

Meta-analysis of ERP investigations of pain empathy underlines methodological issues in ERP research

Michel-Pierre Coll

Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6HG, UK

Correspondence should be addressed to Michel-Pierre Coll, Department of Experimental Psychology, University of Oxford, Anna Watts Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6HG, UK. E-mail: michel-pierre.coll@psy.ox.ac.uk.

Abstract

Empathy has received considerable attention from the field of cognitive and social neuroscience. A significant portion of these studies used the event-related potential (ERP) technique to study the mechanisms of empathy for pain in others in different conditions and clinical populations. These show that specific ERP components measured during the observation of pain in others are modulated by several factors and altered in clinical populations. However, issues present in this literature such as analytical flexibility and lack of type 1 error control raise doubts regarding the validity and reliability of these conclusions. The current study compiled the results and methodological characteristics of 40 studies using ERP to study empathy of pain in others. The results of the meta-analysis suggest that the centro-parietal P3 and late positive potential component are sensitive to the observation of pain in others, while the early N1 and N2 components are not reliably associated with vicarious pain observation. The review of the methodological characteristics shows that the presence of selective reporting, analytical flexibility and lack of type 1 error control compromise the interpretation of these results. The implication of these results for the study of empathy and potential solutions to improve future investigations are discussed.

Key words: empathy; pain; ERP; methods; meta-analysis

Introduction

Empathy is a complex psychological construct that refers to the ability of individuals to share the experience of others (Batson, 2009; Cuff et al., 2016; Coll et al., 2017b). It is of high importance for healthy interactions with others and has been suggested to be altered in several psychiatric conditions (Decety and Moriguchi, 2007; Bird and Viding, 2014). With the hope that understanding the neuronal mechanisms of empathy will bring new insights on this concept, it has been one of the main endeavours of social neuroscience to describe the cerebral processes and computations underlying empathy (Decety and Jackson, 2004; de Vignemont and Singer, 2006; Klimecki and Singer, 2013).

While it is challenging to elicit empathy in a controlled neuroimaging experiment, studies often use cues of nocicep-

tive stimulation in others (i.e. a needle piercing a hand) to study empathy since they are relatively unambiguous, highly salient and easily understood (Vachon-Preseau et al., 2012). In the electroencephalography (EEG) literature, the event-related potential (ERP) technique has mostly been used to study this phenomenon by measuring electrical brain responses to nociceptive cues depicting various levels of pain in others. In a seminal study, Fan and Han (2008) showed participants real pictures or cartoon depictions of hands in painful or neutral situations and asked participants to either judge the intensity of the pain experienced or to count the number of hands present in the stimuli. The results showed an early effect of pain in the N1 and N2 components that was not influenced by task demands and a latter effect of pain in the P3 component that was modulated

Received: 21 May 2018; Revised: 9 August 2018; Accepted: 12 August 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

by task requirements. The authors interpreted these results as the presence of an early automatic response indexing emotional sharing and a late response indexing the cognitive evaluation of others' pain (Fan and Han, 2008).

This framework is now regularly used to study pain empathy in healthy and clinical samples. Broadly, the main results from these studies suggest that these ERP components reflect sensitivity to pain in others. Indeed, both early and late ERP components in response to vicarious pain have been related to self-reported trait empathy (Corbera et al., 2014; Fabi et al., 2016; Vaes et al., 2016) and to the intensity and unpleasantness of the pain perceived in others (Decety et al., 2010; Cheng et al., 2012; Meng et al., 2012). The differentiation between ERP to pain and neutral stimuli is decreased when task demands interfere with the processing of the pain stimuli (Fan and Han, 2008; Kam et al., 2014; Cui et al., 2017a) and increased when task demands favour a deeper processing (Li et al., 2010; Ikezawa et al., 2014). The nature of the target depicted in the stimuli has also been shown to influence early and late ERP responses, with decreased responses for cartoon depiction of pain (Fan and Han, 2008), pain inflicted to robots' hands (Suzuki et al., 2015) and to individuals of a different race (Sessa et al., 2014; Contreras-Huerta et al., 2014; Fabi et al., 2018). In psychopathology research, this paradigm was predominantly used with persons suffering from schizophrenia and indicates that compared to healthy controls, patients with schizophrenia show a decreased capacity to initiate and regulate their empathic response (Ikezawa et al., 2012; Gonzalez-Lienres et al., 2016b).

However, the reliability and relevance of these results are unclear. Indeed, studies using this paradigm are often plagued by several methodological issues commonly observed in ERP research. For example, in the initial study by Fan and Han (2008), the authors analysed multiple time windows at several scalp locations and performed over 100 statistical tests on ERP data without adjusting the significance threshold for those multiple comparisons. This suggests that some of the results have a high probability of being false positives (Kilner, 2013) and that the effect of vicarious pain observation on ERP therefore deserves further scrutiny.

The issue of multiple comparisons is a common problem in neuroimaging studies due to the large amount of data collected (Poldrack et al., 2008; Luck and Gaspelin, 2017). In the ERP literature, this is often made worse by the traditional use of factorial analyses performed in several time windows and at several scalp locations without clear hypotheses on the main effects and interactions (Luck and Gaspelin, 2017). Furthermore, this large amount of data also allows for considerable analytical flexibility; that is, the idea that the same data set can be analysed in different ways with significant changes in the results and interpretations depending on the analytical pipeline chosen (Carp, 2012). The presence of flexibility in design and analysis choices and ambiguity regarding how to best make these choices can lead researchers to compare the results of different analytical pipelines and choose the one that gives the most favourable pattern of result (Simmons et al., 2011; Carp, 2012). When considerable analytical variability is present in a particular field without justification, it can raise doubt regarding the validity of the results and their interpretations.

If the study of pain empathy using ERP is to provide results that are appropriate to further our understanding of empathy in different contexts and populations, it seems imperative to assess (i) the reliability of the effect of the observation of pain in others on the ERP response and (ii) the amount of variability and flexibility in the designs employed to investigate this phe-

nomenon. To reach these aims, a review of the methodological practices used in 40 ERP studies investigating pain empathy and a meta-analytical compilation of their results was performed. The results provide meta-analytical evidence for the association between late ERP components and the observation of pain in others. However, there was considerable variation in the design and analyses and incomplete reporting of results, raising doubts on the validity of these results.

Methods

Study selection

A systematic review of the literature was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). The articles included in this review were selected by searching PubMed for studies that were available online before 1 May 2018 using different combinations of keywords (e.g. 'EEG', 'ERP', 'pain', 'empathy', 'vicarious', see Supplementary Table S1). The reference lists and citation reports of eligible studies were also consulted.

To be included in this report, studies had to report scalp ERP data in response to pictures depicting nociceptive stimulations (e.g. Jackson et al., 2005). Studies using facial expressions stimuli were included only when nociceptive stimulations were visible in the stimuli (e.g. needle piercing the skin of the face). This procedure led to the selection of 40 studies published between 2008 and 2018 in 20 different journals (Supplementary Table S1, Table 2 and asterisks in the Reference list).

From the 40 studies reviewed, four were excluded from quantitative meta-analysis. One was excluded because it used the same data set and analyses as another study (Han et al., 2008), one because it did not report sufficient information (Ikezawa et al., 2014), one because it used non-parametric statistics (Fitzgibbon et al., 2012) and one because it reported incorrect degrees of freedom and F statistics (Sun et al., 2017). The quantitative meta-analysis was therefore performed on 36 studies (marked with double asterisks in the References). The PRISMA flowchart for study selection and rejection is shown in Figure 1. The data reported in this review were manually extracted from the text of the published articles or accompanying Supplementary Materials and available in Supplementary Table S1.

Methodological review

Several variables concerning the participants, materials and procedures, data collection and preprocessing, ERP measurements, statistical analyses and the reporting of results were collected and summarised below. When information was not clearly reported, the value was estimated based on the available information or assumed to have a particular value (e.g. when no post-hoc correction was reported, it was assumed that none were used). When insufficient information was available for a particular variable, it was marked as not reported and treated as a missing value. In order to assess the exhaustiveness of the hypotheses formulated regarding ERPs in each study, hypotheses were classified in one of four categories: *Complete*, *Partial*, *Alternative* and *None*. Hypotheses were rated as *Complete* if they clearly predicted the specific components that were expected to be influenced by all manipulations as well as the direction and the location of this effect. If some predictions were present but were incomplete or unclear, the hypothesis was rated as *Partial*. Hypotheses that were formulated as two alternative outcomes without a clear prediction were labelled *Alternative*

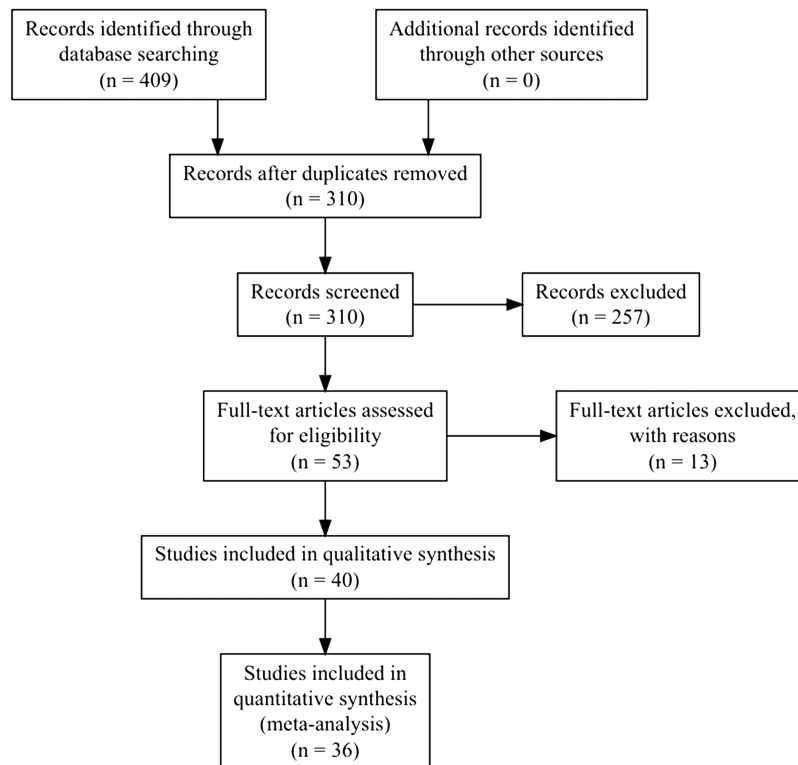


Fig. 1. PRISMA flowchart. Reasons for exclusions are shown in Table S1.

and the absence of prediction regarding ERP effects was labelled *None*. This procedure was applied separately for the factorial analysis of variance (ANOVA) performed on ERP components and for the correlational analysis of ERP components with other variables.

Meta-analysis of ERP components

A quantitative meta-analysis was carried out to assess the evidence for a modulation of different ERP components by the observation of pain. However, this was complicated by the fact that most studies reviewed reported only significant results (section *Results reporting*) and the general lack of clarity and precision of the results section of many studies. Nevertheless, when possible, F-values for the omnibus repeated measures test comparing the ERP response to pain and neutral stimuli were collected for each study.

When several between-subject and within-subject factors were manipulated, the F-value of the baseline condition was selected when available (e.g. in healthy controls or following neutral priming). Similarly, when available, the F-values for individual electrodes were collected. However, in most cases, the omnibus F-value from the main effect of pain in a multi-factorial analysis was collected and attributed to all electrodes included in the analysis. When mean amplitudes and standard deviations or standard errors were reported, the paired sample t-value for the pain effect was calculated assuming a correlation of 0.70 between measurements. When only the exact P-value was reported, the corresponding t-value was found using the t distribution. Following available guidelines (Cooper and Hedges, 1994; Moran et al., 2017), when the effect was reported as non-significant without the information necessary to compute an

effect size, the effect size was calculated assuming $P = 0.5$. This was done for each of the most frequently analysed components [N1, N2, P3 and the late positive potential (LPP)]. All F-values were subsequently converted to t-values by taking their square root (Brozek and Alexander, 1950). In order to compare the effect sizes across studies, the t-values collected for each electrode and component were converted to Hedges's g , a standardised measure of difference that is less biased than Cohen's d , especially for small samples (Hedges, 1981). Effects were scored as positive when the observation of pain led to increased ERP amplitude (i.e. more positive) than the observation of neutral stimuli and as negative when the ERP amplitude was more positive in response to neutral stimuli compared to pain stimuli.

Effect sizes were summarised in different ways. First, the spatial distribution of the effects was assessed by plotting scalp maps of the weighted absolute effect size for each component of interest. The absolute effect was taken to show where the effects were stronger on the scalp independently of their direction. The average effect at each electrode was also weighted by the number of studies including this electrode in their analysis in order to down-weight the effects at electrodes that were only analysed in a small number of studies. Second, the proportion of significant effects and significant interactions with other factors was compiled for each component. Third, a random effect meta-analysis was performed for each component at fronto-central (Fz, FCz, F1, F2, F3, F4, FC1, FC2, FC3, FC4), centro-parietal (Cz, CPz, C1, C2, C3, C4, CP1, CP2, CP3, CP4) and parieto-occipital (Pz, POz, P1, P2, P3, P4, PO1, PO2, PO3, PO4) electrode clusters to estimate summary effect size and the heterogeneity across studies. Finally, potential publication bias was assessed using funnel plots and regression tests (Egger et al., 1997). To assess potential excess significance, the number of studies finding a significant effect for each component was compared to the

Table 1. Ratings of the exhaustiveness of hypotheses for the factorial and correlational analyses of ERP data

Judgement of hypothesis	Factorial analyses on ERP data (% of 40 studies)	Correlations with ERP data (% of 26 studies)
None	10	69.23
Partial	50	26.92
Complete	22.5	0
Alternative	17.5	3.85

expected number of significant studies given the power of each study to detect the summary effect size using exact one-tailed binomial tests and a significance threshold of 0.10 (Ioannidis and Trikalinos, 2007).

Data availability

All data and scripts used to produce this manuscript and accompanying figures, the PRISMA guidelines checklist and supplementary information and figures are available online at <https://osf.io/xtm5c/>. All data processing and analyses were performed using Rstudio (RStudio Team, 2015; R Core Team, 2018) and the Fieldtrip toolbox (Oostenveld et al., 2011) within Matlab R2017a (The MathWorks, Inc., Natick, Massachusetts, United States).

Systematic review of methodological practices

Goals and hypotheses

All studies reviewed aimed at comparing the effect of an experimental manipulation and/or participant characteristics on the ERP to pain stimuli with the goal of furthering the understanding of the mechanisms underlying empathy in the general population and/or in various clinical groups. As shown in Table 1, the majority of studies presented incomplete hypotheses regarding the analysis of ERP components. While some studies provided complete hypothesis for the factorial analysis of ERPs, this was rarely the case for correlational analyses of ERP and other behavioural or physiological variables, suggesting that most of these analyses were exploratory in nature.

Participants

In the 40 studies reviewed, 42.5% used a between-subject design and 57.5% used a within-subject design. Among studies employing a between-subject design, 22.5% compared participants from the general population to participants from a clinical group. These clinical conditions included autism, amputation, bipolar disorder, fibromyalgia, juvenile delinquents and schizophrenia. Only two studies provided a justification for their sample size based on a priori power analyses. The average sample size, sample size per group and participants excluded are shown in Table 2. In order to assess the power of each study to detect a small ($d = 0.2$), medium ($d = 0.5$) or large ($d = 0.8$) effect size (Cohen, 1992), a power analysis was performed for each study and each of these effect size using the sample size per group, a two-sided paired t-test and a significance threshold of 0.05. As shown in Figure 2, most studies were only adequately powered to detect a large effect size equal to or higher than $d = 0.8$. No studies had 80% power to detect a small or medium effect size.

Materials and procedure

Visual stimuli. The majority of studies reviewed (82.5%) used static pictures depicting limbs (hands or feet) in painful and

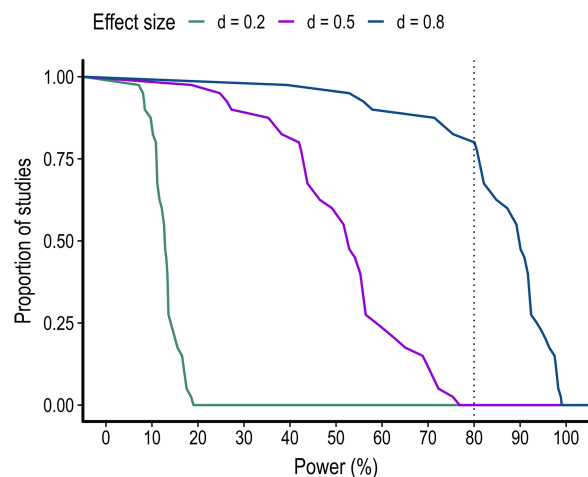


Fig. 2. Proportion of studies as a function of the level of power to detect a small ($d = 0.2$), medium ($d = 0.5$) and large ($d = 0.8$) effect size.

non-painful situations similar to those used initially by Jackson et al. (2005). The other types of stimuli included short three-frame clips of limbs in painful situations (7.5%), static pictures of faces pricked by a needle or touched with a cotton bud (5%), both faces and limbs (2.5%) and anthropomorphised objects pricked by a needle or touched with a cotton bud (2.5%). The average stimulus duration is shown in Table 2.

Experimental task. All studies compared ERPs to painful and non-painful stimuli. Including this pain factor, studies had an average of 2.08 within-subject factors (s.d. = 0.45; range, 1–3) and an average of 4.55 within-subject conditions (s.d. = 1.67; range, 2–8). During the experimental tasks, the participants were either asked to detect the presence of pain in a forced choice format (60% of studies), to assess the intensity of the pain observed using a rating scale (17.5%), to passively observe the pictures (15%) or to perform another behavioural task (7.5%). The average number of trials per condition is shown in Table 2.

Recordings. EEG was collected from 60 to 64 scalp electrodes in the majority of cases (70%) while the remaining studies used 32 (27.58%), 72 (5%) or 128 (5%) scalp electrodes. A total of 30% of the studies did not report the manufacturer of the EEG system used. The majority of studies (27.5%) used an EEG system manufactured by Brain Products. Other manufacturers included Biosemi (BioSemi BV, Amsterdam, Netherlands) (27.5%), EGI Geodesic (Electrical Geodesics Inc., Eugene, OR, USA) (5%) and NuAmps (Compumedics Neuroscan Inc., El Paso, TX) (10%).

Preprocessing. EEG data were high-pass filtered in most cases with a cutoff value of 0.1 Hz (52.5%) and the rest of the studies used a high-pass cutoff between 0.01 and 1 Hz. For low-pass filters, the most used cutoffs were 100 and 30 Hz used in 30% and 40% of the cases, respectively. Other low-pass filter cutoffs

Table 2. Main design, task and EEG characteristics for each study reviewed

Study	Between subjects	Sample size (total)	Number of exclusions	Sample size (per group)	Clinical sample	Task type	Stimulus type	Stimulus duration (ms)	Number of conditions	Number of trials per condition	Number of electrodes	EEG reference	Components analysed
Chen et al., 2012	No	28	2	28		Passive observation	Dynamic limbs	1000	4	64	64	Average reference	N260, N360, LPP
Cheng et al., 2012	Yes	43	2	14.33	Juvenile delinquents	Pain detection	Static limbs	1000	4	40		Single mastoid	N120, P3, N360, LPP
Cheng et al., 2014	Yes	72	22	8.5		Passive observation	Dynamic limbs	2050	2	80	32	Average mastoids	EAC, LPP
Coll et al., 2017b	Yes	48	2	16		Pain assessment	Static limbs	2500	4	40	128	Average reference	N200, P2, LPP
Contreras-Huerta et al., 2014	No	21	0	21		Pain detection	Static faces	1000	8	96	64	Average reference	N1, P3, N170, VPP, P2
Corbera et al., 2014	Yes	37	0	18.5	Schizophrenia	Pain detection	Static limbs	500	4	204	32	Average mastoids	N110, P180, N240, LPP
Cui et al., 2016b	No	20	3	20		Other	Static limbs	1000	4	120	63	Average reference	N1, P2, N2, P3, LPP
Cui et al., 2016a	No	28	0	28		Passive observation	Static limbs	1000	6	60	63	Average reference	N1, P2, N2, P3, LPC
Cui et al., 2016c	No	23	3	20		Other	Static limbs	1000	8	110	63	Average reference	N1, N2, P3
Cui et al., 2017b	No	20	3	17		Other	Static limbs	1000	4	60	63	Average reference	N1, N2, LPP
Cui et al., 2017a	No	22	1	22		Passive observation	Static limbs	1000	4	240	63	Average reference	N1, P2, N2, P3
Decety et al., 2010	Yes	30	3	15		Passive observation	Static limbs	1000	2	150	32	Single mastoid	N110, P3
Decety et al., 2015	Yes	38	1	19.5		Pain assessment	Static limbs	1500	4	50	64	Average mastoids	Early, LPP
Decety et al., 2017	No	30	6	30		Passive observation	Static limbs	2000	4	20	32	Average reference	N2, LPP, Slow wave
Fabi et al., 2016	No	16	3	16		Pain detection	Static limbs	200	4	62	72	Average reference	P1, N1, N240, P3
Fabi et al., 2018	No	24	1	23		Pain detection	Static limbs	400	4	60	72	Average reference	P1, N1, EPN, P3b
Fallon et al., 2015	Yes	37	0	18.5	Fibromyalgia	Pain assessment	Static limbs	3000	2	50	64	Average reference	P1, N1, P2, N2, P3, LPP
Fan & Han, 2008	No	26	5	26		Pain detection	Static limbs	200	8	320	64	Average mastoids	N110, P180, N240, N340, P3, P1, N170, P320
Fan et al., 2014	Yes	40	18	20	Autism Spectrum Disorder	Pain detection	Dynamic limbs	2080	4	60	32	Average reference	N2, LPP
Fitzgibbon et al., 2012	Yes	28	0	9.33	Amputees /Synaesthetes	Pain assessment	Static limbs	3000	4	80	64	Average reference	N110, P180, N240, N340, P1, N170, P320, N3

(Continued)

Table 2. Continued.

Study	Between subjects	Sample size (total)	Sample size (per group)	Clinical sample	Task type	Stimulus type	Stimulus duration (ms)	Number of conditions	Number of trials per condition	Number of electrodes	EEG reference	Components analysed
Gonzalez-Liencres et al., 2016a	Yes	39	19.5	Schizophrenia	Pain detection	Static limbs	500	4	24	32	Average mastoids	N110, P400, P1, P3
Gonzalez-Liencres et al., 2016b	Yes	60	15		Pain assessment	Static limbs	750	2	70	32	Average mastoids	N110, P3
Han et al., 2008	Yes	26	13		Pain detection	Static limbs	200	4	320	64	Average mastoids	N110, P180, N240, N340, P3, P1, N170, P320
Ibáñez et al., 2011	No	13	6.5		Pain detection	Static limbs	500	4	40	128	Average mastoids	N1, P3
Ikezawa et al., 2012	Yes	52	13	Schizophrenia	Pain detection	Static limbs		4	408	32	Average mastoids	N110, P180, P3a, P3b, P1, N170
Ikezawa et al., 2014	No	18	18		Pain detection	Static limbs	350	8	102	32	Average mastoids	P3, LPP
Jiang et al., 2014	Yes	36	18		Pain detection	Static limbs	200	6	120	64	Average mastoids	N2, N320, P3
Kam et al., 2014	No	19	19		Pain detection	Static limbs	300	8	400	64	Not reported	Early, P3, LPP
Li et al., 2010	No	24	24		Pain detection	Static limbs	200	4	160	64	Average mastoids	N110, P160, N240, N320, P3
Iyu et al., 2014	No	28	28		Pain assessment	Static limbs	3000	4	74	64	Single mastoid	N1, P3
Mella et al., 2012	Yes	29	14.5		Pain detection	Static limbs	200	4	160	64	Average reference	N110, N340, LPP
Meng et al., 2012	No	20	20		Pain detection	Static limbs	3000	4	135	64	Single mastoid	N1, P2, N2, P3, LPC
Meng et al., 2013	No	20	20		Pain assessment	Static limbs	1000	4	60	64	Single mastoid	N1, P2, N2, P3, LPC
Sessa et al., 2013	No	12	12		Pain detection	Static faces	250	4	240	64	Average earlobes	N1, P2, N2, N2-N3, P3
Sun et al., 2017	No	36	31		Pain detection	Static limbs and static faces	250	6	150	64	Average mastoids	N1, N170, N1, P1, VPP, N2, LPC
Suzuki et al., 2015	No	15	15		Pain detection	Static limbs	500	4	112	32	Average earlobes	N1, P2, N2, P310, P3
Vaes et al., 2016	Yes	18	9		Pain detection	Other	250	4	90	64	Average earlobes	N1, P2, N2-N3, P3
Wang et al., 2016	No	15	15		Pain detection	Static limbs	1500	4	64.25	66	Average mastoids	N110, P300, LPP, N2
Wexler et al., 2014	No	37	18.5	Schizophrenia	Pain detection	Static limbs	500	8	102	32	Average reference	P3, LPP
Yang et al., 2017	Yes	53	26.5	Bipolar disorder	Pain detection	Static limbs	1000	4	60	64	Average mastoids	N1, P3

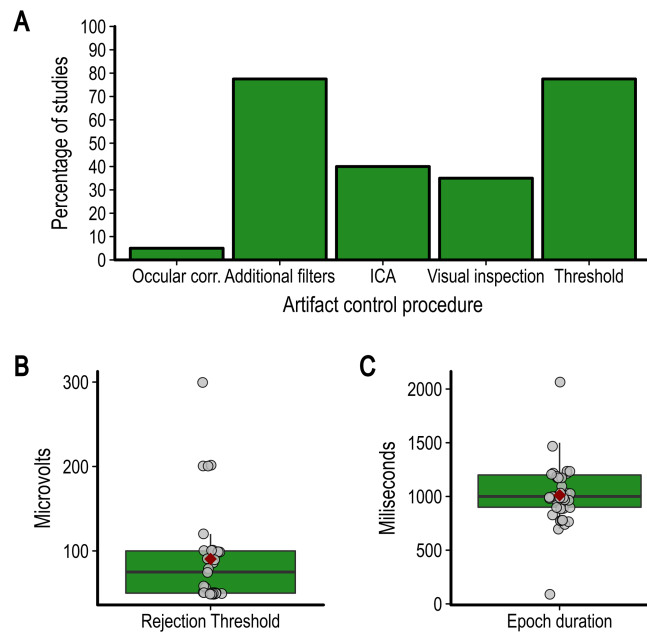


Fig. 3. (A) Percentage of studies reviewed using each type of artifact control procedure identified, (B) value of the rejection threshold if used and (C) duration of epochs.

included values between 40 and 80 Hz. Studies show that 7.5% and 5% did not report using a high-pass filter or a low-pass filter, respectively, while 5% reported using a notch filter to filter out electrical noise.

The average of the mastoid processes and the average of all scalp electrodes were the most popular reference schemes for EEG analyses (37.5% and 40%, respectively). Other studies use the average of the earlobes (7.5%), a single mastoid (12.5%) or did not report the reference used for analysis (2.5%). EEG data were epoched for analyses and in all cases the average pre-stimulus baseline was subtracted from the post-stimulus epoch. The duration of this baseline was on average 193.75 ms (s.d. = 28.17; range, 100–250). The average post-stimulus epoch duration is shown in Figure 3C.

All studies reported using at least one method to remove or correct for artifacts. Artifact rejection procedures included rejecting epochs by visual inspection or using a fixed amplitude threshold. Artifact correction procedures included removing components after independent component analysis or using various algorithms to remove electrooculogram (EOG) activity from the data. Some studies reported using additional filters to remove artifacts without providing further details. The percentage of studies using each of the main procedures is shown in Figure 3A. Automatic rejection using a fixed threshold was the most used method and the average rejection threshold is shown in Figure 3B. When using an artifact rejection procedure, 50% of studies reported the average number of epochs removed. On average, 11.34% of trials were removed (s.d. = 6.04; range, 1.34–29%).

ERP analyses

ERP selection and measurement. The average number of components analysed is shown in Figure 4C. In most cases, the choice of ERP components to analyse was based on previous studies (72.5%), while other studies chose components based on the inspection of the grand average waveform (17.5%), used another analysis to select the components of interest (2.5%) or

did not justify their selection of components (7.5%). As shown in Figure 4A, the most widely analysed ERP components were the N1, N2, P2, P3 and LPP. Note that in some cases, slightly different names were used for these components (e.g. P320 instead of P3). Furthermore, some studies also performed several analyses on the same component (e.g. early and late LPP). See Supplementary Table S1 for the names and all components analysed in each study.

The average number of locations analysed is shown in Figure 4C, and the percentage of studies analysing each scalp location is shown on a 64-electrode montage in Figure 4B. A minority of studies provided a justification for the choice of locations to analyse (40%) and in most cases this choice was based on previous studies (25%). Almost all studies quantified the ERP components using the mean amplitude within a time window (80%), while other studies used the peak amplitude within a time window (7.5%) or point-by-point analyses (i.e. performing analyses in small time windows covering the whole ERP epoch, 7.5%). One study used peak amplitude or mean amplitude depending on the component (5%). The choice of the time window to analyse was either based on visual inspection (47.5%), previous studies (22.5%) or both inspection and previous studies [previous studies (5%) the data itself (i.e. circular analyses, 15%) or not justified (10%)]. The percentage of studies analysing each time point in the poststimulus window for each component is shown in Figure 4D.

ERP statistical analyses. Almost all studies used factorial ANOVAs to assess the statistical significance of the experimental factors on the ERPs (95%). One study used the analysis of covariance and another used the non-parametric Kruskal–Wallis test. Since the few studies using point-by-point analyses (7.5%) performed a large number of ANOVAs compared to the rest of the studies (on average 130 ANOVAs), these studies were not considered in the following description of the factorial analyses. Another study performing 54 ANOVAs was also considered an outlier and was not included in the descriptive statistics. The average number of ANOVAs performed in the remaining studies

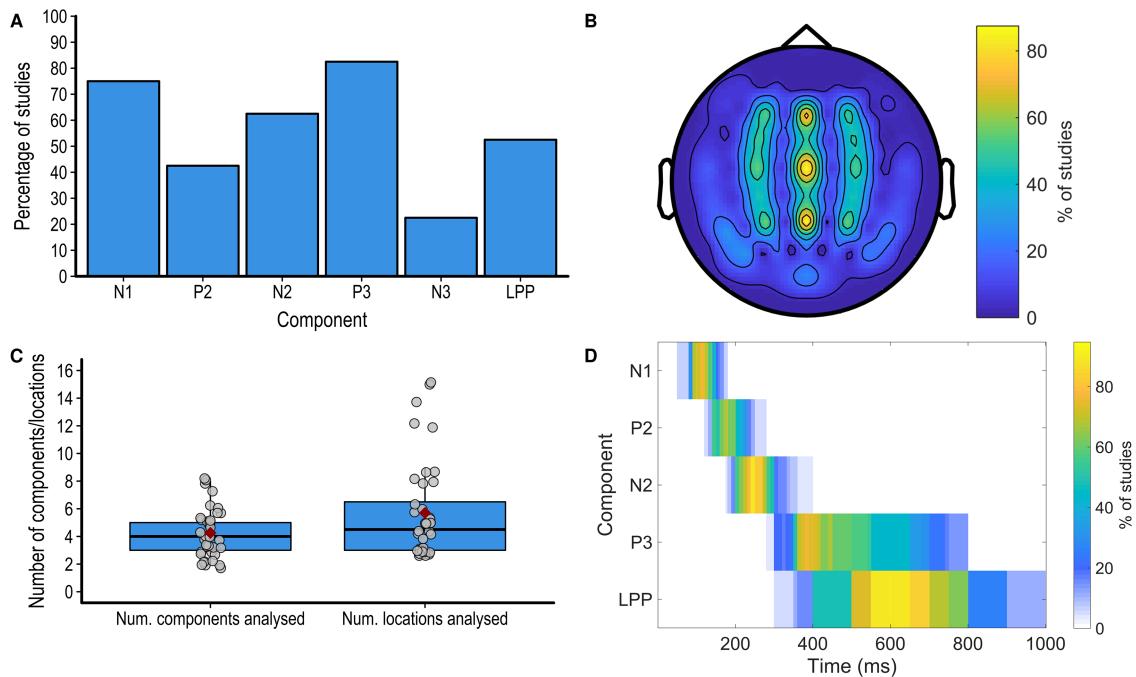


Fig. 4. (A) Number of components and locations analysed in each study. (B) Percentage of studies analysing each of the main component. (C) Percentage of studies analysing each scalp location. (D) Percentage of studies analysing each component analysing each time-point in the post-stimulus window.

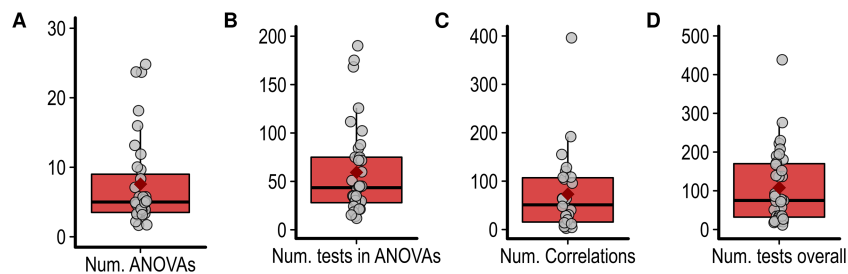


Fig. 5. (A) Number of ANOVAs performed. (B) Total number of statistical tests performed using ANOVAs. (C) Number of correlations. (D) Overall number of statistical tests performed.

($n = 35$) is shown in Figure 5. These ANOVAs had an average of 3.2 factors (s.d. = 0.72; range, 2–5) and 19.6 cells (s.d. = 15.97; range, 4–60). To assess how many statistical tests these ANOVAs represent, the number of main effects and potential interactions was multiplied by the number of ANOVAs (Luck and Gaspelin, 2017). The total number of tests in ANOVAs is shown in Figure 5. No studies corrected the significance threshold for the total number of ANOVAs performed. However, some studies (60%) corrected the significance threshold when performing post-hoc comparisons using the Bonferroni (42.5%), Tukey (10%), Scheffe correction (5%) or false discovery rate (FDR) (2.5%) correction. Several studies used the Greenhouse–Geisser correction when performing repeated-measure analyses (52.5%).

In addition to the factorial analyses, correlations between ERP data and other behavioural, physiological or questionnaire measures, were performed in 65% of studies. The variables most frequently correlated with the ERP data were the four subscales of the Interpersonal Reactivity Index (Davis, 1983, 27.5% of studies) as well as subjective ratings of the intensity of the pain experienced by the models in the stimuli and of the unpleasantness experienced by the participant when watching these stimuli (27.5% of studies). These correlations either used the difference between the ERP amplitude in two conditions (32.5%),

the mean amplitude in a particular condition (27.5%), both mean amplitude and peak amplitude (2.5%) or mean amplitude and peak latency (2.5%) and were often performed in multiple time windows and scalp locations. The average number of correlations per study is shown in Figure 5. In 10% of the studies, the significance threshold was corrected to control for the possibility of a type 1 error in correlation analyses.

Results reporting

A total of 92.5% of the studies reported mainly or exclusively significant results. When reporting the results from factorial analyses, 50% of the studies did not report any estimate of effect size, while the rest reported the partial eta squared (45%), Cohen's d (2.5%) or both (2.5%).

All but one study plotted the ERP data, 32.5% plotted only the time course of the ERP response, while 65% plotted both the time course and scalp maps at particular time points. In most cases, the locations and time points plotted were chosen because they were thought to be representative of the results (57.5%), while other studies plotted all locations analysed (25%) or only the locations showing significant effects (15%). On the time-course

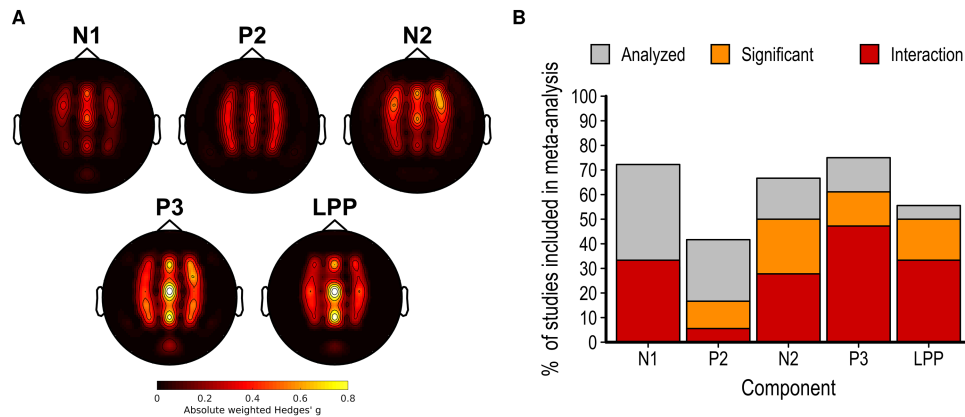


Fig. 6. (A) Scalp map of the average partial eta squared weighted by the number of studies reporting results at this electrode for the main effect of pain observation on each electrode and component. (B) Bar graph showing the proportion of studies included in the meta-analysis analysing each component, reporting a significant main effect of pain and a significant interaction between this effect and another factor.

plots, only one study plotted the error intervals. In addition to the time course and scalp maps, 22.5% of studies also reported the ERP amplitudes in a table and 42.5% in bar graphs.

Meta-analysis of the effect of pain observation

The results from the meta-analysis of the effect of the observation of pain stimulation on the ERP components that were the most frequently analysed are shown in Figures 6 and 7. The complete set of forest plots for each component at each location is shown in Supplementary Figures S1–S8.

N1 component. The effect of pain observation on the N1 component is maximal at frontal electrodes (Figure 6A). Although a minority of studies measuring this component found a significant effect of pain observation, several studies found a significant interaction between the effect of pain observation and another experimental factor (Figure 6B). The random-effect meta-analytic model fitted on the effect of pain on the N1 component at fronto-central sites collected from 22 studies suggests the presence of a high heterogeneity in the effect sizes ($Q = 94.67$; $df = 21$; $P < 0.01$; $I^2 = 79.37\%$). Overall, the random-effect model indicated that of pain observation on the N1 component at fronto-central sites is not significant [$g = 0.07$; $k = 22$; $P < 0.01$; 95% confidence interval (CI), -0.21 to 0.35].

N2 component. Similarly to the N1 component, the effect of pain observation on the N2 component is also maximal at frontal electrodes (Figure 6A). Approximately 50% of the studies included in the meta-analysis found a significant effect and 27.78% found a significant interaction between the effect of pain and another experimental factor (Figure 6B). There is an evidence for significant heterogeneity across the 20 studies ($Q = 108.46$; $df = 19$; $P < 0.01$; $I^2 = 85.13\%$). Interestingly, the direction of the significant effects is highly heterogeneous, with a similar number of studies finding a significant increase or decrease in amplitude during pain observation. This led to a non-significant overall effect of pain observation on the N2 component at fronto-central sites ($g = 0.18$; $k = 20$; $P < 0.01$; $I^2 = 85.13$; 95% CI, -0.15 to 0.5).

P3 component. While the effects of pain observation on the P3 component are distributed across the scalp, the effect is maximal at centro-parietal sites. Most studies measuring this com-

ponent found a significant effect and a significant interaction between the effect of pain and another experimental factor. Although the effect sizes for the P3 component across the 20 studies are considerably less heterogeneous than those found for the early components, there is still significant heterogeneity across studies ($Q = 55.78$; $df = 19$; $P < 0.01$; $I^2 = 66.33\%$). All studies found that pain observation led to a positive shift in P3 amplitude and the overall effect is large and significant ($g = 0.97$; $k = 20$; $P < 0.01$; 95% CI, 0.73 – 1.2).

LPP component. The effect of pain observation on the LPP component is strongest at centro-parietal sites. As for the P3 component, most studies measuring this component found a significant effect of pain observation and 33.33% found a significant interaction between the effect of pain and another experimental factor. There is a significant heterogeneity across the 17 studies ($Q = 63.62$; $df = 16$; $P < 0.01$; $I^2 = 76.39\%$), despite the fact that pain observation led to a positive shift in LPP amplitude for all studies. The overall effect of pain observation on the LPP is large and significant ($g = 1.1$; $k = 17$; $P < 0.01$; 95% CI, 0.8195 – 1.39).

Publication bias and excess significance. Funnel plots illustrating the effect size for each study as a function of study precision are shown in Figure 8. It should be noted that the low variance in precision (due to most studies having a similar sample size) limits the interpretation of these figures. Nevertheless, the funnel plots show that the effect sizes were roughly symmetrically distributed across the summary effect size, suggesting the absence of publication bias. This observation was formally tested using linear regressions to assess the relationship between the magnitude of the effect size and the precision (standard error) of each study. This procedure revealed non-significant relationships between effect size and precision for the N1 component at fronto-central sites [$t(20) = 1.84$; $P = 0.08$], the N2 component at fronto-central sites [$t(18) = -0.76$; $P = 0.45$] and the P3 component at centro-parietal sites [$t(18) = 1.46$; $P = 0.16$]. A significant negative relationship between effect sizes and precision is present for the LPP component at centro-parietal sites ($b = 3.91$; $t(15) = 2.32$; $P = 0.03$). This suggests that the large effect sizes found for the LPP component in some studies with a smaller sample size are likely inflated.

Exact binomial tests assessing the presence of excess significance suggest that the number of significant effects found for the N1 and N2 components is significantly higher than expected

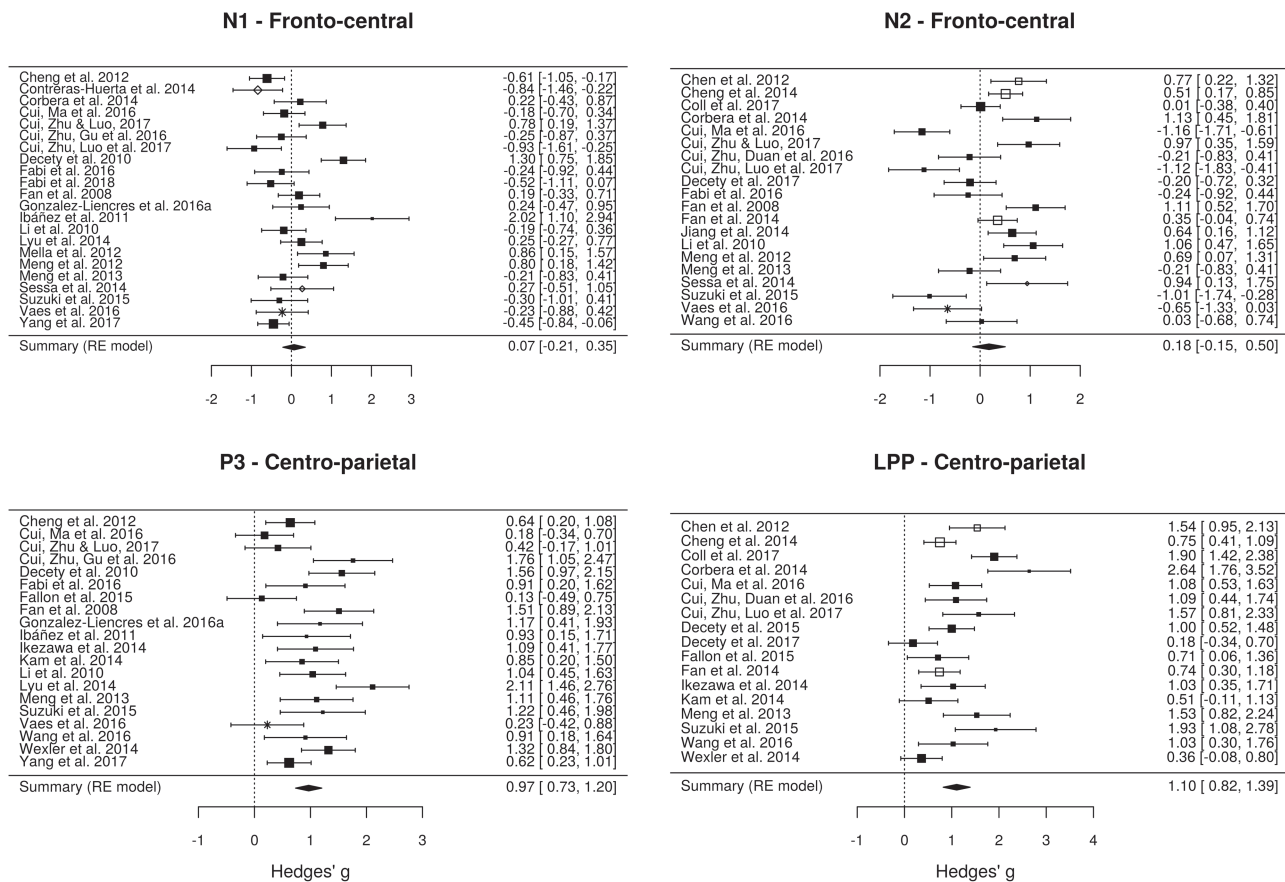


Fig. 7. Forest plots showing the effect size and 90% CI for the effect of pain observation on the ERP amplitude at the frontal electrodes from the N1, N2, P3 and LPP components. Positive effects indicate higher (more positive) amplitudes for pain pictures and negative effects indicate lower (more negative) amplitudes for the pain pictures. Symbols show the type of stimulus used in each study (black square, static limbs; white square, dynamic limbs; white diamond, static faces; asterisk, other). Symbol size shows the relative sample size for this component/location.

given the power of each study and the summary effect size ($P < 0.1$ for both). The proportion of significant effects is, however, similar to the expected number for the P3 and LPP components ($P > 0.9$ for both).

Discussion

The meta-analysis of 36 studies investigating pain empathy using ERPs suggests that the observation of pain in others does not reliably modulate the early N1 and N2 components often measured in these studies. However, large and reliable effects were found for the latter P3 and LPP components.

These findings challenge a popular model arguing that the time course of the ERP to the observation of pain in others is characterised by an early frontal 'affective sharing response' followed by a centro-parietal 'late cognitive reappraisal' of the stimulus (Fan and Han, 2008; Decety et al., 2010). Indeed, the current results suggest that the observation of pain in others does not lead to a reliable modulation of early frontal ERP components thus undermining the idea that they are associated with automatic emotion sharing. The absence of a reliable dissociation of early ERP responses to pain and neutral stimuli suggests that these stages instead reflect the perceptual processing of the stimulus that is necessary to assess its affective content and raises question regarding the commonly accepted time course of the processes involved in the empathic response.

Indeed, the 'affective sharing response' to pain in others might not be as fast and as automatic and might rely more on the initial identification of the stimulus than previously thought (Coll et al., 2017b).

Another possibility is that the vicarious pain stimuli used in the studies reviewed led to a modulation of an early posterior components that is generally observed for emotional stimuli (Schupp et al., 2003a,b). Unfortunately, almost none of the studies reviewed reported early responses at posterior sites, thus preventing the comparison between ERP components that are claimed to reflect empathic processes and the components usually observed in response to emotional stimuli. Future research should further explore the influence of vicarious pain on early ERP components and the meaning of this effect by testing more thoroughly the spatial distribution of this effect and its relationship to other emotional ERP responses.

The results showed reliable effects of vicarious pain on the latter P3 and LPP components that are compatible with a wealth ERP research showing that emotional stimuli modulate the latter centro-parietal components (Schupp et al., 2003a; Hajcak et al., 2010). The P3 and LPP responses are often interpreted in terms of sustained attentional processing and cognitive evaluation of motivationally relevant stimuli (Hajcak et al., 2010; Lang and Bradley, 2010) that, in the context of empathy, might contribute to social understanding and emotional regulation (Decety et al., 2010; Cheng et al., 2012; Fan et al., 2014). The fact that the effect of pain observation on these components often interact with other

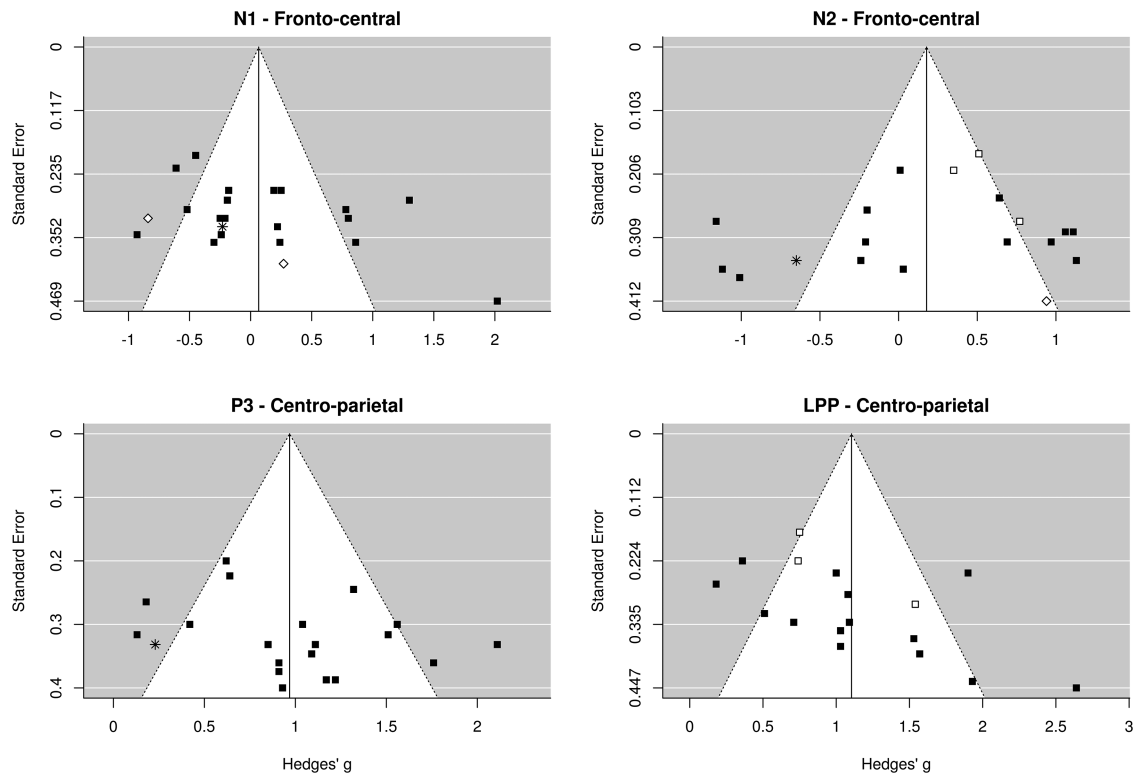


Fig. 8. Funnel plots showing the distribution of studies as a function of effect size and precision (standard error). The white area shows the 95% CI around the summary effect size for each degree of precision. Symbols show the type of stimulus used in each study (black square, static limbs; white square, dynamic limbs; white diamond, static faces; asterisk, other).

experimental or between-subject factors suggest that this process is highly susceptible to interference and that its alteration in some groups reflects social-emotional difficulties. However, since this modulation is commonly observed in response to many types of emotional stimuli, the idea that it represents a specific process of the empathic response in the context of pain empathy and not a general aversive/regulatory response remains to be established. Indeed, none of the study reviewed compared the ERP response to vicarious pain stimuli to non-social-emotional stimuli to assess the specificity and uniqueness of the processes indexed (Happé et al., 2017). It therefore seems imperative for future studies using ERP to investigate pain empathy to carefully evaluate the validity, reliability and specificity of ERP responses to pain in others.

Despite the lack of reliability of the early frontal effects observed here, a considerable proportion of the studies reviewed report significant effects of pain observation on the early frontal N1 and N2 components. This apparent contradiction could be explained by the methodological issues underlined in the systematic review of methodological practices indicating that there was considerable variability in the quantification and statistical analysis of the ERP data. Indeed, while most studies used the mean amplitude between fixed latencies to quantify ERPs, the time windows and the electrode locations used for this measure often varied considerably across studies without a clear rationale underlying this choice. The problems of analytical flexibility, sometimes called researcher's degrees of freedom, have already been discussed elsewhere (Simmons et al., 2011; Luck and Gaspelin, 2017). However, the current study suggests that previous investigations of pain empathy using ERPs might be compromised by this practice.

This analytical flexibility was often combined with a shotgun analytical approach in which a high number of statistical analyses were performed on several locations and time windows and any significant effect was interpreted as meaningful. On average, the studies reviewed here performed on average over 100 statistical tests on ERP data and none corrected the significance threshold to reduce the risk of a false-positive finding. This is problematic since such a high number of statistical tests leads to a high probability that several significant results are in fact false positives (Kilner, 2013; Luck and Gaspelin, 2017). To diminish this risk, researchers can use analytical techniques that can take into account the spatial and temporal distribution of ERP data to reduce the number of comparisons or to adequately control for them (Maris and Oostenveld, 2007; Groppe et al., 2011; Pernet et al., 2015).

Alternatively, researchers can restrict their analyses to scalp regions and time windows for which an effect was predicted. However, a worrying result of this study is the fact that in several cases, the analysis and interpretation of the data were not constrained by the researchers' predictions since no clear hypotheses were formulated. This suggests that despite using the confirmatory analytical approach of null hypothesis significance testing, most of ERP research in the field of pain empathy is exploratory (Wagenmakers et al., 2012), even after 10 years and 40 studies. A solution to this issue would be to require researchers to clearly formulate their research hypotheses and label as exploratory the results of analyses that were not predicted. Ideally, researchers could pre-register their hypotheses online or publish using the pre-registered report format to establish the analysis plan before collecting data (Munafò et al., 2017).

Another striking observation permitted by this review is the lack of comprehensive reporting for the results from the

statistical analysis of ERP components. Indeed, it was found that the vast majority of the studies only reported significant results from a large number of factorial analyses. Therefore, in addition to a potential publication bias (Rosenthal, 1979), the ERP studies reviewed here also show a within-study reporting bias according to which analyses leading to non-significant results are less likely to be reported. Putting aside the fact that negative results can sometimes be informative if statistical power is high enough (Greenwald, 1975), the main consequence of this practice is that any attempt to meta-analytically summarise the results of such studies will be difficult, inevitably biased and of questionable usefulness (Moran et al., 2017). Therefore, it should be noted that the effect size calculated in the meta-analysis performed in this report are probably inflated. In several cases, this practice was justified by the necessity to provide a concise and brief report of the results. The short-term solution to this issue is to make it mandatory for authors to appropriately report all the results of all statistical analyses performed on ERP components in the text or in Supplementary Materials. It would also be beneficial for the field of ERP research to adopt a standard reporting procedure that would enable the automatic extraction of results from published articles and facilitate meta-analyses and large-scale automated summaries of all published studies. For example, the field of functional magnetic resonance imaging research has taken advantage of the standard reporting of activation coordinates in tables to produce automated meta-analytical tools (Yarkoni et al., 2011). A more preferable long-term solution would be to favorise the sharing of ERP data in online repository that would allow the reanalysis and meta-analysis of large data sets and a quick and efficient assessment of the evidence for specific effects (Poldrack and Gorgolewski, 2014).

A related issue to the incomplete reporting of results was incomplete and unclear reporting of several important methodological details. For example, less than half of the studies reported the number of trials left after artifact rejection procedures, meaning that it was unknown how many trials were included in the analyses in a majority of cases. To avoid omitting to report such crucial information, researchers should refer to published guidelines for reporting of EEG experiments (Picton et al., 2000; Keil et al., 2014; Moran et al., 2017) and reviewers should enforce these guidelines in all relevant cases.

Finally, an estimation of the statistical power for each study revealed that most studies reviewed were only adequately powered to detect large effect sizes. The lack of consideration for statistical power in ERP research has been discussed previously. Future research should provide adequate justifications for their sample sizes (see Larson and Carbine, 2017 for a review and suggestions). Low statistical power can not only prevent the detection of small effects (type 2 error) but may also lead to an imprecise estimation of the effect sizes and to type 1 errors (Button et al., 2013). This imprecise estimation of the effect sizes could have contributed to the excess significance and heterogeneous effect sizes observed for the early components. Because of this and the issue of selective reporting discussed above, the results from the meta-analysis are most likely biased. This means that while they can be used to roughly guide future investigations, the precise value of the meta-analytic effect sizes reported here should not be taken at face value. Ideally, future investigations should consider adopting a level of statistical power that allows the detection of the smallest effect size of interest (see Lakens, 2014 for a discussion of the smallest effect size of interest). Other factors such as the

signal-to-noise ratio can also contribute to statistical power and should be considered when planning ERP experiments (Boudewyn et al., 2018).

Limitations

The high heterogeneity in the experimental designs and analytical approaches as well as partial reporting of results in the studies analysed prevented an analysis of factors modulating the ERPs to vicarious pain. It is therefore possible that the lack of effect found for the early components is due to a moderation by other factors that could not be assessed. For example, the impact of the nature of the stimuli used (e.g. limbs versus faces) on the ERP responses should be carefully investigated in future studies. The improvement of methods and reporting in future studies should allow a more comprehensive analysis of such factors potentially modulating the effects observed.

The methodological issues highlighted in this report are not specific to the domain of pain empathy and the present observations could potentially be generalised to many other fields of research in cognitive neuroscience (e.g. Hobson and Bishop, 2017 for similar observations on EEG studies of action observation). Furthermore, while several issues were found to be prevalent in the studies reviewed, the scientific quality and usefulness of all papers cited in this report should be assessed on an individual basis. Finally, the solutions proposed to the issues raised are not exhaustive, nor can they be applied indiscriminately to all ERP research.

Conclusion

In conclusion, this study provides meta-analytic evidence for a robust modulation of the latter, but not early ERP components during pain observation. Furthermore, it suggests that the current framework used in pain empathy research using ERPs is undermined by several methodological problems that raise doubts regarding the reliability, validity and overall usefulness of this research. Researchers in the field should take into account the methodological issues raised here when designing and reviewing ERP experiments. This is of critical importance if this paradigm is to be used to draw conclusions on social-emotional functioning in different clinical populations.

Supplementary data

Supplementary data are available at SCAN online.

Funding

MP Coll is funded by a fellowship from the Canadian Institutes of Health Research and was funded by a fellowship from the Fonds de recherche du Quebec – Sante during this work.

Conflict of interest. None declared.

References

- *Included in the systematic review
- **Included in the systematic review and meta-analysis
- Batson, C.D. (2009). These things called empathy: eight related but distinct phenomena. In: J. Decety & W. Ickes, editors. *The Social Neuroscience of Empathy* (pp. 3–16). Cambridge, MA, US: MIT Press. DOI: 10.7551/mitpress/9780262012973.003.0002.

- Bird, G., Viding, E. (2014). The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520–32. DOI: [10.1016/j.neubiorev.2014.09.021](https://doi.org/10.1016/j.neubiorev.2014.09.021).
- Boudewyn, M.A., Luck, S.J., Farrens, J.L., Kappenman, E.S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. DOI: [10.1111/psyp.13049](https://doi.org/10.1111/psyp.13049).
- Brozek, J., Alexander, H. (1950). The formula $t^2 = f$. *The American Journal of Psychology*, *63*(2), 262–9. DOI: [10.2307/1418933](https://doi.org/10.2307/1418933).
- Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365.
- Carp, J. (2012). The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage*, *63*(1), 289–300. DOI: [10.1016/j.neuroimage.2012.07.004](https://doi.org/10.1016/j.neuroimage.2012.07.004).
- **Chen, C., Yang, C.-Y., Cheng, Y. (2012). Sensorimotor resonance is an outcome but not a platform to anticipating harm to others. *Society for Neuroscience*, *7*(6), 578–90. DOI: [10.1080/17470919.2012.686924](https://doi.org/10.1080/17470919.2012.686924).
- **Cheng, Y., Chen, C., Decety, J. (2014). An EEG/ERP investigation of the development of empathy in early and middle childhood. *Developmental Cognitive Neuroscience*, *10*, 160–9. DOI: [10.1016/j.dcn.2014.08.012](https://doi.org/10.1016/j.dcn.2014.08.012).
- **Cheng, Y., Hung, A.-Y., Decety, J. (2012). Dissociation between affective sharing and emotion understanding in juvenile psychopaths. *Development of Psychopathology*, *24*(2), 623–36. DOI: [10.1017/S095457941200020X](https://doi.org/10.1017/S095457941200020X).
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–9. DOI: [10.1037/0033-2909.112.1.155](https://doi.org/10.1037/0033-2909.112.1.155).
- **Coll, M.-P., Tremblay, M.-P.B., Jackson, P.L. (2017a). The effect of tDCS over the right temporo-parietal junction on pain empathy. *Neuropsychologia*, *100*, 110–19. DOI: [10.1016/j.neuropsychologia.2017.04.021](https://doi.org/10.1016/j.neuropsychologia.2017.04.021).
- Coll, M.-P., Viding, E., Rütgen, M., et al. (2017b). Are we really measuring empathy? Proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews*, *83*, 132–39. DOI: [10.1016/j.neubiorev.2017.10.009](https://doi.org/10.1016/j.neubiorev.2017.10.009).
- **Contreras-Huerta, L.S., Hielscher, E., Sherwell, C. S., Rens, N., Cunnington, R. (2014). Intergroup relationships do not reduce racial bias in empathic neural responses to pain. *Neuropsychologia*, *64*, 263–70. DOI: [10.1016/j.neuropsychologia.2014.09.045](https://doi.org/10.1016/j.neuropsychologia.2014.09.045).
- Cooper, H., Hedges, L. (1994). *The Handbook of Research Synthesis and Meta-Analysis*, New York: Russel Sage Foundation.
- **Corbera, S., Ikezawa, S., Bell, M.D., Wexler, B.E. (2014). Physiological evidence of a deficit to enhance the empathic response in schizophrenia. *European Psychiatry*, *29*(8), 463–72. DOI: [10.1016/j.eurpsy.2014.01.005](https://doi.org/10.1016/j.eurpsy.2014.01.005).
- Cuff, B.M.P., Brown, S.J., Taylor, L., Howat, D. J. (2016). Empathy: a review of the concept. *Emotion Review: Journal of the International Society for Research on Emotion*, *8*(2), 144–53. DOI: [10.1177/1754073914558466](https://doi.org/10.1177/1754073914558466).
- **Cui, F., Ma, N., Luo, Y.-J. (2016a). Moral judgment modulates neural responses to the perception of other's pain: an ERP study. *Scientific Reports*, *6*(1), 20851. DOI: [10.1038/srep20851](https://doi.org/10.1038/srep20851).
- **Cui, F., Zhu, X., Duan, F., Luo, Y. (2016b). Instructions of cooperation and competition influence the neural responses to others' pain: an ERP study. *Society for Neuroscience*, *11*(3), 289–96. DOI: [10.1080/17470919.2015.1078258](https://doi.org/10.1080/17470919.2015.1078258).
- **Cui, F., Zhu, X., Gu, R., Luo, Y.-J. (2016c). When your pain signifies my gain: neural activity while evaluating outcomes based on another person's pain. *Scientific Reports*, *6*(0), 26426. DOI: [10.1038/srep26426](https://doi.org/10.1038/srep26426).
- **Cui, F., Zhu, X., Luo, Y. (2017a). Social contexts modulate neural responses in the processing of others' pain: an event-related potential study. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(4), 850–7. DOI: [10.3758/s13415-017-0517-9](https://doi.org/10.3758/s13415-017-0517-9).
- **Cui, F., Zhu, X., Luo, Y., Cheng, J. (2017b). Working memory load modulates the neural response to other's pain: evidence from an ERP study. *Neuroscience*, *644*, 24–29.
- Davis, M.H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113.
- de Vignemont, F., Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences (Regul Ed)*, *10*(10), 435–41. DOI: [10.1016/j.tics.2006.08.008](https://doi.org/10.1016/j.tics.2006.08.008).
- Decety, J., Jackson, P.L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. DOI: [10.1177/1534582304267187](https://doi.org/10.1177/1534582304267187).
- **Decety, J., Lewis, K.L., Cowell, J.M. (2015). Specific electrophysiological components disentangle affective sharing and empathic concern in psychopathy. *Journal of Neurophysiology*, *114*(1), 493–504. DOI: [10.1152/jn.00253.2015](https://doi.org/10.1152/jn.00253.2015).
- **Decety, J., Meidenbauer, K.L., Cowell, J.M. (2017). The development of cognitive empathy and concern in preschool children: a behavioral neuroscience investigation. *Developmental Science*. DOI: [10.1111/desc.12570](https://doi.org/10.1111/desc.12570).
- Decety, J., Moriguchi, Y. (2007). The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions. *Biopsychosocial Medicine*, *1*, 22. DOI: [10.1186/1751-0759-1-22](https://doi.org/10.1186/1751-0759-1-22).
- **Decety, J., Yang, C.-Y., Cheng, Y. (2010). Physicians down-regulate their pain empathy response: an event-related brain potential study. *Neuroimage*, *50*(4), 1676–82. DOI: [10.1016/j.neuroimage.2010.01.025](https://doi.org/10.1016/j.neuroimage.2010.01.025).
- Egger, M., Davey Smith, G., Schneider, M., Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *The BMJ*, *315*(7109), 629–34. DOI: [10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629).
- **Fabi, S., Leuthold, H. (2016). Empathy for pain influences perceptual and motor processing: evidence from response force, ERPS, and EEG oscillations. *Society for Neuroscience*, *12*, 1–16. DOI: [10.1080/17470919.2016.1238009](https://doi.org/10.1080/17470919.2016.1238009).
- **Fabi, S., Leuthold, H. (2018). Racial bias in empathy: do we process dark-and fair-colored hands in pain differently? An EEG study. *Neuropsychologia*, *114*, 143–157.
- **Fallon, N., Li, X., Chiu, Y., Nurmikko, T., Stancak, A. (2015). Altered cortical processing of observed pain in patients with fibromyalgia syndrome. *J Pain*, *16*(8), 717–26. DOI: [10.1016/j.jpain.2015.04.008](https://doi.org/10.1016/j.jpain.2015.04.008).
- **Fan, Y., Han, S. (2008). Temporal dynamic of neural mechanisms involved in empathy for pain: an event-related brain potential study. *Neuropsychologia*, *46*(1), 160–173. DOI: [10.1016/j.neuropsychologia.2007.07.023](https://doi.org/10.1016/j.neuropsychologia.2007.07.023).
- **Fan, Y.-T., Chen, C., Chen, S.-C., Decety, J., Cheng, Y. (2014). Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. *Social Cognitive and Affective Neuroscience*, *9*(8), 1203–13. DOI: [10.1093/scan/nst101](https://doi.org/10.1093/scan/nst101).
- *Fitzgibbon, B.M., Enticott, P.G., Giummarra, M.J., Thomson, R.H., Georgiou-Karistianis, N., Bradshaw, J.L. (2012). Atypical electrophysiological activity during pain observation in amputees who experience synaesthetic pain. *Social Cognitive and Affective Neuroscience*, *7*(3), 357–68. DOI: [10.1093/scan/nsr016](https://doi.org/10.1093/scan/nsr016).

- **Gonzalez-Liencre, C., Breidenstein, A., Wolf, O.T., Brüne, M. (2016a). Sex-dependent effects of stress on brain correlates to empathy for pain. *International Journal of Psychophysiology*, *105*, 47–56. DOI: [10.1016/j.ijpsycho.2016.04.011](https://doi.org/10.1016/j.ijpsycho.2016.04.011).
- **Gonzalez-Liencre, C., Brown, E.C., Tas, C., Breidenstein, A., Brüne, M. (2016b). Alterations in event-related potential responses to empathy for pain in schizophrenia. *Psychiatry Research*, *241*, 14–21. DOI: [10.1016/j.psychres.2016.04.091](https://doi.org/10.1016/j.psychres.2016.04.091).
- Greenwald, A.G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*(1), 1–20. DOI: [10.1037/h0076157](https://doi.org/10.1037/h0076157).
- Groppe, D.M., Urbach, T.P., Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields i: a critical tutorial review. *Psychophysiology*, *48*(12), 1711–25. DOI: [10.1111/j.1469-8986.2011.01273.x](https://doi.org/10.1111/j.1469-8986.2011.01273.x).
- Hajcak, G., MacNamara, A., Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Developmental Neuropsychology*, *35*(2), 129–55. DOI: [10.1080/87565640903526504](https://doi.org/10.1080/87565640903526504).
- *Han, S., Fan, Y., Mao, L. (2008). Gender difference in empathy for pain: an electrophysiological investigation. *Brain Research*, *1196*, 85–93. DOI: [10.1016/j.brainres.2007.12.062](https://doi.org/10.1016/j.brainres.2007.12.062).
- Happé, F., Cook, J.L., Bird, G. (2017). The structure of social cognition: in(ter)dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243–67. DOI: [10.1146/annurev-psy-010416-044046](https://doi.org/10.1146/annurev-psy-010416-044046).
- Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107–28.
- Hobson, H.M., Bishop, D.V.M. (2017). The interpretation of mu suppression as an index of mirror neuron activity: past, present and future. *Royal Society Open Science*, *4*(3), 160662. DOI: [10.1098/rsos.160662](https://doi.org/10.1098/rsos.160662).
- **Ibáñez, A., Hurtado, E., Lobos, A., et al. (2011). Subliminal presentation of other faces (but not own face) primes behavioral and evoked cortical processing of empathy for pain. *Brain Research*, *1398*, 72–85. DOI: [10.1016/j.brainres.2011.05.014](https://doi.org/10.1016/j.brainres.2011.05.014).
- *Ikezawa, S., Corbera, S., Liu, J., Wexler, B. E. (2012). Empathy in electrodermal responsive and nonresponsive patients with schizophrenia. *Schizophrenia Research*, *142*(1-3), 71–6. DOI: [10.1016/j.schres.2012.09.011](https://doi.org/10.1016/j.schres.2012.09.011).
- **Ikezawa, S., Corbera, S., Wexler, B.E. (2014). Emotion self-regulation and empathy depend upon longer stimulus exposure. *Social Cognitive and Affective Neuroscience*, *9*(10), 1561–8. DOI: [10.1093/scan/nst148](https://doi.org/10.1093/scan/nst148).
- Ioannidis, J.P., Trikalinos, T.A. (2007). An exploratory test for an excess of significant findings. *Clinical. Trials*, *4*(3), 245–53. DOI: [10.1177/1740774507079441](https://doi.org/10.1177/1740774507079441).
- Jackson, P.L., Meltzoff, A.N., Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *Neuroimage*, *24*(3), 771–9. DOI: [10.1016/j.neuroimage.2004.09.006](https://doi.org/10.1016/j.neuroimage.2004.09.006).
- **Jiang, C., Varnum, M.E.W., Hou, Y., Han, S. (2014). Distinct effects of self-construal priming on empathic neural responses in chinese and westerners. *Society for Neuroscience*, *9*(2), 130–8. DOI: [10.1080/17470919.2013.867899](https://doi.org/10.1080/17470919.2013.867899).
- **Kam, J.W.Y., Xu, J., Handy, T.C. (2014). I don't feel your pain (as much): the desensitizing effect of mind wandering on the perception of others' discomfort. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 286–96. DOI: [10.3758/s13415-013-0197-z](https://doi.org/10.3758/s13415-013-0197-z).
- Keil, A., Debener, S., Gratton, G., et al. (2014). Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, *51*(1), 1–21. DOI: [10.1111/psyp.12147](https://doi.org/10.1111/psyp.12147).
- Kilner, J. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology*, *124*(10), 2062–3. DOI: [10.1016/j.clinph.2013.03.024](https://doi.org/10.1016/j.clinph.2013.03.024).
- Klimecki, O., Singer, T. (2013). Empathy from the perspective of social neuroscience. In J. Armony & P. Vuilleumier (Eds.), *The Cambridge Handbook of Human Affective Neuroscience* (pp. 533–550). Cambridge: Cambridge University Press. DOI: [10.1017/CBO9780511843716.029](https://doi.org/10.1017/CBO9780511843716.029).
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*(7), 701–10.
- Lang, P.J., Bradley, M.M. (2010). Emotion and the motivational brain. *Biological Psychology*, *84*(3), 437–50.
- Larson, M.J., Carbine, K.A. (2017). Sample size calculations in human electrophysiology (EEG and ERP) studies: a systematic review and recommendations for increased rigor. *International Journal of Psychophysiology*, *111*, 33–41. DOI: [10.1111/psyp.13049](https://doi.org/10.1111/psyp.13049).
- **Li, W., Han, S. (2010). Perspective taking modulates event-related potentials to perceived pain. *Neuroscience Letters*, *469*(3), 328–32. DOI: [10.1016/j.neulet.2009.12.021](https://doi.org/10.1016/j.neulet.2009.12.021).
- Luck, S.J., Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–57. DOI: [10.1111/psyp.12639](https://doi.org/10.1111/psyp.12639).
- **Lyu, Z., Meng, J., Jackson, T. (2014). Effects of cause of pain on the processing of pain in others: an ERP study. *Experimental Brain Research*, *232*(9), 2731–9. DOI: [10.1007/s00221-014-3952-7](https://doi.org/10.1007/s00221-014-3952-7).
- Maris, E., Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–90. DOI: [10.1016/j.jneumeth.2007.03.024](https://doi.org/10.1016/j.jneumeth.2007.03.024).
- **Mella, N., Studer, J., Gilet, A.-L., Labouvie-Vief, G. (2012). Empathy for pain from adolescence through adulthood: An event-related brain potential study. *Frontiers in Psychology*, *3*, 501. DOI: [10.3389/fpsyg.2012.00501](https://doi.org/10.3389/fpsyg.2012.00501).
- **Meng, J., Hu, L., Shen, L., et al. (2012). Emotional primes modulate the responses to others' pain: an ERP study. *Experimental Brain Research*, *220*(3-4), 277–86. DOI: [10.1007/s00221-012-3136-2](https://doi.org/10.1007/s00221-012-3136-2).
- **Meng, J., Jackson, T., Chen, H., et al. (2013). Pain perception in the self and observation of others: an ERP investigation. *Neuroimage*, *72*, 164–73. DOI: [10.1016/j.neuroimage.2013.01.024](https://doi.org/10.1016/j.neuroimage.2013.01.024).
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. DOI: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097).
- Moran, T.P., Schroder, H.S., Kneip, C., Moser, J.S. (2017). Meta-analysis and psychophysiology: a tutorial using depression and action-monitoring event-related potentials. *International Journal of Psychophysiology*, *111*, 17–32. DOI: [10.1016/j.ijpsycho.2016.07.001](https://doi.org/10.1016/j.ijpsycho.2016.07.001).
- Munafo, M., Nosek, B., Bishop, D., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. DOI: [10.1038/s41562-016-0021](https://doi.org/10.1038/s41562-016-0021).
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. DOI: [10.1155/2011/156869](https://doi.org/10.1155/2011/156869).
- Pernet, C.R., Latinus, M., Nichols, T.E., Rousselet, G.A. (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: a simulation

- study. *The Journal of Neuroscience Methods*, 250, 85–93. DOI: [10.1016/j.jneumeth.2014.08.003](https://doi.org/10.1016/j.jneumeth.2014.08.003).
- Picton, T.W., Bentin, S., Berg, P., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37(2), 127–52.
- Poldrack, R.A., Fletcher, P.C., Henson, R.N., Worsley, K.J., Brett, M., Nichols, T.E. (2008). Guidelines for reporting an fMRI study. *Neuroimage*, 40(2), 409–14. DOI: [10.1016/j.neuroimage.2007.11.048](https://doi.org/10.1016/j.neuroimage.2007.11.048).
- Poldrack, R.A., Gorgolewski, K.J. (2014). Making big data open: data sharing in neuroimaging. *Nature Neuroscience*, 17(11), 1510–17. DOI: [10.1038/nn.3818](https://doi.org/10.1038/nn.3818).
- R.Core Team (2018). *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–41. RStudio Team (2015). RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Inc.
- Schupp, H.T., Junghöfer, M., Weike, A.I., Hamm, A. O. (2003a). Attention and emotion: an ERP analysis of facilitated emotional stimulus processing. *Neuroreport*, 14(8), 1107–10. DOI: [10.1097/01.wnr.0000075416.59944.49](https://doi.org/10.1097/01.wnr.0000075416.59944.49).
- Schupp, H.T., Junghöfer, M., Weike, A.I., Hamm, A.O. (2003b). Emotional facilitation of sensory processing in the visual cortex. *Psychological Science*, 14(1), 7–13. DOI: [10.1111/1467-9280.01411](https://doi.org/10.1111/1467-9280.01411).
- **Sessa, P., Meconi, F., Castelli, L., Dell’Acqua, R. (2014). Taking one’s time in feeling other-race pain: an event-related potential investigation on the time-course of cross-racial empathy. *Social Cognitive and Affective Neuroscience*, 9(4), 454–63. DOI: [10.1093/scan/nst003](https://doi.org/10.1093/scan/nst003).
- Simmons, J.P., Nelson, L.D., Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–66. DOI: [10.1177/0956797611417632](https://doi.org/10.1177/0956797611417632).
- *Sun, Y.-B., Lin, X.-X., Ye, W., Wang, N., Wang, J.-Y., Luo, F. (2017). A screening mechanism differentiating true from false pain during empathy. *Scientific Reports*, 7(1), 11492. DOI: [10.1038/s41598-017-11963-x](https://doi.org/10.1038/s41598-017-11963-x).
- **Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports*, 5, 15924. DOI: [10.1038/srep15924](https://doi.org/10.1038/srep15924).
- Vachon-Preseu, E., Roy, M., Martel, M.O., et al. (2012). Neural processing of sensory and emotional-communicative information associated with the perception of vicarious pain. *Neuroimage*, 63(1), 54–62. DOI: [10.1016/j.neuroimage.2012.06.030](https://doi.org/10.1016/j.neuroimage.2012.06.030).
- **Vaes, J., Meconi, F., Sessa, P., Olechowski, M. (2016). Minimal humanity cues induce neural empathic reactions towards non-human entities. *Neuropsychologia*, 89, 132–40. DOI: [10.1016/j.neuropsychologia.2016.06.004](https://doi.org/10.1016/j.neuropsychologia.2016.06.004).
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L.J., Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–8. DOI: [10.1177/1745691612463078](https://doi.org/10.1177/1745691612463078).
- **Wang, Y., Song, J., Guo, F., Zhang, Z., Yuan, S., Cacioppo, S. (2016). Spatiotemporal brain dynamics of empathy for pain and happiness in friendship. *Frontiers in Behavioral Neuroscience*, 10, 45. DOI: [10.3389/fnbeh.2016.00045](https://doi.org/10.3389/fnbeh.2016.00045).
- **Wexler, B.E., Ikezawa, S., Corbera, S. (2014). Increasing stimulus duration can normalize late-positive event related potentials in people with schizophrenia: possible implications for understanding cognitive deficits. *Schizophrenia Research*, 158(1-3), 163–169. DOI: [10.1016/j.schres.2014.07.012](https://doi.org/10.1016/j.schres.2014.07.012).
- **Yang, J., Hu, X., Li, X., et al. (2017). Decreased empathy response to other people’s pain in bipolar disorder: evidence from an event-related potential study. *Scientific Reports*, 7, 39903. DOI: [10.1038/srep39903](https://doi.org/10.1038/srep39903).
- Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8), 665–70. DOI: [10.1038/nmeth.1635](https://doi.org/10.1038/nmeth.1635).