



Neurofunctional characterization of early prefrontal processes contributing to interpersonal guilt

Jose Sánchez-García¹ · Javier Espuny¹ · David Hernández-Gutiérrez¹ · Pili Casado^{1,2} · Francisco Muñoz^{1,2} · Laura Jiménez-Ortega^{1,2} · Sabela Fondevila^{1,2} · Manuel Martín-Loeches^{1,2}

Published online: 28 August 2019
© The Psychonomic Society, Inc. 2019

Abstract

Guilt is a social emotion that plays a central role in promoting prosocial behavior. Despite its relevance, it remains poorly understood. The present study aimed to fill this gap by verifying and characterizing a frontal negative fluctuation of the event-related brain potentials (ERP) emerging in conditions of interpersonal guilt. Paired participants would earn money if both performed correctly a dot estimation task (both right); otherwise, both would lose a similar amount (self wrong, partner wrong, and both wrong conditions). The reported feeling of guilt was noticeable in the self wrong condition, which yielded a frontal negativity between 300 and 500 ms after the onset of performance feedback. The amplitude of this fluctuation, however, did not correlate with the amount of guilt reported by the participants, whereas both these values did so with standard measures of empathy. Neither anxiety (trait or state) nor arousal (skin conductance response) seemed to relate to this negativity. A neural source (LORETA) analysis established its generators in the dorsal medial prefrontal cortex (mPFC), a region linked to guilt in fMRI studies but also, importantly, to empathy. The frontal negative fluctuation thus might reflect empathic processes contributing to achieve feelings of interpersonal guilt.

Keywords Interpersonal guilt · Frontal negativity · LORETA · ERP · Medial prefrontal cortex · Empathy

The study of interpersonal guilt has recently gained interest because of 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 (Amodio, Devine, & Harmon-Jones, 2007; Haidt, 2003). 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, Smith, Dufwenberg, & Sanfey, 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, Stillwell, A. M., & Heatherton, 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 (Basile et al., 2011; Bastin, Harrison, Davey, Moll, & Whittle, 2016; Kédia, Berthoz, Wessa, Hilton, & Martinot, 2008; Mclatchie, Giner-Sorolla, & Derbyshire, 2016; Morey et al., 2012; Shin et al., 2000; Takahashi et al., 2004; Zahn, de Oliveira-Souza, Bramati, Garrido, & Moll, 2009).

✉ Jose Sánchez-García
jossan07@ucm.es

¹ Center for Human Evolution and Behavior ISCIII–UCM, Madrid, Spain

² Psychobiology and Methods in Behavioral Sciences Department, Complutense University of Madrid, Madrid, Spain

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, Hu, Hu, and Zhou (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 three conditions. These were as follows: 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 Hooze, Nelissen, Breugelmanns, & Zeelenberg, 2011; Nelissen & Zeelenberg, 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 the 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, Wang, Cao, and Li (2017), 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 around 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. (2017), 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. (2017) 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. (2017), 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. (2017) 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 (Anolli & Pascucci, 2005; Wallbott & Scherer, 1995). 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, Esslen, Kochi, & Lehmann, 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. (2017) 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. (2017). 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. (2017) 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, López-Pérez, & Márquez, 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, Liu, & Lai, 2014). Therefore, we measured SCR throughout the experiment in each participant in order to estimate whether SW trials increased arousal as compared with the other conditions, and its possible relationship to ERP fluctuations linked to guilt. Furthermore, and along these same lines, we used the Spielberger State-Trait Anxiety Inventory (STAI; 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, Ortiz, Conejero, & Pascual, 2009; Silfver & Helkama, 2007), while they are functionally equivalent for both women and men (Monteith, 1993; Monteith, Ashburn-Nardo, Voils, & Czopp, 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.

Material and method

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 years, $SD = 1.54$ years). 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.

Procedure

Before the ERP experiment, the participating couple individually completed two different questionnaires: STAI (Spielberger, 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 (see Fig. 1) began with a fixation cross appearing at the center of the screen for 500 ms, followed by a black screen with 20 white dots (dot size = 3×3 pixels), randomly displayed, in a 300×300 pixel 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 1,500 ms. After the dots disappeared, a number (19, 20, or 21, randomly chosen) appeared on the screen for 3,000 ms. Meanwhile, participants had to press one of two buttons as soon as possible using the index or middle finger, respectively

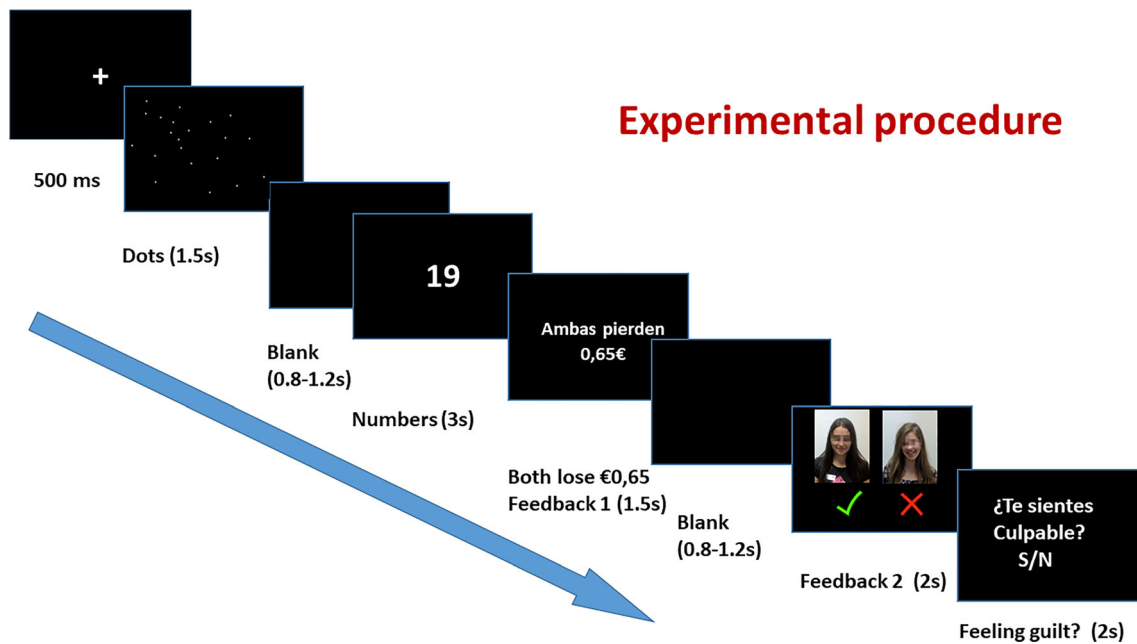


Fig. 1 Experimental procedure. A fixed cross appears for 500 ms at the center of the screen before the onset of random dots for 1,500 ms, where participants estimate the number. After a blank screen, 19, 20, or 21 appears for 3 s while participants have to choose whether the actual number was greater 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

(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.

After the presentation of a blank screen for 800, 1,000 or 1,200 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 1,500 ms and followed by a blank screen. Then the feedback of the performance of each player was presented for 2,000 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 2,000 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. (2017). 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 prestudy (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, Piazza, & Espinosa, 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., nonresponded 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 would 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 one quarter 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 did not know 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, 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 years, $SD = 2.21$ years), 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 the Emotion \times Condition interaction, $F(12, 204) = 37.6, p < .001$, 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 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.

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; off-line, they were rereferenced to the average of the right and left mastoids. The impedance of all electrodes was kept below 5 k Ω . 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 off-line from 0.1 to 30 Hz.

The continuous EEG was segmented into 1,200-ms epochs, starting 200 ms before the Feedback 2 screen onset. Eye movements were corrected using Independent Component Analysis (ICA; Makeig, Enghoff, Jung, & Sejnowski, 2000) as implemented in Brain Vision Analyzer®. Remaining artifacts were further removed by a semiautomatic rejection procedure, eliminating epochs exceeding $\pm 100 \mu\text{V}$ in any of the channels. The mean rejection rate of epochs was 4.8%, and there were no significant differences between conditions, $F(3, 69) = 1.98, p = 0.142, \eta_p^2 = 0.079$. Mean rejected epochs (and SDs) for conditions were as follows: BR ($M = 3.69, SD = 4.768$), PW ($M = 5.29, SD = 6.931$), SW ($M = 5.50, SD = 8.119$), and BW ($M = 4.75, SD = 6.732$).

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 1,000 ms in 100-ms-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; see 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

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

	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
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

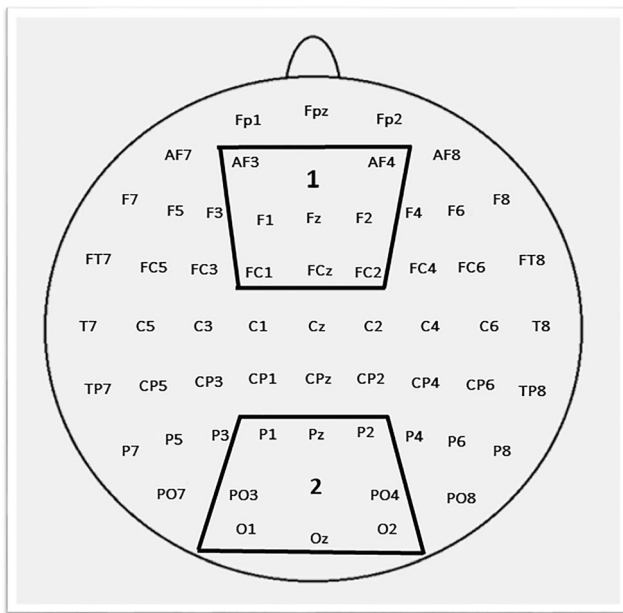


Fig. 2 Location of the electrodes covered by the two main regions of interest (ROIs); a frontal ROI (ROI1) and a parieto-occipital one (ROI2)

correction for nonsphericity was applied when necessary (Greenhouse & Geisser, 1959). Post hoc tests were corrected with the Bonferroni procedure.

SCR recording and analyses

The skin conductance response (SCR) was monitored throughout the experiment in each participant. An eight-channel Multibox polygraph (Brain Products, Munich) was connected to the index and middle fingers of the nonresponding hand in the experimental task. SCR results were analyzed with LEDALAB software (Benedek & Kaernbach, 2010b). Downsampling was applied to the original data to 125 Hz, and preprocessed 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, 2010a). Phasic data were considered with a minimum SCR amplitude of 0.01 μ S; the examining window covered from 1 to 5 s after the onset of Feedback 2. Individual measures of SCR in μ S were separately averaged for each participant. Six out of 24 participants were discarded due to artifacts 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.

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., 2017), as it is the result of subtracting two degrees of relatively comparable emotions (see Performance results) without the contamination of other confounding emotional responses.

Results

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 standard deviation for the individual ratings obtained were as follows: 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, $F(3, 69) = 41.174$, $p < .001$. Post hoc analyses showed significant differences between all conditions ($ps < .001$), with the exception of BR versus PW ($p = .054$).

SCR

Feedback 2 did not generate significant differences in the phasic activity between conditions, $F(3, 51) = 0.137$, $p = .86$; $\eta_p^2 = 0.008$.

ERPs and LORETA analyses

A visual inspection of ERP overall averages (see Fig 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 (see Figs. 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.

An Electrode (59) \times Condition (BR, PW, SW, BW) overall ANOVA showed significant effects of condition between 200

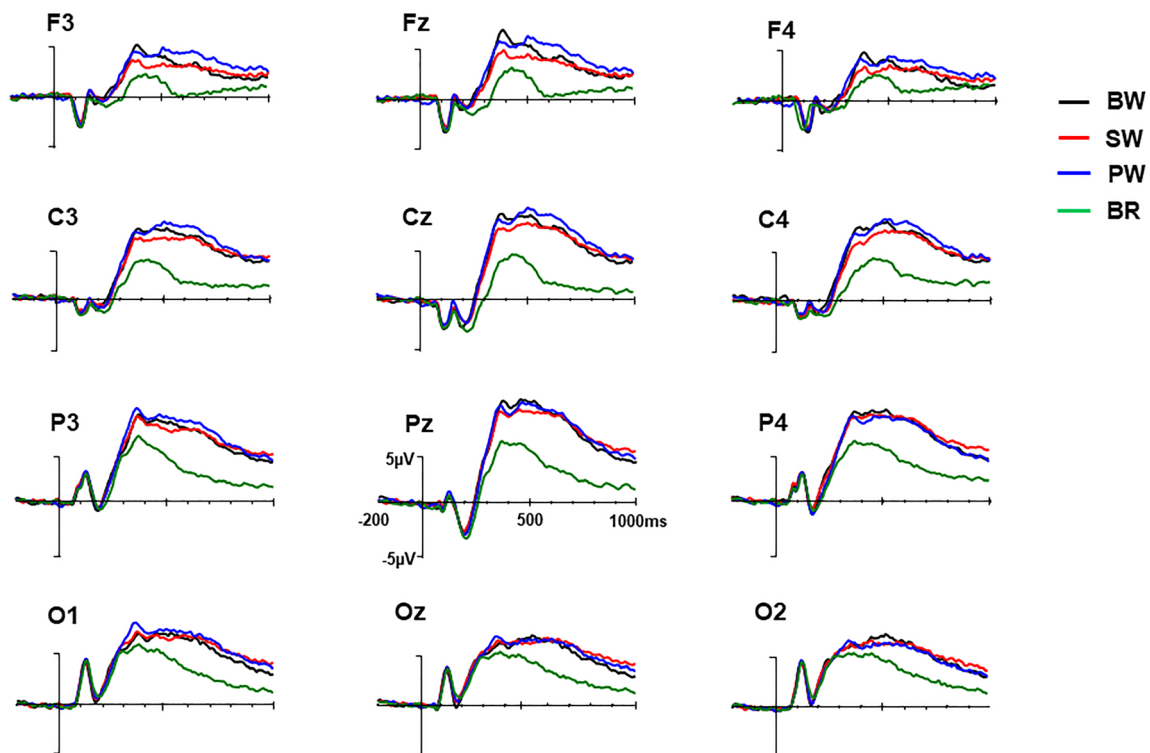


Fig. 3 ERP grand means for BW (both wrong), SW (self wrong), BR (both right), and PW (partner wrong) conditions at selected electrodes

and 1,000 ms, with all $F_s(3, 69) > 2.36$, and all $p_s < .05$, endorsing subsequent ANOVAs at selected ROIs and time windows. An ANOVA at ROI1 for the 350–450 ms window revealed significant effects of condition, $F(3, 69) = 17.373$, $p < .001$, $\eta_p^2 = 0.430$, and Electrode \times Condition interaction, $F(21, 483) = 6.808$, $p < .001$; $\eta_p^2 = 0.228$. Post hoc analyses exhibited significant differences between all conditions BR

versus PW, BR versus SW, BR versus BW, PW versus SW, and SW versus BW (all $p_s < .05$), with the exception of PW versus BW ($p = .10$).

An ANOVA at ROI2 within the 750–950 ms window showed significant effects of condition, $F(3, 69) = 16.86$, $p < .001$, $\eta_p^2 = 0.423$, and Electrode \times Condition interaction, $F(21, 483) = 7.04$, $p < .001$, $\eta_p^2 = 0.235$. Post hoc analyses in

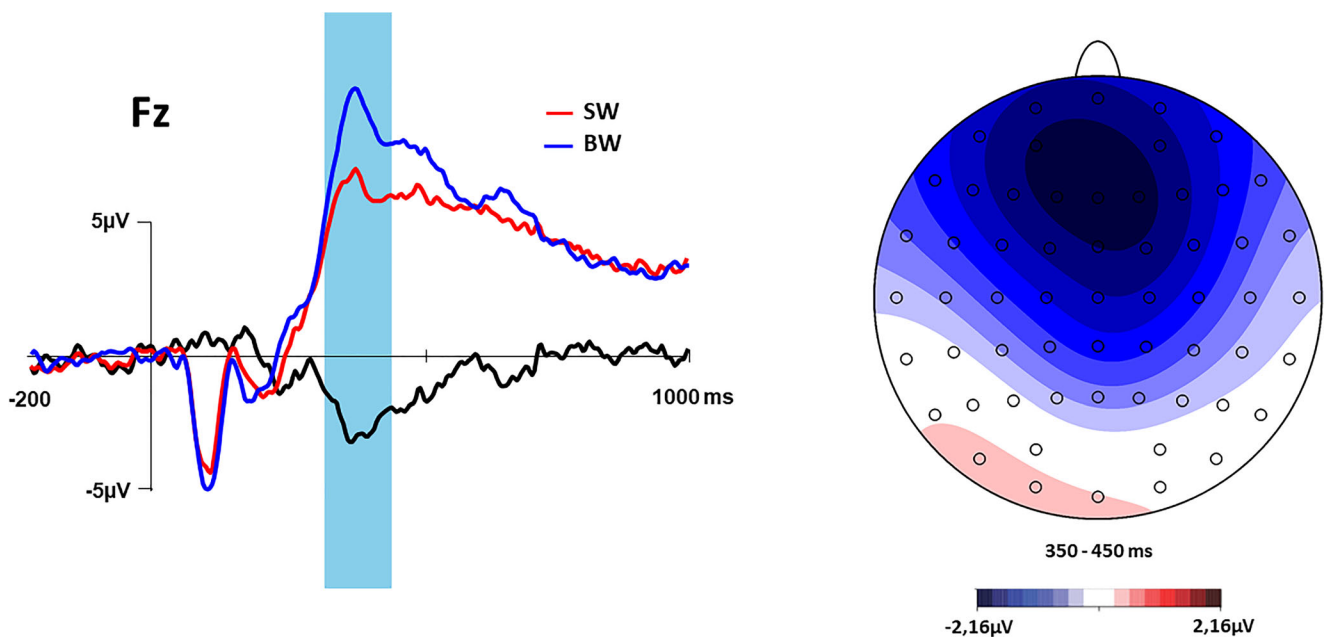


Fig. 4 Guilt effect as the difference wave between SW (self wrong) and BW (both wrong) in electrode Fz, and a map representing the 350–450 ms interval

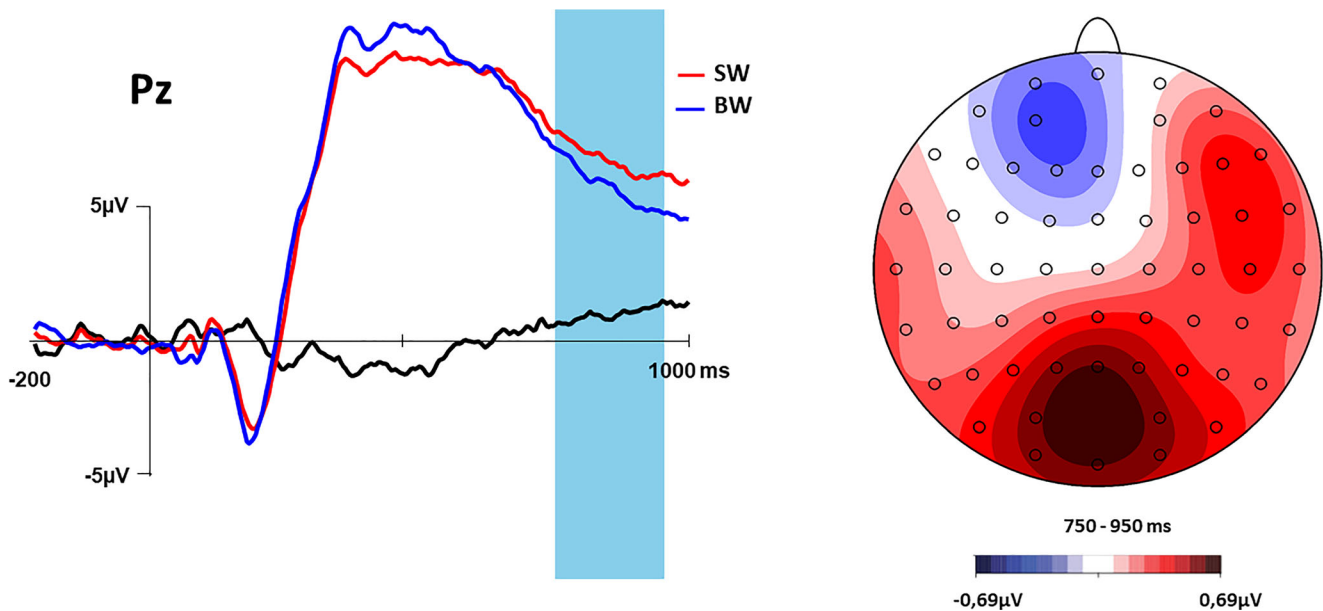


Fig. 5 Guilt effect as the difference wave between SW (self wrong) and BW (both wrong) in electrode Pz, and a map representing the 750–950 ms interval (LPC)

ROI2, however, revealed that these effects were the consequence of the differences between the BR condition and the other conditions (all $ps < .001$), while the latter did not differ significantly ($ps > .10$).

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/m^2) at MNI coordinates ($x = -10, y = 45, z = 50$), corresponding to the dorsal medial prefrontal cortex (mPFC; see 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, nonsignificant effects.

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 FN did not significantly correlate to each other. 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).

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. (2017) 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. (2017). The components found

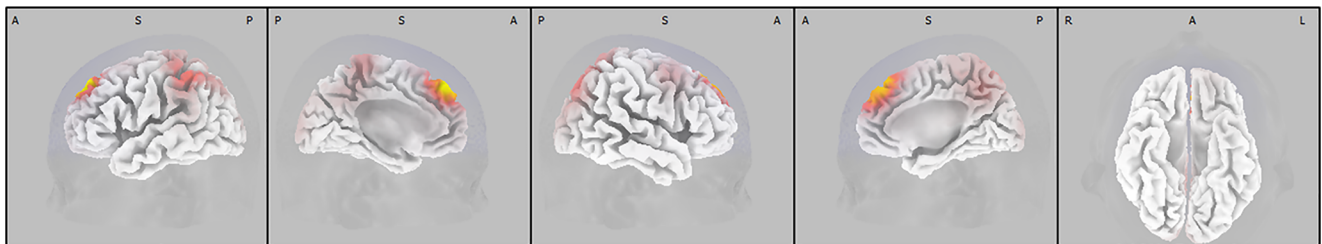


Fig. 6 Neural generators (LORETA analysis) of frontal negativity, performed for the difference SW minus BW (self wrong vs. both wrong) conditions, around its peak of about 400 ms (380–416 ms window) to increase the specificity of the solution

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 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., Anolli & Pascucci, 2005; Wallbott & Scherer, 1995).

Leng et al. (2017) 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 nonexistent in the other conditions. This suggests 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 Material and Method 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 (Baumeister et al., 1994; Eisenberg, 2000; Hoffman, 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, Nickel, & Azari, 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.

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.

Funding Research funded by grant PSI2017-82357-P from the Ministry of Science, Research and Universities, Government Program for High Level Scientific and Technical Research, Spain

References

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.

- 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.
- 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.
- 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.
- Baumeister, R., Stillwell, A. M., & Heatherton, T. F. (1994). Guilt: An interpersonal approach. *Psychological Bulletin, 115*(2), 243–267.
- Benedek, M., & Kaernbach, C. (2010b). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91.
- Benedek, M., & Kaernbach, C. (2010a). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*(4), 647–658.
- 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.
- 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), 462–473.
- Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51*(1), 665–697.
- 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.
- Fernández-Pinto, I., López-Pérez, B., & Márquez, M. (2008). Empatía: Medidas, teorías y aplicaciones en revisión [Empathy: Measures, theories and applications under review]. *Anales de psicología, 24*(2), 284–298.
- 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.
- 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.
- Haidt, J. (2003). The moral emotions. *Handbook of Affective Sciences, 11*(2003), 852–870.
- Hoffman, M. L. (2000). Empathy and moral development implications for caring and justice. Cambridge, UK: Cambridge University Press.
- 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.
- 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: Embarrassment, pride, shame, and guilt. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 573–691). New York, NY: Guilford Press.
- Liu, F., Liu, G., & Lai, X. (2014). Emotional intensity evaluation method based on Galvanic skin response signal. *2014 Seventh International*

- Symposium on Computational Intelligence and Design (ISCID)*, (Vol. 1, pp. 257–261). New York, NY: IEEE.
- 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*. Paper presented at the Second International Workshop on Independent Component Analysis and Signal Separation, Helsinki, Finland.
- 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.
- 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), 1029–1050.
- 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.
- Nelissen, R., & Zeelenberg, M. (2009). When guilt evokes self-punishment: Evidence for the existence of a Dobby effect. *Emotion*, 9(1), 118–122.
- 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 & Clinical Pharmacology*, 24(Suppl. C), 91–95.
- Seitz, R. J., Nickel, J., & Azari, N. P. (2006). Functional modularity of the medial prefrontal cortex: involvement in human empathy. *Neuropsychology*, 20(6), 743–751.
- Shin, L. M., Dougherty, D. D., Orr, S. P., Pitman, R. K., Lasko, M., Macklin, M. L., . . . 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), 1157–1167.
- Silfver, M., & Helkana, K. (2007). Empathy, guilt and gender: A comparison of two measures of guilt. *Scandinavian Journal of Psychology*, 48(3), 239–246.
- Skapinakis, P. (2014). Spielberger State-Trait Anxiety Inventory. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research*. Dordrecht, Netherlands: Springer.
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory STAI (form Y) ("self-evaluation questionnaire")*. Boca Raton, FL: CPP.
- 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), 967–974.
- Tangney, J. P., & Dearing, R. L. (2002). *Emotions and social behavior: Shame and guilt*. New York, NY: Guilford Press.
- 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: 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:<https://doi.org/10.1037/0022-3514.45.4.884>
- 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 cingulate activity reflects individual differences in empathic concern. *Neuroscience Letters*, 457(2), 107–110.
- 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*, 1–14. Advance online publication. doi:<https://doi.org/10.1080/17470919.2017.1391119>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.