the possible sources of brain areas participating in guilt, shame and pride is based in inverse problems algorithms (Pascual-Marqui et al., 2002). Although the areas defined in our three studies match the ones extensively exposed in the literature using technologies with high spatial resolution we are cautious about the accuracy and validity of our data. It appears necessary to combine the approach used here with other neuroimaging techniques.

Inherent to this field, we are still far from knowing what an emotion is in the brain. It is plausible that there is no such a discrete distinction between emotions and thoughts, basic or social, and any intent to search for a fingerprint of emotions is, in its roots, incompatible with the dynamic, hyperconnectivity and high variability of the neural processes.

Chapter 7. Conclusions

1. Our results indicate that a frontal ERP modulation plausibly reflects empathic processes necessary for a final outcome of feelings of interpersonal guilt, though not the feelings of guilt themselves. Those processes are presumably occurring in dorsal mPFC areas.
2. Our results suggest that shame emotional content processes imply an early ERP modulation that extends to late reanalysis. These processes do not relate to the feelings of shame themselves. Those processes are presumably occurring at the precuneus areas.
3. Our results indicate that the dual-time pattern consisting of an early negativity and a late positivity for pride and shame is replicable. The brain regions seemingly involved here apply to social cognition and, consequently, how we place ourselves in relation to others.
4. Our results show that it is possible and advisable to prompt genuine feelings in laboratory conditions, at variance with less realistic procedures used previously, such as imagining or remembering.

References

Abe, J. A. A., & Izard, C. E. (1999). The developmental functions of emotions: An analysis in terms of differential emotions theory. *Cognition & emotion*, *13*(5), 523–549. https://psycnet.apa.org/doi/10.1080/026999399379177

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669-672.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science*, 18, 524–530.

Anderson, S., Bechara, A., Damasio, H. *et al.* Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nat Neurosci* **2,** 1032–1037 (1999). https://doi.org/10.1038/14833

Anolli, L. & Pascucci, P. (2005). Guilt and guilt-proneness, shame and shame-proneness in Indian and Italian young adults. *Personality and Individual Differences 39* (4), 763-773.

Awan, F. G.; Saleem, O. & Kiran, A. (2019). Recent trends and advances in solving the inverse problem for EEG source localization. *Inverse problems in science and engineering, 27*, 1521–1536. https://doi.org/10.1080/17415977.2018.1490279

Bastin, C., Harrison, B. J., Davey, C. G., Moll, J., & Whittle, S. (2016). Feelings of shame, embarrassment and guilt and their neural correlates: A systematic review. *Neuroscience & biobehavioral reviews*, *71*, 455–471. https://doi.org/10.1016/j.neubiorev.2016.09.019

Basile B., Mancini F., Macaluso E., Caltagirone C., Frackowiak R. S., & Bozzali M. (2011). Deontological and altruistic guilt: Evidence for distinct neurobiological substrates. *Human Brain Mapping, 32* (2), 229- 239.

Barrett, H.C. (2014) The Shape of Thought: How Mental Adaptations Evolve, Oxford University Press

Barrett, L. F. (2012). Emotions are real. *Emotion*, *12*(3), 413.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion*, *23*(7), 1284-1306.

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, *15*(2), 79-83.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Social cognitive and affective neuroscience, 12(1), 1-23.

Baumeister, R., Stillwell, A.M., Heatherton, T.F. (1994). Guilt: an interpersonal approach. *Psychological Bulletin, 115*(2), 243.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior*, *52*(2), 336-372.

Beer, J. S., & Keltner, D. (2004). What is unique about self-conscious emotions?. *Psychological inquiry*, *15*(2), 126–129. http://www.jstor.org/stable/20447214

Benedek, M. & Kaernbach, C. (2010a). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*(4)*,* 647-658.

Benedek, M. & Kaernbach, C. (2010b). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190* (1), 80-91.

Brusini, P., Dehaene-Lambertz, G., van Heugten, M., de Carvalho, A., Goffinet, F., Fievet, A. C., & Christophe, A., (2017). Ambiguous function words do not prevent 18-month-olds from building accurate syntactic category expectations: An ERP study. *Neuropsychologia, 98*, 4–12. https://doi.org/10.1016/j.neuropsychologia.2016.08.015

Cannon, W. B. (1987). The James-Lange theory of emotions: a critical examination and an alternative theory. *The American journal of psychology*, *100*(3/4), 567-586.

Carver, C. S., Sinclair, S., & Johnson, S. L. (2010). Authentic and hubristic pride: Differential relations to aspects of goal regulation, affect, and self-control. *Journal of research in personality*, *44*(6), 698–703. https://doi.org/10.1016/j.jrp.2010.09.004

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, *70* (3), 560-572.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Clark, L., Bechara, A., Damasio, H., Aitken, M. R. F., Sahakian, B. J., & Robbins, T. W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. *Brain*, *131*(5), 1311-1322.

Cory, G. A. (2002). Reappraising MacLean's triune brain concept. *The evolutionary neuroethology of Paul MacLean. Convergences and frontiers. Prager, Westport, CT, London*, 9-27.

Davey, C. G., Pujol, J., & Harrison, B. J. (2016). Mapping the self in the brain's default mode network. *NeuroImage, 132*, 390–397. https://doi.org/10.1016/j.neuroimage.2016.02.022

De Ridder, D., & Vanneste, S. (2013). The predictive brain and the "free will" illusion. *Frontiers in psychology*, *4*, 131.

Dror, O. E. (2014). The Cannon–Bard thalamic theory of emotions: A brief genealogy and reappraisal. *Emotion Review*, *6*(1), 13-20.

Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press, USA

De Hooge, I. E., Nelissen, R., Breugelmans, S. M., & Zeelenberg, M. (2011). What is moral about guilt? Acting "prosocially" at the disadvantage of others. *Journal of personality and social psychology*, *100*(3), 462.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of cognitive Neuroscience*, *24*(8), 1742-1752.

Etxebarria, I. & Apodaca, P. (2008). Both Freud and Hoffman are right: anxious aggressive and empathic dimensions of guilt. *The Spanish Journal of Psychology, 11* (1), 159-171.

Etxebarria, I., Ortiz, M.J., Conejero, S. & Pascual, A. (2009). Intensity of habitual guilt in men and women: differences in interpersonal sensitivity and the tendency towards anxious-aggressive guilt. *The Spanish Journal of Psychology, 12* (2), 540-554.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51* (1), 665-697.

Eisenberger, N.I. et al. (2011) The neural sociometer: brain mechanisms underlying state self-esteem. J. Cogn. Neurosci. 23, 3448–3455

Fields, E. C., & Kuperberg, G. R. (2019). Having your cake and eating it too: Flexibility and power with mass univariate statistics for ERP data. *Psychophysiology, 57*(2), e13468. https://doi.org/10.1111/psyp.13468

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and human behavior, 26*(1), 10–46. https://doi.org/10.1016/j.evolhumbehav.2004.08.005

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science*, *18*(12), 1050-1057.

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., & Mayberg, H. (2003). In search of the emotional self: an fMRI study using positive and negative emotional words. *American Journal of Psychiatry*, *160*(11), 1938-1945.

Fernández-Pinto, I., López-Pérez, B., & Márquez, M. (2008). Empatía: Medidas, teorías y aplicaciones en revisión. *Anales de psicología*, *24*(2).

Fliessbach, K.; Weber, B.; Trautner, P.; Dohmen, T.; Sunde, U.; Elger, C.E. & Falk, A. (2007). Social Comparison Affects Reward-Related Brain Activity in the Human Ventral Striatum. *Science 318* (5854), 1305-1308.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138.

Frith, C. D. (2014). Action, agency and responsibility. *Neuropsychologia*, *55*, 137-142.

Frith, C. D. (2007). The social brain?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 671-678.

Gilead, M., Katzir, M., Eyal, T., & Liberman, N. (2016). Neural correlates of processing “self-conscious” vs “basic” emotions. *Neuropsychologia*, *81*, 207–218. https://doi.org/10.1016/j.neuropsychologia.2015.12.009

Giner-Sorolla, R., Piazza, J., & Espinosa, P. (2011). What do the TOSCA guilt and shame scales really measure: Affect or action? *Personality And Individual Differences*, *51* (4), 445-450.

Gilbert, P. (2000). The relationship of shame, social anxiety and depression: The role of the evaluation of social rank. *Clinical Psychology & Psychotherapy, 7*(3), 174–189. https://doi.org/10.1002/1099-0879(200007)7:3<174::AID-CPP236>3.0.CO;2-U

Gillihan, S. J., & Farah, M. J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological bulletin*, *131*(1), 76.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., ... & Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of neuroengineering and rehabilitation*, *5*(1), 25.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology, 48*(12), 1711-1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology, 48*(12), 1726–1737. https://doi.org/10.1111/j.1469-8986.2011.01272.x

Hari, R., & Puce, A. (2017). *MEG-EEG Primer*. Oxford University Press.

Hoemann, K., Devlin, M., & Barrett, L. F. (2020). Emotions are abstract, conceptual categories that are learned by a predicting brain. *Emotion Review*. DOI: 10.1177/1754073919897296.

Hong, Y. J., Park, S., Kyeong, S., & Kim, J. J. (2019). Neural basis of professional pride in the reaction to uniform wear. *Frontiers in human neuroscience, 13*, 253. https://doi.org/10.3389/fnhum.2019.00253

Haidt, J. (2003). The moral emotions. *Handbook of Affective Sciences, 11* (2003), 852–870.

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology*, *84*(3), 451-462.

Hoffman, M. L. (2001). *Empathy and moral development implications for caring and justice.* Cambridge: University Press.

Holbrook, C., Piazza, J., & Fessler, D. M. (2014). Conceptual and empirical challenges to the "authentic" versus "hubristic" model of pride. *Emotion*, *14*(1), 17.

Jiménez-Ortega, L., Ramos-Badaya, E., Hernández-Gutiérrez, D., Silvera, M., Espuny, J., Garcia, J. S., ... & Martín-Loeches, M. (2020). Effects of reader's facial expression on syntactic processing: A brain potential study. *Brain Research*, 146745.

Johnson, M. H., Griffin, R., Gergely Csibra, H. H., Farroni, T., De Haan, & Richards, J. (2005). The emergence of the social brain network: Evidence from typical and atypical development. *Development and psychopathology*, *17*(3), 599.

Keuken, M. C., Hardie, A., Dorn, B. T., Dev, S., Paulus, M. P., Jonas, K. J., … & Pineda, J. A. (2011). The role of the left inferior frontal gyrus in social perception: an rTMS study. *Brain research, 1383*, 196–205. https://doi.org/10.1016/j.brainres.2011.01.073

Kincses WE, Braun C, Kaiser S, Elbert T (1999) Modeling extended sources of event-related potentials using anatomical and physiological constraints. Hum Brain Mapp 8:182–193.

Koide-Majima, N., Nakai, T., & Nishimoto, S. (2020). Distinct dimensions of emotion in the human brain and their representation on the cortical surface. *NeuroImage*, *222*, 117258.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological psychology*, *11*(2), 99-116.

Kédia G., Berthoz S., Wessa M., Hilton D. & Martinot J.L. (2008). An agent harms a victim: A functional magnetic resonance imaging study on specific moral emotions. *Journal of cognitive neuroscience, 20* (10), 1788–1798.

Kiehl, K. A. (2006). A cognitive neuroscience perspective on psychopathy: Evidence for paralimbic system dysfunction. *Psychiatry research*, *142*(2-3), 107-128.

LaBar, K. S., & LeDoux, J. E. (1996). Partial disruption of fear conditioning in rats with unilateral amygdala damage: correspondence with unilateral temporal lobectomy in humans. *Behavioral neuroscience*, *110*(5), 991.

Larsen, R. J., & Diener, E. (1992). *Promises and problems with the circumplex model of emotion.* In M. S. Clark (Ed.), *Review of personality and social psychology, No. 13. Emotion* (p. 25–59). Sage Publications, Inc.

LeDoux, J. (2012). A neuroscientist's perspective on debates about the nature of emotion. *Emotion Review*, *4*(4), 375-379.

Leng, B., Wang, X., Cao, B., & Li, F. (2017). Frontal negativity: An electrophysiological index of interpersonal guilt. *Social neuroscience*, *12*(6), 649-660.

Lewis, M. (2000). Self-conscious emotions: Embarrasment, pride, shame, and guilt. In M- Lewis and J.M. Haviland-Jones (Eds.), *Handbook of Emotions*. New York: Guilford Press, pp. 573-691.

Lieberman, D. et al. (2007) The architecture of human kin detection. Nature 445, 727–731

Liu, F., Liu, G., & Lai, X. (2014). Emotional intensity evaluation method based on Galvanic skin response signal. In *2014 Seventh International Symposium on Computational Intelligence and Design* (ISCID), (Vol. 1, pp. 257-261). IEEE.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., ... & Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of cognitive neuroscience*, *22*(7), 1623-1635.

Lopes da Silva, F.; Niedermeyer, E. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Lippincott Williams & Wilkins.

Lou, H. C., Luber, B., Crupain, M., Keenan, J. P., Nowak, M., Kjaer, T. W., Sackeim, H. A., & Lisanby, S. H. (2004). Parietal cortex and representation of the mental Self. *Proceedings of the National Academy of Sciences of the United States of America, 101*(17), 6827–32. https://doi.org/10.1073/pnas.0400049101

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique.* MIT Press.

Luck, S. J., & Hillyard, S.A. (1984) Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*(3), 291–308. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x

Makeig, S., Enghoff, S., Jung, T. P. & Sejnowski, T. J. (2000). Moving-window ICA decomposition of EEG data reveals event-related changes in oscillatory brain activity. In *2nd Int. Workshop on Independent Component Analysis and Signal Separation* (pp. 627-632).

Mclatchie, N., Giner-Sorolla, R., & Derbyshire, S. W. G. (2016). Imagined guilt' versus 'recollected guilt': Implications for fMRI. *Social cognitive and affective neuroscience, 11* (5), 703–711.

Menesini, E., & Camodeca, M. (2008). Shame and guilt as behaviour regulators: Relationships with bullying, victimization and prosocial behaviour. *British journal of developmental psychology*, *26*(2), 183–196. https://psycnet.apa.org/doi/10.1348/026151007X205281

Michel CM, Grave de Peralta R, Lantz G, Gonzalez Andino S, Spinelli L, Blanke O, Landis T, Seeck M (1999) Spatiotemporal EEG analysis and distributed source estimation in presurgical epilepsy evaluation. J Clin Neurophysiol 16:239–266

Michl, P., Meindl, T., Meister, F., Born, C., Engel, R. R., Reiser, M., & Hennig-Fast, K. (2014). Neurobiological underpinnings of shame and guilt: a pilot fMRI study. *Social cognitive and affective neuroscience*, *9*(2), 150–157. https://doi.org/10.1093/scan/nss114

Monteith, M.J. (1993). Self-regulation of stereotypical responses: Implications for progress in prejudice reduction. *Journal of Personality and Social Psychology, 65, 469-485.*

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of personality and social psychology*, *83*(5), 1029.

Morey, R. A., McCarthy, G., Selgrade, E. S., Seth, S., Nasser, J. D., & LaBar, K. S. (2012). Neural systems for guilt from actions affecting self versus others. *Neuroimage*, *60* (1), 683-692.

Mochizuki, A. A., Sudo, M.M., Kirino, E, & Itoh, K. (2014). Brain activation associated with motor imagery of coordination exercises and social abilities. *European journal of sport science, 14*(7), 671–677. https://doi.org/10.1080/17461391.2014.893019

Northoff, G., Heinzel, A., De Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Näätänen and Tiitinen, 1998

Nelissen, R., & Zeelenberg, M. (2009). When guilt evokes selfpunishment: Evidence for the existence of a Dobby Effect. *Emotion, 9* (1), 118.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

Oldfield, R.C. (1971). The assessment and analysis of handiness: the Edinburgh inventory. *Neuropsychologia, 9*(1) 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Papez, J. W. (1937). A proposed mechanism of emotion. *Archives of Neurology & Psychiatry*, *38*(4), 725-743.

Pascual-Marqui R. D., Esslen M., Kochi K., & Lehmann D. (2002). Functional imaging with low resolution brain electromagnetic tomography (LORETA): a review. *Methods and findings in experimental and clinical pharmacology,* 24 Suppl C, 91–95.

Pascual-Marqui RD, Esslen M, Kochi K, Lehmann D (2002) Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. Methods Find Exp Clin Pharmacol 24(Suppl C):91–95

Peterson N.N, Schroeder C. E, Arezzo J. C, et al. Neural generators of early cortical somatosensory evoked potentials in the awake monkey. Electroencephalogrphy and Clinical Neurophysiology. 1995;96:248–260.

Phillips C, Rugg MD, Friston KJ (2002) Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. Neuroimage 16:678–695

Phillips C, Rugg MD, Fristont KJ (2002) Systematic regularization of linear inverse solutions of the EEG source localization problem. Neuroimage 17:287–301

Pascual-Marqui R.D., Esslen M., Kochi K., & Lehmann D. (2002). Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods and Ffindings in Experimental & Clinical Pharmacology, 24* (Suppl C), 91-95.

Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage, 57*(3), pp. 1221–1233. https://doi.org/10.1016/j.neuroimage.2011.05.028

Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of personality and social psychology*, *79*(2), 286.

Robertson, T. E., Sznycer, D., Delton, A. W., Tooby, J., & Cosmides, L. (2018). The true trigger of shame: Social devaluation is sufficient, wrongdoing is unnecessary. *Evolution and human behavior*, *39*(5), 566–573. https://doi.org/10.1016/j.evolhumbehav.2018.05.010

Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional–anatomic organization in right hemisphere. *Brain and language*, *104*(1), 51-74.

Roth, L., Kaffenberger, T., Herwig, U., & Brühl, A.B. (2014). Brain activation associated with pride and shame. *Neuropsychobiology, 69*(2), 95–106. https://doi.org/10.1159/000358090

Sánchez-García, J., Espuny, J., Hernández-Gutiérrez, D., Casado, P., Muñoz, F., Jiménez-Ortega, L., ... & Martín-Loeches, M. (2019). Neurofunctional characterization of early prefrontal processes contributing to interpersonal guilt. *Cognitive, affective, & behavioral neuroscience*, *19*(5), 1192–1202. https://doi.org/10.3758/s13415-019-00748-6

Scherer, K. (2009). The dynamic architecture of emotion:Evidence for the component process model. Cognition and Emotion, 23, 1307-1351.

Scherer, K. R. (2005). What are emotions? And how can they be measured? Social Science Information, 4(44), 293-317.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience and biobehavioral reviews, 42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Shaw, P., Lawrence, E. J., Radbourne, C., Bramham, J., Polkey, C. E., & David, A. S. (2004). The impact of early and late damage to the human amygdala on 'theory of mind'reasoning. *Brain*, *127*(7), 1535-1548.

Simon-Thomas, E. R., Godzik, J., Castle, E., Antonenko, O., Ponz, A., Kogan, A., & Keltner, D. J. (2012). An fMRI study of caring vs self-focus during induced compassion and pride. *Social cognitive and affective neuroscience,* 7(6), 635–648. https://doi.org/10.1093/scan/nsr045

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J., & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropsychologia*, *41*(2), 209-220.

Sutton, S. (1979). P300--thirteen years later. In *Evoked brain potentials and behavior* (pp. 107-126). Springer, Boston, MA.

Seitz, R. J., Nickel, J., & Azari, N. P. (2006). Functional modularity of the medial prefrontal cortex: involvement in human empathy. *Neuropsychology*, *20*(6), 743.

Shin L.M., Dougherty D.D., Orr S.P., Pitman R.K., Lasko M., Macklin M.L., Alpert N.M., Fischman A.J., & Rauch S.L. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological psychiatry 48* (1), 43–50.

Singer, T., Seymour B., O´Doherty, J., Kaube, H., Dolan, R.J. & Frith, C.D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science 303* (5661), 1157-67.

Silfver M. and Helkana, K. (2007). Empathy, guilt and gender: a comparison of two measures of guilt. *Scandinavian Journal of Psychlogy, 48* (3), 239-246.

Skapinakis P. (2014). Spielberger State-Trait Anxiety Inventory. In: Michalos A.C. (eds) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht

Spielberger, C. D. (1983). Manual for the State-Trait Anxiety Inventory STAI (form Y)(" self-evaluation questionnaire").

Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial psychiatry journal*, *18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K. & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *Neuroimage, 23* (3), 967–74.

Tangney, J. P., & Dearing, R. L. (2002). Emotions and social behavior. Shame and guilt. New York, NY, US.

Takahashi, H., Matsuura, M., Koeda, M., Yahata, N., Suhara, T., Kato, M., & Okubo, Y. (2008). Brain activations during judgments of positive self-conscious emotion and positive basic emotion: pride and joy. *Cerebral cortex*, *18*(4), 898–903. https://doi.org/10.1093/cercor/bhm120

Tangney, J.P., & Dearing, R.L. (2003). *Shame and Guilt*. Guilford Press.

Tangney, J. P., Wagner, P. E., Fletcher, C., & Gramzow, R. (1992). Shamed into anger? The relation of shame and guilt to anger and self-reported aggression. *Journal of personality and social psychology, 62*(4), 669–675. https://doi.org/10.1037//0022-3514.62.4.669

Tangney, J. P. (1995). Recent advances in the empirical study of shame and guilt. *American Behavioral Scientist*, *38*(8), 1132-1145.

Tangney, J. P., Miller, R., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions?. *Journal of personality and social psychology, 70*(6), 1256–1269. https://doi.org/10.1037//0022-3514.70.6.1256

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual review of psychology, 58*, 345–372. https://doi.org/10.1146/annurev.psych.56.091103.070145

Tiitinen, H., May, P., & Näätänen, R. (1997). The transient 40-Hz response, mismatch negativity, and attentional processes in humans. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *21*(5), 751-771.

Tracy, J.L., & Randles, D. (2011). Four models of basic emotions: a review of Ekman and Cordaro, Izard, Levenson,and Panksepp and Watt. *Emotion review, 3*(4), 397–405. https://doi.org/10.1177/1754073911410747

Tracy, J. L., & Robins, R. W. (2007). The psychological structure of pride: a tale of two facets. *Journal of personality and social psychology*, *92*(3), 506-525 https://doi.org/10.1037/0022-3514.92.3.506

Tracy, J. L., Robins, R. W., & Tangney, J. P. (Eds.). (2007). *The self-conscious emotions: Theory and research*. Guilford Press.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564-584.

Verduyn, P., Delaveau, P., Rotgé, J. Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review*, *7*(4), 330-335.

Williams, K.D. and Nida, S.A. (2011) Ostracism: consequences and coping. Curr. Dir. Psychol. Sci. 20, 71–75

Wallbott, H. G., & Scherer, K. R. (1995). Cultural determinants in experiencing shame and guilt. In J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 465-487). New York, NY, US: Guilford Press.

Wertheim, E.H. & Schwartz, J.C. (1983). Depression, guilt and self-management of pleasant and unpleasant events. *Journal of Personality and Social Psychology, 45*, 884-889. Doi:10.1037/0022-3514.45.4.884

Whitehouse, H., Francois, P., Savage, P. E., Currie, T. E., Feeney, K. C., Cioni, E., ... & Ter Haar, B. (2019). Complex societies precede moralizing gods throughout world history. *Nature*, *568*(7751), 226-229.

Yao D, He B (2001) A self-coherence enhancement algorithm and its application to enhancing three-dimensional source estimation from EEGs. Ann Biomed Eng 29:1019–1027

Yik, M. S., Russell, J. A., Ahn, C. K., Dols, J. M. F., & Suzuki, N. (2002). Relating the five-factor model of personality to a circumplex model of affect. In *The five-factor model of personality across cultures* (pp. 79-104). Springer, Boston, MA.

Yu, H., Hu, J., Hu, L., & Zhou, X. (2014). The voice of conscience: Neural bases of interpersonal guilt and compensation. *Social Cognitive and Affective Neuroscience, 9*(8), 1150–1158.

Zahn R., de Olivera-Souza, R., Bramati I., Garrido, G., & Moll, J. (2009). Subgenual cingulated activity reflects individual differences in empathic concern. *Neuroscience letters 457* (2), 107–110.

Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E. D., & Grafman, J. (2009). The neural basis of human social values: evidence from functional MRI. *Cerebral cortex,* 19(2), 276–283. https://doi.org/10.1093/cercor/bhn080

Zahn, R., Garrido, G., Moll, J., & Grafman, J. (2014). Individual differences in posterior cortical volume correlate with proneness to pride and gratitude. *Social cognitive and affective neuroscience, 9*(11), 1676–1683. https://doi.org/10.1093/scan/nst158

Zhu, R., Wu, H., Xu, Z., Tang, H., Shen, X., Mai, X., & Liu, C. (2017). Early distinction between shame and guilt processing in an interpersonal context. *Social neuroscience, 14*(1), 1–14. https://doi.org/10.1080/17470919.2017.1391119

Zhu, R., Feng, C., Zhang, S., Mai, X., & Liu, C. (2019). Differentiating guilt and shame in an interpersonal context with univariate activation and multivariate pattern analyses. *NeuroImage*, *186*, 476–486. https://doi.org/10.1016/j.neuroimage.2018.11.012