Author(s): Li, Xueqiao; Pesonen, Janne; Haimi, Elina; Wang, Huili; Astikainen, Piia

Title: Electrical brain activity and facial electromyography responses to irony in dysphoric and non-dysphoric participants

Year: 2020

Version: Accepted version (Final draft)

Please cite the original version:

Li, X., Pesonen, J., Haimi, E., Wang, H., & Astikainen, P. (2020). Electrical brain activity and facial electromyography responses to irony in dysphoric and non-dysphoric participants. Brain and Language, 211, Article 104861. https://doi.org/10.1016/j.bandl.2020.104861

Electrical Brain Activity and Facial Electromyography Responses to Irony in Dysphoric and Non-Dysphoric Participants

Xueqiao Li*[a], Janne Pesonen[a], Elina Haimi[a], Huili Wang[b], Piia Astikainen[a]

Affiliation:
[a] Department of Psychology
University of Jyvaskyla
P. O. Box 35
FIN-40100 Jyväskylä
Finland

[b] School of Foreign Languages
Dalian University of Technology
Dalian
China

Xueqiao Li:
E-mail: xueqiao.x.li@jyu.fi

Janne Pesonen:
E-mail: janne.matias.pesonen@gmail.com

Elina Haimi:
E-mail: eea.haimi@gmail.com

Huili Wang:
E-mail: huiliw@dlut.edu.cn

Piia Astikainen:
E-mail: piia.astikainen@jyu.fi

* *Corresponding author*: Xueqiao Li

Abstract

We studied irony comprehension and emotional reactions to irony in dysphoric and control participants. Electroencephalography (EEG) and facial electromyography (EMG) were measured when spoken conversations were presented with pictures that provided either congruent (non-ironic) or incongruent (ironic) contexts. In a separate session, participants evaluated the congruency and valence of the stimuli. While both groups rated ironic stimuli funnier than non-ironic stimuli, the control group rated all the stimuli funnier than the dysphoric group. N400-like activity, P600, and EMG activity indicating smiling were larger after the ironic stimuli than the non-ironic stimuli for both groups. Further, in the dysphoric group the irony modulation was evident in the electrode cluster over the right hemisphere, while no such difference in lateralization was observed in the control group. The results suggest a depression-related alteration in the P600 response associated to irony comprehension, but no alterations were found in emotional reactivity specifically related to irony.

Keywords: event-related potentials; facial electromyography; N400; P600; irony; depressive symptoms.

1 Introduction

Generally, irony has been considered to be evoked from a comparison: a disparity between what is said and what is intended to be said (Grice, 1975). Imagine that you are on a sailing trip. Your boat is not moving, as there is no wind. The trip is becoming boring because the condition has remained windless for several hours. Your friend says, “What an exciting day for sailing!” This comment, which contrasts with the existing feeling, is a good example of an ironic statement.

The present study aimed to investigate irony comprehension and irony-related emotional reactions, and identify possible differences between dysphoric (individuals with elevated number of depressive symptoms) and non-dysphoric participants in these aspects. Facial electromyography (EMG) and event-related potentials (ERPs) were used to measure emotional facial expressions and cognitive processing related to irony, respectively. Differences in the irony-related emotional reactions and irony comprehension were expected between dysphoric and control groups, because previous studies have found alterations in emotional reactivity (for reviews, see Bylsma et al., 2008; Cusi et al., 2012), and cognitive function (for a review, see Gotlib and Joormann, 2010), such as verbal fluency (Henry and Crawford, 2005) and mentalizing (Bora and Berg, 2016) in clinical and preclinical depression. Impairments in irony comprehension have been found in schizophrenia and schizotypy (e.g., Del Goleto et al., 2016; Herold et al., 2018; Varga et al., 2013); however, whether or how irony comprehension is altered in depressed individuals has not yet been explored.

1.1 Emotional reactions to irony

Irony is widely used in everyday conversations to fulfill an implied goal in social communications. It can elicit positive feelings in the receiver, such as inducing amusement and humor (Calmus and Caillies, 2014; Dews, Kaplan, and Winner, 1995; Dynel, 2009; Martin, 2010; Matthews, Hancock, and Dunham, 2006), but it can also increase or reduce the listeners' negative feelings (Leggitt and Gibbs, 2000; Toplak and Katz, 2000). Relationship between conscious valence ratings and emotional reactions to irony can also be complex: when irony was directed towards the participants, behavioral evaluations showed that the participants perceived the ironic stories to be more humorous than the literal ones, but the ironic stories also elicited more negative emotions than the literal stories in the participants (Akimoto et al., 2014).

In addition to behavioral ratings, psychophysiological reactivity, such as facial muscle activity (i.e., facial EMG), is of interest when attempting to understand irony-evoked emotions (Thompson, Mackenzie, and Leuthold, 2016). Facial reactivity in the zygomaticus major muscle (cheek area) and in the corrugator supercilii muscle (brow region) is an index of emotional expressions corresponding to smiling and frowning, respectively (Van Boxtel, 2010; Dimberg, 1990). Facial EMG has been measured in response to humor (e.g., Fiacconi and Owen, 2015), but only a few studies have investigated facial EMG responses to irony. In one study, facial EMG was measured in relation to written ironic praise and criticism, with and without emoticons (Thompson et al., 2016); it was found that ironic criticism reduced frowning and enhanced smiling in comparison to literal criticism, indicating weakening of the negative emotional response due to irony. However, no previous studies have investigated facial reactivity to spoken ironic statements.

1.2 Theoretical models and empirical evidence of irony comprehension

Different models have been used to explain irony comprehension (for a review, see Attardo, 2000). The one-stage model (also called the direct access view, e.g., Gibbs, 2002) in general proposes that the figurative language (written or oral), such as irony, can be accessed directly as the literal language. In contrast, two-stage models, such as the traditional standard pragmatic view (e.g., Grice, 1975) and the graded salience hypothesis (e.g., Giora and Fein, 1999; Giora, 2003), suggest that irony comprehension requires access to both literal and ironic meanings. The former expects that, in irony comprehension, one should first access the literal meaning, detect and distinguish the discrepancy in the semantic context, and then reconstruct the information to retrieve the intended ironic meaning. The graded salience hypothesis assumes that there are two stages in irony comprehension, but it makes no explicit hypothesis about their sequential order.

Most empirical studies have supported the two-stage processing of irony. Behavioral reading paradigms have shown that ironic comments require longer reading times than literal ones (Giora, Fein, and Schwarts, 1998; Schwoebel, Dews, Winner, and Srinivas, 2000). Along the same lines, in eye tracking studies (e.g., Filik and Moxey, 2010; Kaakinen et al., 2014), the participants' fixation times were longer, and they spent more time in rereading ironic sentences than non-ironic sentences. These findings can be interpreted as evidence of additional processing demands related to irony, thus supporting the two-stage models.

The cognitive processes in irony comprehension have been investigated using time-sensitive ERP method in healthy participants (e.g., Balconi and Amenta, 2008; Baptista et al., 2018; Caillies et al., 2019; Cornejol et al., 2007; Filik et al., 2014; Regel et al., 2011, 2014; Spotorno et al., 2013). These studies have shown that comprehending ironic sentences requires additional inferential processes in comparison to non-ironic sentences; this finding is consistent with the traditional standard pragmatic view (e.g., Regel et al., 2011) or the graded salience hypothesis (e.g., Filik et al., 2014). In many of the studies (Cornejol et al., 2007; Filik et al., 2014; Caillies et al., 2019), two ERP components, N400 and P600, were elicited consecutively, and both were larger in amplitude for irony than non-irony, reflecting the two-stage processing of irony comprehension.

The N400 is an ERP component that is triggered by semantic violations and modulated by the context given to the statement (for reviews, see Kutas and Federmeier, 2011; Lau et al., 2008). It is observed as a shift towards negative polarity at 200–600 ms after a stimulus onset at the centro-parietal electrode sites (e.g., Kutas and Federmeier, 2000; Kutas and Hillyard, 1980; Van Petten and Luka, 2006). N400 amplitude modulations have been interpreted to reflect either an additional effort in semantic integration (Kutas and Federmeier, 2011; Lau et al., 2008) or the difficulty of retrieving the stored knowledge associated with the stimulus (Brouwer et al., 2012). Larger N400 amplitude has been observed during the processing of nonliteral language, for instance, humor (e.g., Coulson and Kutas, 2001; Feng, Chan, and Chen, 2014; Marinkovic et al., 2011) and irony (Baptista et al., 2018; Caillies et al., 2019; Cornejol et al., 2007; Filik et al., 2014). Several studies (Cornejol et al., 2007; Filik et al., 2014; Caillies et al., 2019) have demonstrated that the modulation of N400 reflects difficulties in integrating the meaning of an irony-related incongruent word into the context; thus, N400 modulation has been found to be one of the indicators reflecting the two-stage processing of irony. However, the findings related to irony are inconsistent, as the N400 effect is not always found in response to irony (e.g., Balconi and Amenta, 2008; Regel et al., 2011, 2014).

In contrast to the inconsistent findings on the N400 elicitation to irony, the amplitude of the P600 has been repeatedly found to be larger for ironic than non-ironic sentences (Filik et al., 2014; Regel et al., 2011, 2014; Spotorno et al., 2013). P600 is a shift towards positive polarity, reaching its maximum amplitude approximately at 600 ms latency at the centro-parietal electrode sites (for a review, see Bornkessel-Schlesewsky and Schlesewsky, 2008). It was initially observed when the participants encountered syntactic anomalies (e.g., "The woman encouraged to write a blog."; "the fancy very car") and it was interpreted as a reflection of syntactic reanalysis (e.g., Gouvea et al., 2010; Osterhout and Holcomb, 1992). The P600 effect has also been found in response to semantic violations (e.g., Nieuwland and Van Berkum, 2005; Sanford et al., 2011), which was interpreted as continued analysis to achieve conflict resolution or the updating of mental representations (for reviews, see Bornkessel-Schlesewsky and Schlesewsky, 2008; Brouwer et al., 2012; Kuperberg, 2007) in language comprehension. Increased P600 amplitude is often observed in response to irony (Filik et al., 2014; Regel et al., 2011, 2014; Spotorno et al., 2013), humor (e.g., Feng et al., 2014; Canal, et al., 2019; Shibata, et al., 2017; Marinkovic, et al., 2011) and figurative language in general (e.g., Bambini, et al., 2016). It has been suggested that the modulation of the P600 reflects the cognitive effort made when interpreting the implied meaning of nonliteral language (for irony, see e.g., Regel et al., 2011, 2014).

1.3 Depression and possible emotional and cognitive alterations related to irony comprehension

Depression is a severe mental health disorder characterized by both affective and cognitive symptoms, including sadness, tiredness, disturbances in sleep and appetite, and feelings of guilt or low self-worth (World Health Organization, 2010). Depressive individuals possess persistent mood-congruent rumination that negatively affects their ability to process emotional information (for reviews, see Gotlib and Joormann, 2010; Peckham, McHugh, and Otto, 2010). One of the most prominent features in depression is blunted reactivity to emotional stimuli (see a review, Bylsma et al., 2008), such as to pleasant pictures (e.g., Allen, Trinder, and Brennen, 1999; Dunn et al., 2004; Sloan, Strauss, and Wisner, 2001), positive words (Canli et al., 2004), and affective audiovisual videos (Rottenberg et al., 2002; 2005).

Concerning nonliteral language processing, humor comprehension has been found to be altered in depression (Uekermann et al., 2008). Depressed patients performed worse than the controls in selecting correct punchlines to humorous discussions, and they rated the humorous punchlines as being less funny than the controls (Uekermann et al., 2008). Alterations in humor processing have been associated with impairments in executive functions and mentalizing in depressed participants (Bora and Berg, 2016; Uekermann et al., 2008). Deficits in mentalizing have also been suggested to underlie the impairments in irony processing in individuals with schizophrenia and schizotypal personality disorder (e.g., Del Goleto et al., 2016; Herold et al., 2018; Varga et al., 2013), but no studies have investigated irony processing in people with depression. In addition to mentalizing, irony comprehension requires verbal fluency and verbal memory (e.g., Gaudreau et al., 2015; Spotorno et al., 2012), both of which are impaired in depression (Basso and Bornstein, 1999; Henry and Crawford, 2005). Based on the previous findings on the blunted emotional reactions and the cognitive deficits in depression, it is very likely that both emotional reactions to ironic stimuli and the cognitive processing aspect of irony are affected by depressive symptoms.

1.4 The present study

In the current study, conversational irony was investigated by presenting the participants with spoken dyadic conversations that resembled the natural occurrence of irony in the daily life. The keywords in the conversations were allocated to different positions, which made the stimuli less predictable and more naturalistic in comparison to previous studies in which the keyword was always presented as the last word of the sentence (e.g., Filik et al., 2014; Regel et al., 2011, 2014). The conversations were combined with pictures that were either congruent or incongruent with the statements; thus, the conversation was defined as either non-ironic or ironic, respectively. This approach allowed the responses to be contrasted to the same spoken sentences (non-ironic vs. ironic), avoiding the confounding caused by possible differences in the physical features of the stimuli, e.g., the position of the keywords.

In the first measurement day, facial EMG (corrugator and zygomaticus activity) and ERPs (N400 and P600) were measured while participants were passively observing the stimuli. To complement these measurements, evaluations of the congruency of the picture-sentence pairs and valence ratings (unpleasant vs. funny) of the conversations were also recorded in a separate measurement session on a different day. These data were collected on different days, as the rating task can influence participants' spontaneous emotional reactivity (Hutcherson et al., 2005). The use of relatively naturalistic stimulus conditions and a combination of different measures provide a comprehensive understanding of irony processing and how depressive symptoms possibly affects cognitive and/or emotional aspect of irony processing.

We hypothesized that the control participants would rate the ironic stimuli funnier than the non-ironic stimuli, because most of the previous studies have demonstrated this (Akimoto et al., 2014; Calmus and Caillies, 2014; Dews et al., 1995). However, it is also possible that the ironic stimuli would be rated as more unpleasant than the non-ironic stimuli, because several stimulus- and participant-related factors might affect the emotional reactions to irony, but these are not well known (Leggitt and Gibbs, 2000). Regardless, the ratings should be to the same direction with the EMG responses; that is, if participants rate ironic stimuli as more funny than non-ironic stimuli, a higher zygomaticus reactivity, reflecting smiling (Van Boxtel, 2010), should also be observed to ironic comparing to non-ironic stimuli; and if participants rate ironic stimuli more unpleasant than non-ironic stimuli, a higher activity in the corrugator supercili, reflecting frowning (Van Boxtel, 2010; Dimberg, 1990), should be observed.

Due to biases in emotional information processing in depression (e.g., Beck, 1967, 2008; Bylsma et al., 2008; Gotlib and Joormann, 2010; Xu et al., 2018), the valence ratings of the ironic stimuli were expected to be more negative for the dysphoric group than the control group. We also expected, based on previous findings of low emotional reactivity to both positive and negative stimuli in depression (e.g., Bylsma et al., 2008), that the difference in facial EMG responses between the non-ironic and ironic stimuli would be smaller in the dysphoric group than the control group. Since no difficulty related to detecting semantic violation is expected in the dysphoric group (Deldin et al., 2006; Klumpp et al., 2010), it was hypothesized that the accuracy of the congruency detection of the picture-sentence pairs should be high in both groups.

For the ERP responses, we hypothesized that an N400-like effect (Filik et al., 2014) would be similar in the dysphoric and control groups, as reported in previous studies where the depression group and the controls showed similar N400 amplitudes in semantic processing of congruent and incongruent sentence endings (Deldin et al., 2006; Klumpp et al., 2010). However, it is possible that the N400-like activity would not be elicited, as the recognition of semantic incongruence may not be a necessary processing stage for irony comprehension (Balconi and Amenta, 2008; Regel et al., 2011, 2014).

In addition, we hypothesized that irony would elicit a larger P600 response than non-ironic stimuli, as has been consistently reported in previous irony studies (e.g., Filik et al., 2014; Regel et al., 2011, 2014), and it is possible that the P600 response would be decreased in amplitude in the dysphoric group. While no previous studies have investigated the P600 response to ironic stimuli in depressed participants, one study has shown that P600 elicited in a working memory task can be used to distinguish between depressed and control participants through using a machine learning approach (Kalatzis et al, 2004). In schitzotypyal personality disorder, a decrease in P600 modulation in relation to irony has been associated mainly with difficulties in mentalizing (Del Goleto et al., 2016). Since there is also a deficit in mentalizing (Bora and Berk, 2016; Inoue et al., 2004; Lee et al., 2005) and in other cognitive functions in depression, such as verbal fluency and verbal memory (Basso and Bornstein, 1999; Henry and Crawford, 2005), it is possible that a decrease in P600 will be found in the dysphoric group.

2 Materials and methods

2.1 Participants

As there was no previous study investigating cognitive and affective aspects of irony processing in control and dysphoric or depressed participants, sample size in the present study was estimated based on a standard medium effect size (η_p^2 = .060; Cohen, 1988). Power analysis, conducted with G*Power 3 (Faul et al., 2007), showed a requirement of 17 participants in each group (control and dysphoric) with a statistical power of (1 - β) = .80 and a significance level of alpha = .050.

Two groups of volunteers were recruited in this study: dysphoric group and control group. Potential participants were informed about the study via email lists from the University of Jyväskylä and by distributing advertisement flyers in Jyväskylä. The study protocol was in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the University of Jyväskylä. All the participants gave written informed consent before the measurements started.

For the dysphoric group, participant with current depressive symptoms (13 or higher score in the Beck's Depression Inventory-II; BDI-II; Beck et al., 1996) were included. The participants in the control group had a BDI-II score below 10. All participants were right-handed native speakers of Finnish, with normal or corrected-to-normal vision and normal hearing. Participants were included if they reported no current or previous neurological or psychiatric disorders, except depression or anxiety disorders in the dysphoric group.

In total, twenty-four dysphoric participants and twenty-eight control participants (age range 18–40 years) volunteered for the study. Seven participants were excluded from the dysphoric group and another seven participants were excluded from the control group according to the exclusion criteria or because of extensive ocular artefacts in their EEG data. Therefore, 17 dysphoric participants (3 male, 14 female) and 21 control participants (6 male, 15 female) were included in the final sample. Sixteen out of 17 dysphoric participants reported having an existing diagnosis of depression. One of them had been diagnosed with a mixed anxiety and depressive disorder. Demographics and clinical information of the dysphoric and control group are presented in Table 1.

There was no significant difference in age, gender or education (all p-values > .254) between the two groups (Table 1). The cognitive capabilities were also evaluated in the dysphoric and control groups. The two groups did not differ in any of the cognitive tests measuring memory, executive functions and verbal fluency. Details of the cognitive test scores are shown in the Supplementary Table 1.

2.2 Procedure

Psychophysiological and behavioral measurements for each participant were conducted on two separate days.

On the first measurement day, facial EMG and EEG data were collected. A set of cognitive tests (see list in Supplementary Table 2) were conducted on the same day after the recording of the psychophysiological signals. The participants were informed that they were going to be presented with conversations between two people, along with related pictures. The purpose of the study (i.e., irony processing) was not revealed until the participants completed the whole study. During the measurement, participants were seated in a soundproofed room while the stimuli were presented. No task was designed for the participants during the first day's measurement. Participants were instructed to focus on the stimuli and avoid frequent eye blinks and unnecessary movements during the recording.

On the second measurement day, participants were seated at the same room. Congruency evaluation and valence rating (unpleasant vs. funny) for picture-sentence pairs were collected. The Humor Styles Questionnaire (HSQ; Martin et al., 2003) was filled out after the behavioral task. The analysis and results of Humor Styles Questionnaire are presented in Supplementary materials.

2.3 Stimuli

The same stimuli were applied for both facial EMG/EEG and behavioral measurements. Stimuli in each trial consisted of an introductory sentence, a contextual picture and a commenting sentence. The sentences were spoken in Finnish, with a neutral intonation by two female and two male speakers. The spoken conversations were presented from a ceiling loudspeaker above the participants, approximately one meter from participants' ears. The contextual pictures were taken from the internet and the International Affective Picture System (IAPS; Lang et al., 2008). They were color pictures depicting real-life scenes and presented to a screen with a resolution of 1024×768 pixels to a 23-inch screen (Asus VG236H, 1920×1080 pixels) approximately one meter in front of the participant.

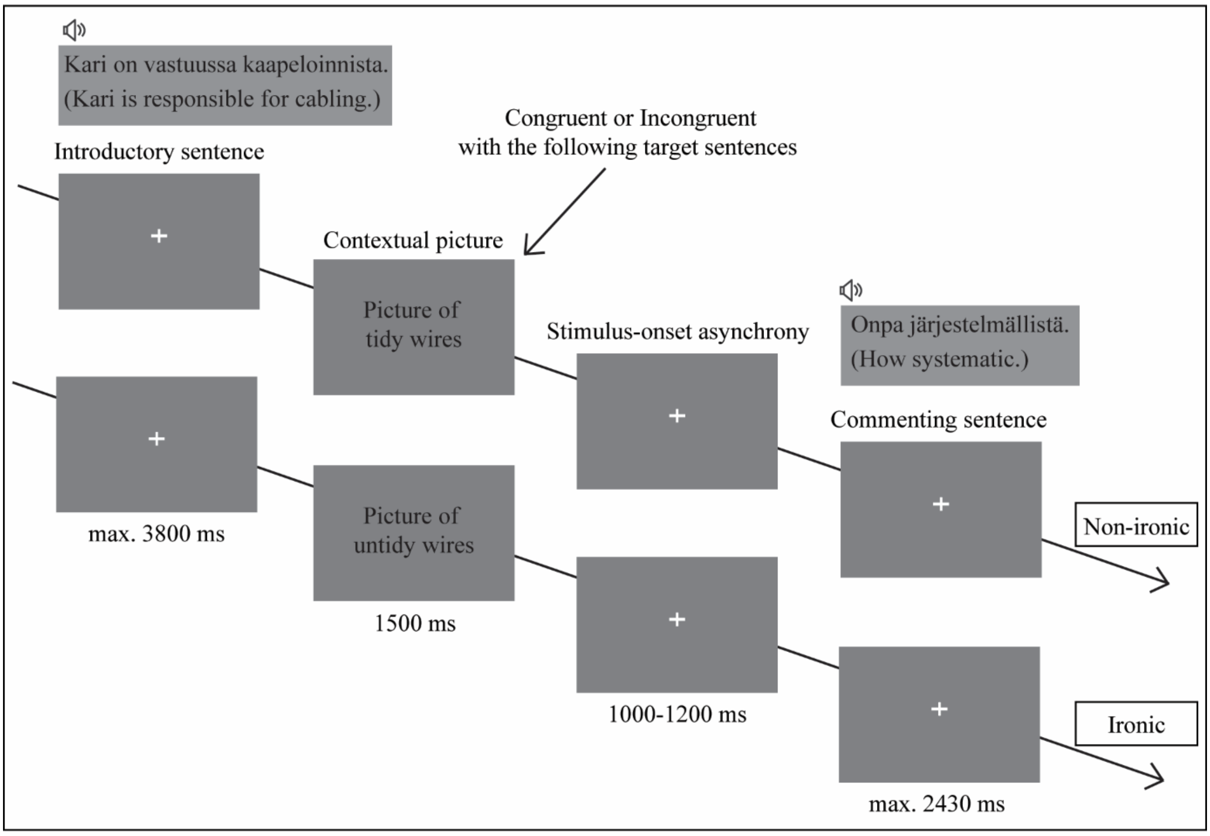


Figure 1. An example of one pair of trials. The sentences were spoken in Finnish, but for illustrative purposes, the translations are provided here. The changes of the contextual pictures brought different conditions (non-ironic or ironic), while the commenting sentences were always the same for each non-ironic and ironic pair.

The presentation of each experimental trial is illustrated in Figure 1. First, an introductory sentence was presented as a setup for the conversation. Next, a contextual picture was displayed on the center of the screen. Then, a fixation mark appeared to the screen, after which a spoken commenting sentence was presented together with the fixation mark. The commenting sentence was spoken by a different person than the introductory sentence providing a comment to the introductory sentence. The same conversations were presented twice during the experiment: with a picture which provided a congruent context to the commenting sentence (non-ironic condition) and with a picture which provided an incongruent context to the commenting sentence (ironic condition). The inter-trial interval (from the offset of the last word in the commenting sentence to the onset of the first word in the introductory sentence) was randomized between 4000–4200 ms. Identities of the four speakers in the audio files changed trial-by-trial, and the keywords were allocated at different positions of the commenting sentences making the sentences more variable and less predictable.

Both negative (50%) and positive (50%) valence keywords were applied. The commenting ironic sentence with a positive keyword can be defined as ironic criticism and ironic sentence with negative keyword can be defined as ironic praise. However, due to the limited number of trials in each sub-type, we did not analyze possible effects related to different sub-types.

In total, 100 conversations with two contrasting pictures were created. A pilot study aimed to validate the agreement in the congruency rating was conducted before the formal experiment. Seven participants rated one half of the conversations, and six participants rated the other half. These participants were different from the participants of the actual ERP study. Ten conversations were excluded from the stimuli after the pilot study because the accuracy of congruency in each was less than 95%.

Therefore, 90 conversations with two contrasting pictures (90 trials for ironic and non-ironic conditions, respectively) were applied in the EMG and EEG measurement. They were divided into four blocks of 45 trials. Stimulus presentation order was pseudorandom: the ironic and non-ironic stimuli were presented in a random order, but to avoid immediate repetitions, the same conversations were separated with at least 45 other conversations. In order to keep the measurement time reasonable and participants better focused on the task at hand, for the measurement of the behavioral responses participants were presented with a half of the same stimuli they were presented in the EMG and EEG measurement: a half of the participants rated the first two stimulus blocks, and the other half rated the last two stimulus blocks applied in the EMG and EEG measurement. For all the measurements, the presentation of the stimuli was controlled by E-Prime 2.0 software (Psychology Software Tools, Inc., Sharpsburg, PA, USA).

2.4 Facial EMG and EEG data

2.4.1 Measurement of facial EMG and EEG data

For facial EMG responses, data were recorded continuously with two disposable bipolar electrode pairs (Ag/AgCl), and in each pair, there was 0.5 cm distance between the electrodes. Pads with disinfectant were used to clean the skin in contact with the electrodes. The bipolar electrodes were placed on the right side of each participant's face, one over the corrugator supercilii muscle region (above the eyebrow) and another one over the zygomaticus major muscle region (around the cheek).

For EEG recording, a 128-Channel Net (HydroCel Geodesic Sensor Net, Electric Geodesic Inc, USA) was applied. The vertex electrode (Cz) was used as the online reference electrode.

Continuous facial EMG and EEG signals were both recorded simultaneously by a NeurOne system (Bittium Biosignals Ltd, Kuopio, Finland). All signals were recorded using AC mode at a sampling rate of 1000 Hz and were filtered online with a high cut-off of 250 Hz.

2.4.2 Analysis and statistics for facial EMG data

Facial EMG responses were pre-processed and analyzed according to a previous study (Thompson et al., 2016). The facial EMG data were first filtered with a 20–400 Hz band-pass filter (24 dB/octave) and a 50 Hz notch filter in BrainVision Analyzer 2.1 (Brain Products GmbH, Munich, Germany). Then, facial EMG data were segmented into 4400-ms epochs, starting from 400 ms before the onset of keywords. After this, a custom MATLAB script (MATLAB, R2015b, version 8.6.0) was used to further process the data. A rectified facial

EMG segment was omitted if the amplitude was larger than 250 μV in any of the time points. Then, each facial EMG segment was computed into 11 consecutive 400-ms time windows. The facial EMG activity of 400-ms before each keyword was regarded as a baseline for each segment, and for each participant, the facial EMG amplitude was presented as a percentage relative to the baseline.

In order to reduce the levels of the time windows for analysis of variance (ANOVA; Akechi et al., 2013), facial EMG amplitude was averaged over a 1200-ms time window. The percentage of mean amplitude relative to the baseline was calculated for three intervals: 0–1200 ms (Time 1), 1200–2400 ms (Time 2) and 2400–3600 ms (Time 3) after the onset of the keyword. The values for facial EMG were analyzed separately for facial reactivity in the zygomaticus major muscle and in the corrugator supercilii muscle by using three-way repeated measures ANOVAs with Stimulus type (non-ironic vs. ironic) and Time window (1 vs. 2 vs. 3) as within-subject factors and Group (control vs. dysphoric) as a between-subject factor.

2.4.3 Analysis and statistics for EEG data

EEG data were analyzed by using Brain Vision Analyzer 2.1 (Brain Products GmbH, Munich, Germany). A new reference was calculated offline as an average over all channels. The Gratton and Coles method (Gratton et al., 1983) was applied to detect and correct the interference of eye-blinks, and Channel 8, which is above the midpoint of the right eye, was chosen as the vertical electro-oculogram (VEOG) channel. Channels with an extensive amount of noise were interpolated using a spherical spline model. The EEG signal was filtered with a low cut-off of 0.1 Hz and a high cut-off of 20 Hz, accompanied by a 24 dB/octave roll-off. A 50 Hz notch filter was also applied. After this, EEG data were extracted into 1100-ms long segments relative to the onset of keywords, starting from 100 ms before the presentation of keywords. To exclude noisy segments, gradient criterion and max–min amplitude criterion were both applied. Specifically, the maximum allowed difference in amplitude between two consecutive time points was 75 μV, and the maximum allowed voltage was −150 μV and 150 μV at an interval of 200 ms for all the channels. Segments exceeding the specified values were omitted from the averages. The mean voltage during a period of 100 ms prior to the onset of the keyword was used as a baseline for each segment. Data were averaged separately for the two stimulus types (ironic vs. non-ironic) and for each participant.

For the ERP data, two responses were analyzed: N400-like activity and P600. Similarly to a previous study (Filik et al., 2014), we use here the term N400-like instead of N400, because there was no clear peak for the observed response unlike the classical N400 (e.g., Kutas and Federmeier, 2000; Kutas and Hillyard, 1980; Van Petten and Luka, 2006).

The time windows for statistical analysis of the N400-like activity and P600 were defined a priori based on previous studies (N400-like activity: 300–500 ms after the onset of the target word, Kutas & Federmeier, 2011; P600: 500–800 ms after the onset of the target word, Bornkessel-Schlesewsky & Schlesewsky, 2008).

The selection of the electrodes for the analysis was based on cluster-based permutation statistics (Maris and Oostenveld, 2007). This data-driven method was used as a tool to restrict the number of electrodes for the subsequent analyses conducted with repeated measures of ANOVAs (see also, e.g., Hämäläinen et al., 2018; Strömmer et al., 2017). This procedure is described in Supplementary materials. According to the results of the permutation statistics, the N400-like response was analyzed from one region of interest (ROI) (the electrodes 5, 6, 7, 12, 31, 80, 106 and 112), and the P600 response was analyzed from two ROIs (Left ROI: the electrodes 50, 51, 58 and 59; Right ROI: the electrodes 91, 96, 97 and 102). Figure 3 and Supplementary Figure 1 depict the electrode locations.

Mean amplitude values of the N400-like activity and P600 from the pre-defined time windows and electrodes selected with cluster-based permutation tests were applied in separate statistical analyses for each component. For the N400-like activity, a two-way repeated measures of ANOVA with Stimulus type (non-ironic vs. ironic) as a within-subject factor and Group (control vs. dysphoric) as a between-subject factor was conducted. For P600 responses, a three-way repeated measures ANOVA with within-subject variables Stimulus type (non-ironic vs. ironic) and ROI (left vs. right), and a between-subject variable Group (control vs. dysphoric) was applied. For P600, whenever a three-way interaction effect (Stimulus type × ROI × Group) was revealed, two follow-up two-way repeated measures ANOVAs were conducted to investigate the interaction effect: one separately for the two stimulus types (ROI × Group) and the other separately for the two groups (Stimulus type × ROI).

2.5 Behavioral responses

2.5.1 Measurement of behavioral responses

During the behavioral testing, participants were instructed to keep both hands on the keyboard (standard Finnish QWERTY keyboard), and press buttons with their index fingers. Response buttons (C, V, B, N, M in the keyboard) were labelled and covered by stickers with numbers (-2, -1, 0, 1, 2). Participants were first asked to evaluate as quickly as possible whether the picture and the spoken commenting sentence after it were congruent or incongruent. Then participants were asked to evaluate the valence of the conversations by using a scaling of -2 to 2 by pressing buttons that were labelled: hyvin ikävä (very unpleasant) = -2; ikävä (somewhat unpleasant) = -1; siltä väliltä (between unpleasant and funny) = 0; jokseenkin hauska (somewhat funny) = 1; hyvin hauska (very funny) = 2. The valence rating was thus designed to measure especially perceived funniness of the stimuli and the opposite negative aspect. “Ikävä”, the word chosen to be the opposite (the other end of the rating scale) for funny can imply that there is a certain dislikeable aspect. Even if the promptness of the responding was requested, there was no time limit for responding. Before the actual evaluation task, the participants practiced with six rehearsal stimuli, and the experimenter confirmed that participants understood the task correctly.

2.5.1 Analysis and statistics for behavioral data

As the accuracy of the congruency detection is considered as binary and categorical data (correct: 1, incorrect: 0) in the present study, a multilevel model (mixed model) was applied (Jaeger, 2008), using Mplus software (Version 8). The model included Accuracy by Stimulus type in each trial at the within-level, and Group at the between-level (with a random slope). The full-information maximum likelihood method (MLR estimation in Mplus) was conducted to estimate the parameters. This model was aimed at analyzing the interaction between the Stimulus type (non-ironic vs. ironic) and the group (control vs. dysphoric) variables. The results of detection accuracy are reported as the percentage of correct responses.

The response times for congruency detection were calculated for trials with correct responses from the onset of the keywords. In the given example (Figure 1), “systemaattista” (“systematic”) was defined as the keyword of the commenting sentence. For each participant, response times above 2.5 standard deviations from the mean response time were excluded from the analyses. As a result of the trimming procedures, approximately 2.7% of all trials for all participants were lost for the analysis of response times. Each participant had more than 32 trials for each stimulus type (ironic or non-ironic) for the reaction time measure.

Each participant's mean values for response time of the congruency detection and valence rating of conversations were calculated for both stimulus types (non-ironic and ironic) and were analyzed by conducting a two-way repeated measures of ANOVA, with Stimulus type (non-ironic vs. ironic) as a within-subject factor and with Group (control vs. dysphoric) as a between-subject factor, respectively.

2.6 General information on statistics

For all ANOVA models, whenever the sphericity assumption was violated, Greenhouse-Geisser correction was applied. The *p*-values for ANOVA results are reported based on the Greenhouse-Geisser correction, but the degrees of freedom are reported as uncorrected. The significance level for all statistics was $p < .050$, but marginally significant ($p \leq .080$) interaction effects were also further studied. Significant interactions found in the ANOVAs or follow-up ANOVAs (for P600 responses) were investigated by using independent samples *t*-tests for between-subject comparisons or two-tailed paired t-tests for within-subject comparisons. Both types of t-tests were applied with a bootstrapping method using 1000 permutations (Good, 2005) as implemented in IBM SPSS Statistics 24.0 (Armonk, NY: IBM corporated). Partial eta-squared η_p^2 and Cohen's d are reported for the estimates of effect size for ANOVAs and t-tests, respectively. Cohen's d was computed using pooled standard deviations (Cohen, 1988).

Pearson's correlation coefficients (two-tailed, computed at subject level) were applied to examine the relationships between the comprehension-related variables (the N400-like activity and the P600, and accuracy of congruency detection, respectively), and between the emotion-related variables (facial EMG corrugator and zygomaticus, and valence ratings of conversation, respectively). The correlations were calculated for ironic and non-ironic stimulus conditions separately. Multiple correlations were controlled by applying false discovery rate (Benjamini and Yekutieli, 2001) at 0.050.

For exploratory purposes, the Humor Styles Questionnaire scores were applied as a covariate to original analyses conducted for valence ratings and facial EMG activity (repeated measures of analysis of covariance, ANCOVA). These measures of affective aspects of irony could be expected to be influenced by individual's humor style. Here we used the Aggressive humor style from the Humor Styles Questionnaire because irony is close to sarcasm (Kreuz and Glucksberg, 1989), which can be categorized as aggressive humor (Martin et al., 2003). Cognitive test scores were also applied as a covariate in P600 analysis, for which we found a group difference. It can be assumed that cognitive deficits can influence irony comprehension. In order to reduce variables for analysis of covariance, principal component analysis was conducted to extract factors of the cognitive tests (see, Supplementary materials). Three cognitive factors (executive function, list memory, and semantic processing) were extracted and applied separately as covariates in a two-way repeated measures ANCOVA of P600 responses. In order to explore the effect of current medication status on the behavioral responses, ERPs, and facial EMG reactivities, additional repeated measures ANCOVA analyses were also applied with Medication (medicated vs. non-medicated) as a covariate for the dysphoric group. Since there were no interaction effects with the covariates found in all the above mentioned ANCOVA analysis, we report in Results only the changes the covariates caused to the original group effects, that is, when the covariates revealed or concealed a main effect of Group or an interaction effect with it.

In addition to ANOVA/ANCOVA analyses, effect of depressive symptoms on responses were investigated with simple linear regression analyses. A simple linear regression was calculated for the whole sample with the amount of depressive symptoms (BDI-II scores) as a predictor of ERPs, facial EMG amplitudes, and behavioral evaluations separately. Results of these analyses are reported in Supplementary materials.

3 Results

3.1 Behavioral results

The results of the accuracy and the response time of the congruency detection, as well as the valence ratings of ironic and non-ironic conversations, are presented in Figure 2.

3.1.1 Accuracy of congruency detection

The multilevel model analysis revealed a significant main effect of Stimulus type, $p = .007$. The detection accuracy of the incongruent stimuli ($M = 97.78\%$, $SD = 0.14$) was slightly higher than that of the congruent stimuli ($M = 96.2\%$, $SD = 0.19$). Neither the interaction effect of Stimulus type × Group nor the main effect of Group was significant (all p-values > .172).

3.1.2 Response time of congruency detection

The two-way repeated measures ANOVA showed no significant effects, the main effect of Stimulus type being closest to significant, $F(1,36) = 3.869$, $p = .057$, $\eta_p^2 = .097$. The response time was descriptively longer for the ironic stimuli ($M = 2041.99$ ms, $SD = 453.79$) than for the non-ironic stimuli ($M = 1992.62$ ms, $SD = 419.14$). No interaction effect of Stimulus type × Group and main effect of Group were observed (all p-values > .116).

3.1.3 Valence ratings of conversations

For the valence ratings of ironic and non-ironic conversations, repeated measures ANOVA revealed a significant main effect of Stimulus type, $F(1,36) = 49.404$, $p < .001$, $\eta_p^2 = .578$. Ironic conversations ($M = 0.27$, $SD = 0.42$) were rated as funnier than non-ironic conversations ($M = -0.11$, $SD = 0.31$). The main effect of group was also significant, $F(1,36) = 5.098$, $p = .030$, $\eta_p^2 = .124$. Valence ratings were more negative in the dysphoric group ($M = -0.05$, $SD = 0.42$) than in the control group ($M = 0.19$, $SD = 0.19$) over the stimulus types. Interaction effect of Stimulus type × Group was non-significant, $p = .112$.

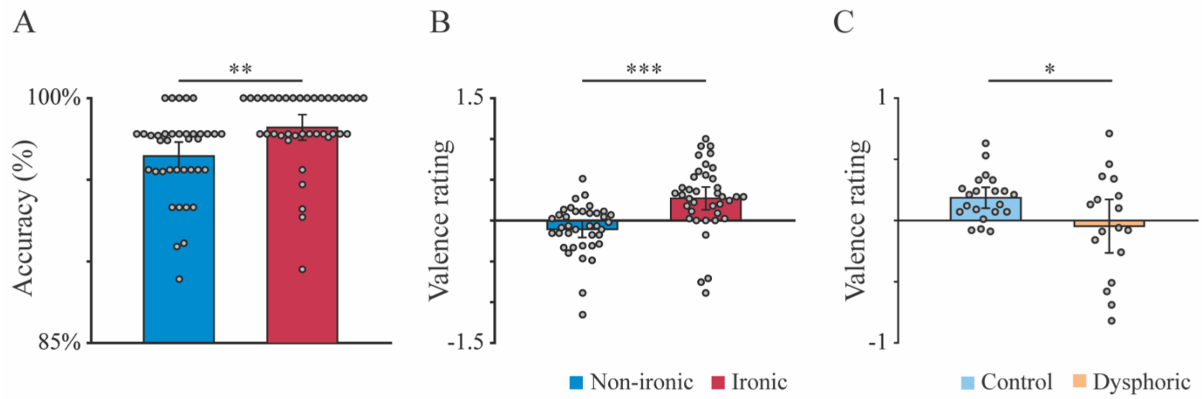


Figure 2. Behavioral responses on the second day's measurement. (A) Mean accuracy of congruency detection and 95% confidence intervals to non-ironic (blue) and ironic (red) conversations. Values of individual responses are presented as a scatterplot. ***p* < .01. (B) Mean values of valence ratings and 95% confidence intervals to non-ironic (blue) and ironic (red) conversations. Individual responses are presented as a scatterplot. ****p* < .001. (C) Mean values of valence ratings and 95% confidence intervals in control (light blue) and dysphoric (light orange) groups (averaged over non-ironic and ironic). Individual responses are presented as a scatterplot. **p* < .05. The scale for valence ratings ranged from -2 (very unpleasant) to 2 (very funny), with 0 meaning neutral. Please note that the scale is different for B and C.

3.2 Facial EMG responses results

Results of the repeated measures ANOVAs for facial EMG corrugator and zygomaticus activity are reported in Table 2. The percentages of averaged corrugator and zygomaticus activity relative to their baseline activity for each condition are shown in Figure 3.

3.2.1 Facial EMG corrugator activity

For facial EMG corrugator activity, three-way repeated measures ANOVA with within-subject variables Stimulus type (non-ironic vs. ironic) and Time window (1 vs. 2 vs. 3), and between-subject variable Group (control vs. dysphoric) revealed that the main effect of Time window was significant, $p = .009$. The corrugator activity was smaller between 0–1200 ms ($M = 100.4$ %, $SD = .022$), compared with the activity in the period of 1200–2400 ms ($M = 101.3\%$, $SD = .032$), $t(37) = 2.599$, $p = .013$, d = 0.327, and with the activity during 2400–3600 ms ($M = 102\%$, $SD = .040$), $t(37) = 2.738$, $p = .009$, d = 0.495. The difference between the activity in the period of 1200–2400 ms and 2400–3600 ms was not found, $p = .051$. Other main effects or interactions were not found for corrugator responses (all p-values > .312).

3.2.2 Facial EMG zygomaticus activity

For facial EMG zygomaticus activity, three-way repeated measures ANOVA showed that the main effect of Time window was significant, $p = .002$. The zygomaticus activity was larger in the time window of 1200–2400 ms ($M = 105.8\%$, $SD = .115$) and 2400–3600 ms ($M = 105.3\%$, $SD = .098$), compared with the activity between 0–1200 ms ($M = 101.3\%$, $SD = .033$), $t(37) = 3.082$, $p = .004$, d = 0.532; $t(37) = 2.739$, $p = .009$, d = 0.547, respectively. There was no significant difference between the activity in the period of 1200–2400 ms and 2400–3600 ms, $p = .674$. The main effect of Stimulus type was significant, $p = .039$. The facial EMG zygomaticus amplitude was larger after the ironic stimuli ($M = 105.3\%$, SD

= .105) than after the non-ironic stimuli (*M* = 103%, *SD* = .051). There were neither group differences nor interactions found in zygomaticus responses, all *p*-values > .091.

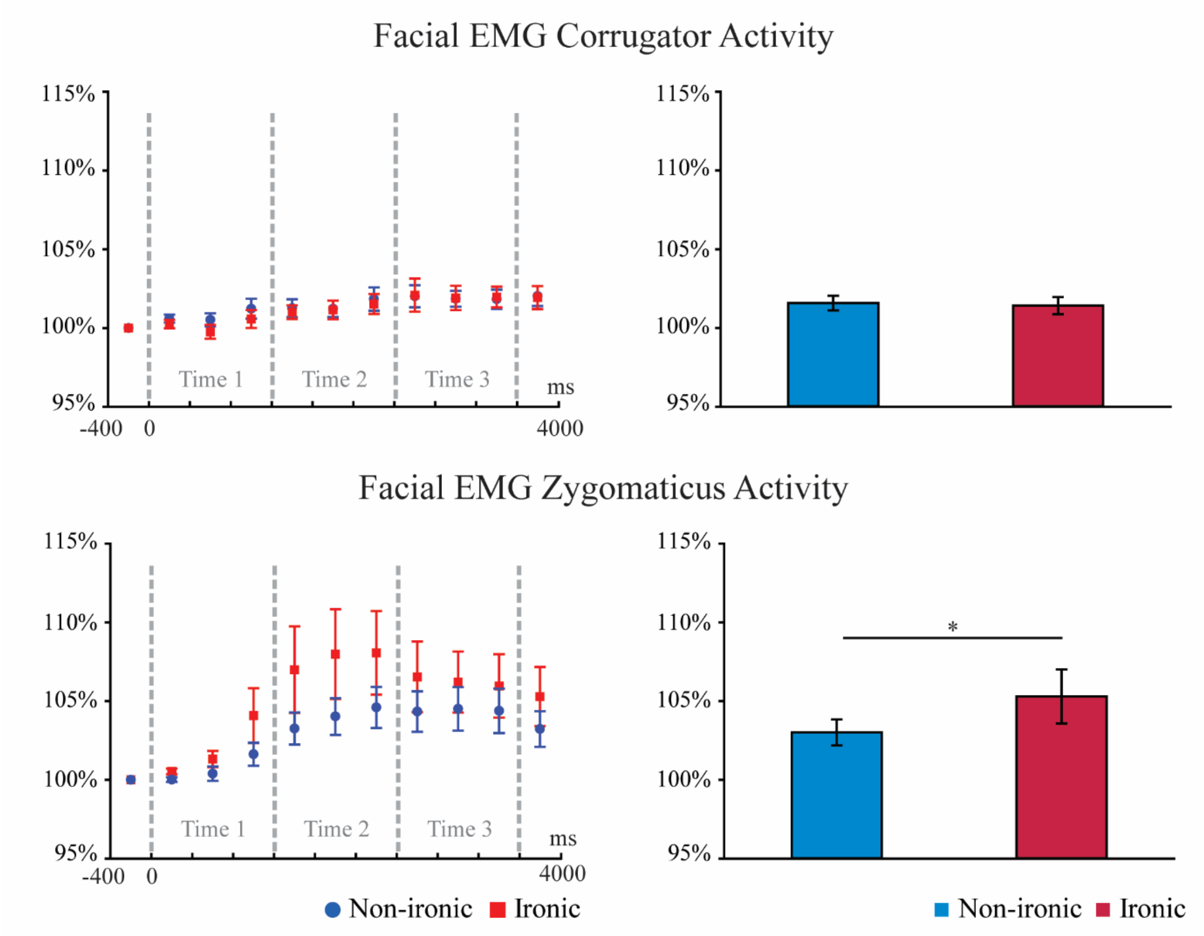


Figure 3. The grand-averages of the mean EMG amplitude percentage relative to the baseline for the corrugator (upper) and for the zygomaticus (lower) to non-ironic (blue) and ironic (red) stimulus. Left: Average percentage values and standard error of mean (SE) are presented for descriptive purposes in 400-ms segments, and the grey dotted lines show the segments applied in the ANOVA (Time 1: 0–1200 ms; Time 2: 1200–2400 ms; Time 3: 2400–3600 ms). Right: The grand-averages of the mean values and SD for non-ironic and ironic responses averaged over the time segments reflect the main effect of Stimulus type (*p < .05) for the zygomaticus. The responses are averaged over the two groups, since there were no group differences (see Supplementary Figure 3 for the responses in each group).

3.3 ERPs results

ERPs results are presented in Figure 4 (N400-like activity) and Figure 5 (P600). Grand-averaged responses showed larger responses for the ironic than the non-ironic stimuli at 300–

500 ms and 500–800 ms after the onset of the keyword reflecting N400-like activity and P600, respectively. Results of repeated measures ANOVAs for N400-like activity and P600 activity are reported in Table 3.

3.3.1 N400-like activity

The analysis of N400-like activity revealed that the main effect of Stimulus type was significant, $p = .007$. The amplitude of the N400-like response was larger (toward negative polarity) for ironic ($M = -0.73$ µV, $SD = 0.69$) than non-ironic ($M = -0.37$ µV, $SD = 0.75$) stimuli. There were no other main or interaction effects, all p-values > .681.

3.3.2 P600

The ANOVA analysis of the P600 amplitude showed that the main effect of Stimulus type was significant, $p < .001$. Ironic conversations elicited larger amplitude toward positive polarity ($M = 1.15$ µV, $SD = 0.83$) than non-ironic conversations ($M = 0.72$ µV, $SD = 0.62$). There was also a three-way interaction effect of Stimulus type × ROI × Group, $p = .034$. No other main or interaction effects were observed, all p-values > .150.

Follow-up ANOVAs investigating the three-way interaction of Stimulus type × ROI × Group were conducted. The amplitude values for each ROI and each stimulus type in the control and the dysphoric groups for P600 are reported in Table 4.

First, follow-up two-way repeated measures ANOVAs with within-subjects variables Stimulus type (non-ironic vs. ironic) and ROI (left vs. right) were conducted separately for the control and the dysphoric group. In the control group, the analysis showed a significant main effect of Stimulus type, $F(1,20) = 5.771$, $p = .026$, $\eta_p^2 = .224$. The amplitude of the P600 response was larger for ironic ($M = 1.13$ µV, $SD = 0.97$) than non-ironic ($M = 0.80$ µV, $SD =$

0.65) stimuli. There were no other main or interaction effects in the control group, all *p*-values > .257. In the dysphoric group, a main effect of Stimulus type was observed, $F(1,16) = 22.864$, $p < .001$, $\eta_p^2 = .588$. The amplitude of the P600 response was larger for ironic (M = 1.17 µV, SD = 0.65) than non-ironic (M = 0.61µV, SD = 0.57) stimuli. The main effect of ROI was not significant, $p > .297$. A marginally significant Stimulus type × ROI interaction effect, $F(1,16) = 3.912$, $p = .065$, $\eta_p^2 = .196$, was found in the dysphoric group. Paired *t*-tests revealed that, in the right ROI, P600 amplitude was larger to ironic than to non-ironic stimuli in the dysphoric group, $t(16) = 4.453$, $p < .001$, d = 0.832. In the left ROI, there was no significant difference between the responses to ironic and non-ironic stimuli, $p = .139$.

Second, follow-up two-way repeated measures ANOVAs with a within-subjects variable ROI (left vs. right) and a between-subjects variable Group (control vs. dysphoric) were conducted for the ironic and non-ironic stimuli separately. For the non-ironic stimuli, there were neither a main effect of ROI, a main effect of Group nor their interaction effect, all *p*-values > .340. For the ironic stimuli, the main effect of ROI and the main effect of Group were non-significant, both *p*-values > .776. There was a significant interaction effect of ROI × Group, $F(1,36) = 6.978$, $p = .012$, $\eta_p^2 = .162$, however. Post hoc tests based on independent samples t-tests comparing the groups in P600 amplitude to ironic stimuli separately at the left and at the right ROI indicated no group differences, both *p*-values > .135. Paired samples t-tests exploring the amplitude difference to ironic stimuli between the left and the right ROIs within each group were implemented. The analysis showed that the P600 amplitude to ironic stimuli was larger on the right ROI than on the left ROI in the dysphoric group, $t(16) = 2.399$, $p = .029$, d = 0.674, but there was no difference in the control group, $p = .132$.

For exploratory purposes, the measures of cognitive skills, which could be relevant on the cognitive aspect of irony processing, were applied in the ANCOVA of P600 responses.

Values of three factors of the cognitive tests (executive function, list memory, semantic processing) were added as covariates independently in the ANCOVA model for P600 responses. Only list memory as a covariate in the ANCOVA changed the original results: the three-way interaction of Stimulus type × ROI × Group was not anymore significant, $p = .091$.

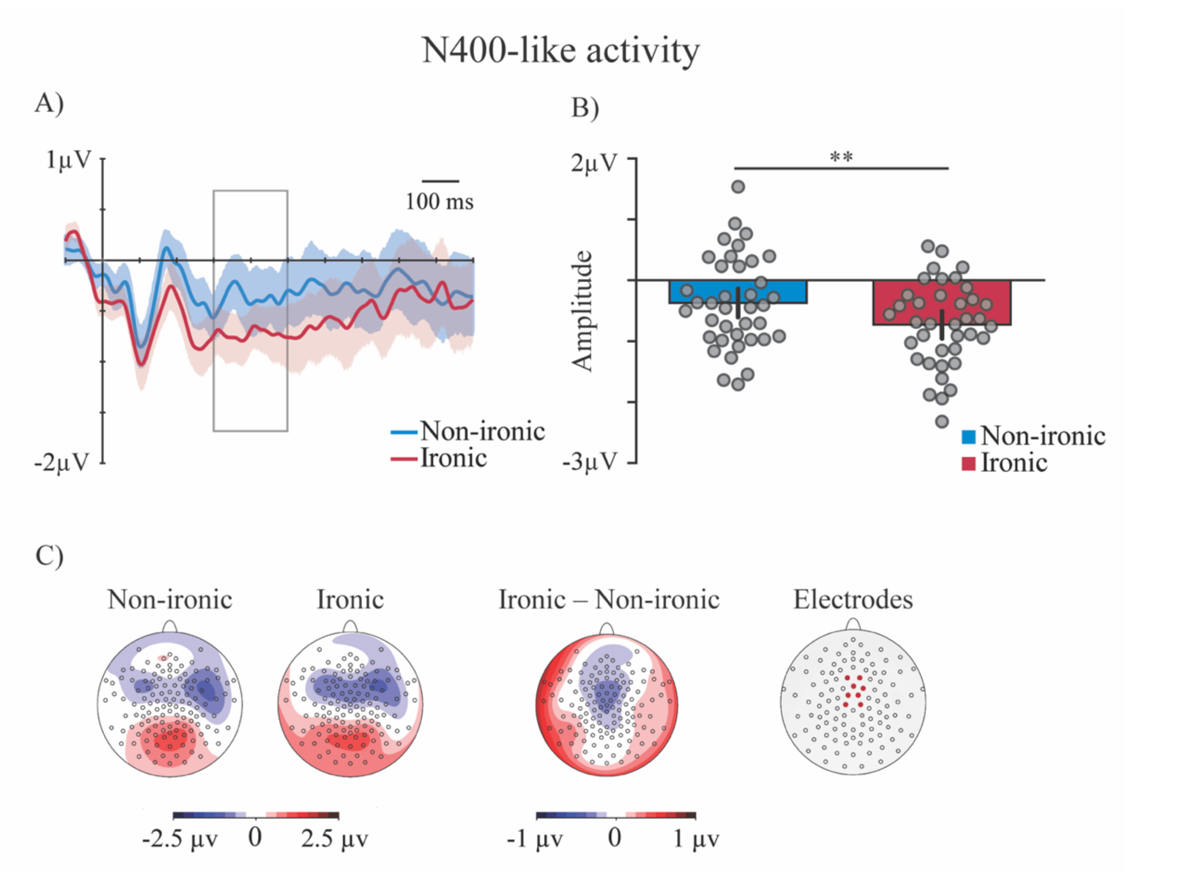


Figure 4. Grand-averaged waveforms, mean amplitude values, topographical maps, and selections of electrodes for N400-like activity. A) Grand-averaged waveforms to non-ironic and ironic stimuli are presented. The blue and red shadows in the waveforms represent 95% confidence intervals. The grey rectangle shows the time window applied in the analysis for N400-like activity (mean amplitude values between 300–500 ms). B) Mean amplitudes for N400-like activity to non-ironic (blue) and ironic (red) stimulus are illustrated. The error bars represent 95% confidence intervals. $^{**}p < .01$, a main effect of Stimulus type. The scatterplots overlaid with the histograms represent individual participants' amplitudes in each condition. C) Grand-averaged topographical maps for non-ironic and ironic stimuli, and a difference between the two stimulus types in topographies and selections of electrodes (marked with red) for N400-like activity are shown. The responses are averaged over the two groups, since there were no group differences (see Supplementary Figure 2 for the responses in each group).

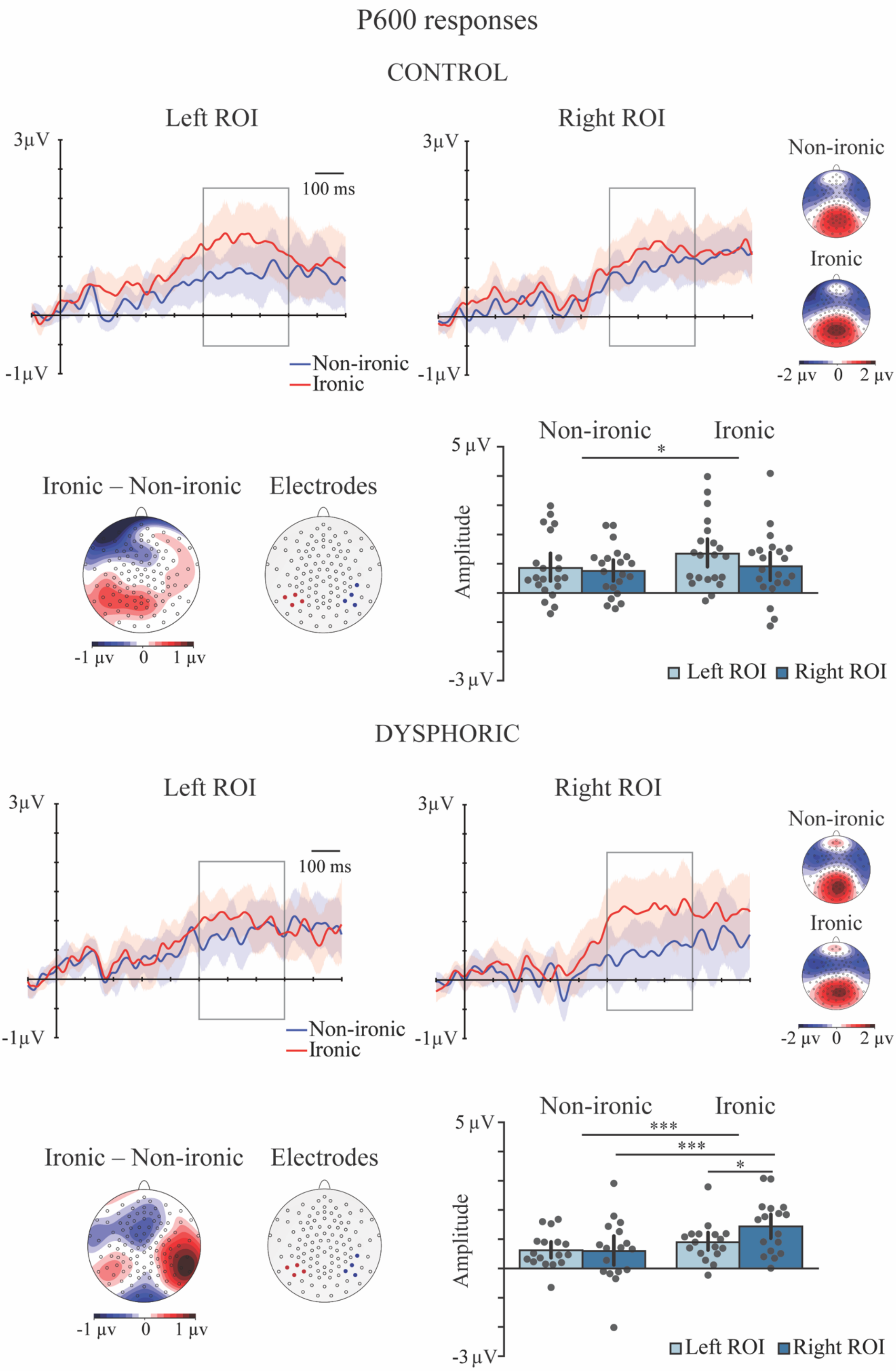

P600 responses
CONTROL
Left ROI
Right ROI
3µV
-1µV
100 ms
Non-ironic
Ironic
-2 µv 0 2 µv
Ironic – Non-ironic
Electrodes
-1 µv 0 1 µv
5 µV
-3 µV
Amplitude
*
Left ROI
Right ROI
DYSPHORIC
***
***
*

Figure 5. Grand-averaged ERP waveforms, mean amplitude values, topographical maps, and selections of electrodes for P600 on left and right ROIs separately for the control and the dysphoric group. The blue and red shadows in the waveforms represent 95% confidence intervals. The grey rectangles in waveform figures show the time window applied in the analysis for P600 (mean amplitude values between 500–800 ms). The red dots and the blue dots represent the electrode cluster on the left and right ROI, respectively. The histograms represent mean amplitudes of P600 to non-ironic and ironic stimulus on the left and right ROIs. The error bars represent 95% confidence intervals (*$p < .05$; ***$p < .001$, post hoc paired t-tests investigating Stimulus type × ROI × Group interaction). The scatterplots overlaid with the histograms represent individual participants' amplitudes in each condition.

3.4 Correlations

There were no correlations between ERPs (the N400-like activity and the P600) and behavioral accuracy of congruency detection for either ironic or non-ironic stimulus condition, all p-values > .198. The correlations between facial EMG (corrugator and zygomaticus activity) and valence ratings of conversations in response to ironic and non-ironic stimulus condition were not found either, all p-values > .101.

4 Discussion

We investigated irony comprehension and emotional reactions related to irony by measuring behavioral responses, facial EMG, and ERPs. The effect of depressiveness was investigated by comparing responses between the dysphoric group and the non-dysphoric control group.

The results showed that the participants in both groups found the ironic conversations to be funnier than the non-ironic conversations. However, overall, the dysphoric group rated both types of conversations as being less funny than the control group. Facial EMG activity in the zygomaticus major, which is an indication of smiling, was also greater in response to the ironic stimuli than the non-ironic stimuli in both groups. As expected, irony processing was reflected in the ERPs as enlarged amplitudes of the N400-like activity and the P600, indexing cognitive processing of irony. A difference in the P600 responses was found between the dysphoric group and the control group. Next, the results are discussed in details.

Behavioral responses indicated that the accuracy in the congruence detections was slightly better for the ironic stimuli than for the non-ironic stimuli for both groups, and the response time was descriptively longer, although the latter effect was not statistically significant. The descriptively longer response time may reflect higher cognitive effort and cognitive complexity in processing ironic stimuli. Consistent with previous studies that found no difference in the detection of semantic violations between the depressed and control participants (Deldin et al., 2006; Klumpp et al., 2010), we found no differences in the accuracy of the congruence detection between the dysphoric and control groups. Moreover, no difference in response time was found between the two groups.

Both groups rated the ironic conversations funnier than the non-ironic ones, which is compatible with the findings reported in previous studies that compared the ratings of ironic and literal sentences (Calmus and Caillies, 2014; for ironic criticism see, Dews et al., 1995) and the degree of perceived humor in ironic and literal stories in healthy participants (Akimoto et al., 2014). In comparing the two groups, the dysphoric group’s valence ratings (unpleasant vs. funny) were lower than the control group’s for both ironic and non-ironic stimuli. This finding is in line with previous studies that reported lower reactivity to different kinds of emotional stimuli in depression (Bylsma et al., 2008; Moran et al., 2012; Rottenberg et al., 2005). Notably, the difference in valence ratings between the two groups may not be explained by differences in the groups’ preference of humor styles, because the groups did not differ in any of the measured humor styles (for the results on humor style, see the Supplementary materials).

Consistent with the valence ratings, the activity of the zygomaticus muscle, which is known to indicate smiling (Van Boxtel, 2010), showed a relatively larger response to the ironic stimuli than to the non-ironic stimuli. This most probably indicated positive emotions related to irony at the whole sample level. A previous study using the facial EMG method have also demonstrated positive emotions to ironic criticism (Thompson et al., 2016).

There were no differences in the facial EMG responses between the dysphoric group and the control group. This finding was surprising, taking into account the well-documented alterations in emotional information processing in depression (e.g., Gotlib and Joormann, 2010; Leppänen, 2006) and the reduced tendency to react with exhilaration to funny stimuli (Uekermann et al., 2008; Falkenberg et. al., 2011). It is possible that the two groups did not differ in their facial EMG responses because similar to the dysphoric group, the ironic stimuli were not found to be very funny in the control group, although they were found to be relatively more funny than the non-ironic conversations at the whole sample level. Another possible reason for the lack of group difference could be the large variance we observed especially in the valence ratings and zygomaticus facial EMG for the dysphoric group. The variance can be expected within dysphoric participants due to the heterogeneity of depression related to different symptom profiles (e.g., more or less vegetative vs. affective symptoms), differences in amount of symptoms or diagnosis type (e.g. atypical depression vs. melancholic depression).

Regarding the ERP data, modulation of both the N400-like activity and P600 related to irony were observed. This finding is compatible with the traditional standard pragmatic view of irony comprehension (e.g., Grice, 1975), which states that, for irony comprehension, detecting and distinguishing the incongruence in a semantic context (reflected by the N400-like activity) and inferring and interpreting the speakers' implied meaning (reflected by the

P600) are required to understand ironic statements. Because we did not evaluate the participants' familiarity with the words in the commenting sentences, our study was unable to test the graded salience hypothesis (e.g., Giora and Fein, 1999; Giora, 2003). However, since modulation of both the N400-like activity and P600 in response to irony was found, our results support the two-stage model of irony processing (Giora, 2003; Grice, 1975) rather than the one-stage model (the direct access view, e.g., Gibbs, 2002).

The N400-like activity was more negative in polarity in response to the ironic stimuli than in response to the non-ironic stimuli. Unlike traditional N400 (e.g., Kutas and Federmeier, 2011), the response did not seem to have a clear peak (see also, Filik et al., 2014) and was slightly frontally-distributed. Here, we applied naturally spoken sentences in which the position of the keyword varied between the sentences. This may have affected the morphology and the topography of the N400 response (Kutas and Federmeier, 2011), which is usually measured in a condition where the keyword is in the end of the sentence (e.g., Filik et al., 2014; Regel et al., 2011, 2014).

The observed N400-like activity to irony is in line with previous findings of N400 elicited to irony (Cornejol et al., 2007; Filik et al., 2014; Caillies et al., 2019), humor (Coulson and Kutas, 2001; Coulson and Wu, 2005), and other nonliteral language (metaphor: Coulson and Van Petten, 2002) in showing enhanced response amplitude to nonliteral stimuli (here ironic) in comparison to literal stimuli (here non-ironic). This finding may suggest that the participants had difficulty processing the meaning of the commenting sentence in the context of the opposing contextual picture. This difficulty may stem from the holistic strategy the participants applied during comprehension (Cornejol et al., 2007) or the stimuli, which were unfamiliar to the participants, because a previous study only found N400-like modulation for unfamiliar ironic utterances (Filik et al., 2014). Some studies have not found N400 modulation for irony, however (Balconi and Amenta, 2008; Regel et al., 2011, 2014). Regel et al. (2011, 2014) reported that the lack of N400 modulation to ironic stimuli means that the participants were capable of comprehending irony based on the supportive context, with no difficulty in semantic interpretation. In our study, even though we did not measure the difficulty of semantic interpretation directly with behavioral tests, the results of congruence detection showed that the participants were very accurate in categorizing both congruent (i.e., non-ironic) and incongruent (i.e., ironic) picture-sentence pairs, and we found the amplitude of the N400 was still modulated by the irony. However, the results regarding the N400 effect should be interpreted with caution, because the effect size of the irony effect is small and as mentioned earlier, it is different in morphology from traditional N400 effects.

There was no difference in the N400-like activity in response to ironic stimuli between the dysphoric and control groups. This result suggests that semantic processing related to irony is not altered in dysphoria. This finding is consistent with previous results showing that patients with mood disorders have normal semantic processing in a passive sentence-viewing task, as indexed by N400 (Deldin et al., 2006).

As expected, P600, which displayed a centro-parietal distribution, was also modulated by irony in our study. The amplitude of P600 was more positive for the ironic punchlines than the non-ironic punchlines; this result has also been reported in previous studies on irony comprehension (Baptista et al., 2018; Filik et al., 2014; Regel et al., 2011, 2014; Caillies et al., 2019). Visual observation of the grand-averaged difference topographies (ironic minus non-ironic) for the P600 responses in Filik et al. (2014) and those in the healthy controls in our study shows a remarkable similarity: they both show irony-related activity in the central and in the left parietal electrode sites. Moreover, the results in our study and in previous studies are similar in that no differences were found in irony modulation between the left and right ROIs in the healthy controls (Filik et al., 2014; Regel et al., 2011, 2014).

In the present study, we defined the EEG-electrode sites for the analysis based on a data-driven method, and we found two ROIs for the P600 modulation: one in the left parietal electrode cluster and the other in the right parietal electrode cluster. Previous ERP studies investigating irony comprehension have selected the electrodes for the analysis either based on previous literature (Caillies et al., 2019) or have applied several fixed ROIs for statistical analysis (Filik et al., 2014; Regel et al., 2011, 2014). The studies that had many different fixed ROIs in analysis (i.e., anterior vs. posterior vs. central vs. left vs. right ROIs) have found maximum irony-related P600 activity over the posterior electrode sites (left mastoid reference: Regel et al., 2011; average reference: Filik et al., 2014) or over the right central and the left and right parietal electrode sites (left mastoid reference: Regel et al., 2014). Thus, it seems to be that in the present study, in which a data-driven method was used to select the electrode sites, we found the irony modulation at approximately same area as those of studies that used fixed ROIs (Filik et al., 2014; Regel et al., 2011).

Here, the participants did not engage in any task related to the stimuli during the EEG/EMG-measurement, and the behavioral evaluations were conducted on separate days. Therefore, it is unlikely that the present study’s finding of enhanced P600 in relation to irony is elicited by the requirements of the comprehension tasks or categorization tasks, which were possible reasons for the P600 modulation in studies by Regel et al. (2011) and Filik et al. (2014), respectively. Regel et al. (2011) also suggested that an increased P600 amplitude to irony may reflect the processing of emotional information expressed by ironic utterances. Partly supporting this idea, a recent study found that greater P600 modulation was observed in relation to ironic criticism than ironic praise (Caillies et al., 2019). In the present study, we had positive keywords in half of the sentences and negative keywords in another half of the sentences. Due to the limited number of trials in each sub-type (ironic criticism or ironic praise), we were unable to analyze the possible effects related to the different sub-types. However, we found no correlations between the amplitude of P600 and valence ratings of the stimuli. Thus, it seems an unlikely explanation that, in the present study, the modulation of P600 is due to the processing of emotional information. Consequently, we considered that the larger amplitude of P600 likely reflects the greater inferential effort required for the resolution of the ironic punchlines than for the non-ironic punchlines arising from the conflict between the meaning of the keyword and the contextual picture in the irony trials.

A difference in P600 was also found between the dysphoric and control groups. In the dysphoric group, irony modulated the amplitude more in the right ROI than the left ROI; in the control group, irony modulation was found equally in both ROIs. Underlying cause of the different hemispheric balance between the dysphoric and control groups and its functional significance is unknown and needs further investigations. However, Kalatzis et al. (2004) applied a machine learning technique to classify the control and depressed individuals based on P600 responses to numbers during a working memory test, and they found that the discrimination accuracy to distinguish depressed individuals from controls was higher using the electrodes at the right hemisphere in comparison to the electrodes at the left hemisphere. The authors suggested that this result could be related to a right hemispheric dysfunction in depression (Kalatzis et al., 2004), but this assumption requires further investigations.

In addition to the analysis investigating group differences categorically, we also calculated simple linear regressions for P600 (separately for the left and right ROI), facial EMG activities, and valence ratings with BDI-II scores as a predictor. However, the regression analyses did not show any significant effects (see Supplementary materials). One reason for this result could be that self-assessment questionnaires, such as the BDI-II, are not the most accurate measures of depressive symptoms because some depressed individuals do not have a clear awareness of their symptoms; thus, they can inaccurately estimate their symptoms, which in turn can lead to non-significant linear regressions between amount of symptoms and responses to irony.

In the present study, the cognitive test results showed that there were no differences between the dysphoric and control groups in terms of memory, executive functions, or semantic processing (see Supplementary materials). This indicates that, in our sample, cognitive abilities were well-preserved in the dysphoric group. It is possible that the ERPs that reflect the cognitive aspect of irony processing (N400-like activity and P600) could better show depression-related alterations in a sample where alterations in cognition exist. In previous studies, cognitive dysfunction has been mostly associated with recurrent depression (e.g., Fossati et al., 2004; Talarowska et al., 2015). In our sample, the dysphoric participants were young adults, and only one participant reported being diagnosed with recurrent depression. When applying values of the factor list memory as a covariate in the analysis of covariance (ANCOVA), the original results were changed for P600: the three-way interaction of Stimulus type × ROI × Group was no longer observed. This suggests that the difference between the groups is, at least to some extent, driven by differences in memory functions. However, this interpretation needs to be considered cautiously because there were no interactions between the factor list memory and the other variables, and there was no group difference in the list memory.

Several limitations in the present study are worth noting. The dysphoric participants self-reported their diagnostic status; not all of them had been recently diagnosed with depression. However, BDI-II scores were used to measure the depressive symptoms at the time of the measurement, and all the participants in the dysphoric group had scores of 14 or more, reflecting at least mild depression (Beck et al., 1996). Still, our results may not be generalized to clinical depression. Moreover, there were more female participants than male participants in both the dysphoric and the control groups. Therefore, our results cannot unconditionally be generalized to both genders. However, the proportion of gender distribution in the dysphoric and control groups was similar. Furthermore, ten dysphoric participants were taking antidepressant medication while participating in the study, but analyses using the medication status as a covariate did not reveal any interactions between the medication status and the dependent variables. It is also notable that our stimuli included both ironic praise and ironic criticism, and as mentioned before, we were not able to analyze the effects separately for each sub-type. It is possible that valence ratings could have been more positive, at least in the control group, if only ironic criticism had been used (Dews et al., 1995; Thompson et al., 2016), and this could have led to group differences also in facial EMG responses. Last, the ERP analysis was based on sensor-level analysis, and we did not utilize any source localization methods. Therefore, we cannot accurately estimate the sources of the activity in the two groups. Future studies should confirm our findings with a larger sample, and also investigate the sources of the brain activity related to irony comprehension.

One advantage of the present study is its design in which the non-ironic and ironic conditions were defined by the previous contextual picture, which provided either a non-ironic or ironic context for the commenting sentence. This arrangement allows for a valid comparison of the non-ironic and ironic conversations irrespective of the low-level stimulus features, the position of the keywords, or other potentially confounding factors because the comparison was always made between sentences that were physically identical. A definite strength of this study is also its use of multimodal recordings and two measurement sessions conducted on separate days. Namely, the ratings of congruency recognition were collected separately after the EEG and EMG measurements. This allowed the participants to focus on the stimuli during the EEG/EMG recordings without responding.

5 Conclusions

To summarize, facial EMG activity in the zygomaticus major was greater after ironic stimuli than after non-ironic stimuli, which corresponds to the behavioral evaluations in which the participants rated the ironic conversations funnier than the non-ironic ones. Thus, the conversational irony applied in our study seemed to evoke positive emotions. However, the valence ratings for all the stimuli were generally lower in the dysphoric group than in the control group, probably reflecting blunted emotional reactivity in dysphoria. The amplitudes of the irony-related ERPs, N400-like activity and P600, were greater for the ironic stimuli than the non-ironic stimuli, reflecting difficulties in integrating the irony-related keyword to the context and the cognitive effort required to interpret the ironic meaning, respectively. P600 had a different hemispheric balance in the dysphoric group and the control group; while the irony-related activity was larger in the right ROI than left ROI in the dysphoric group, no such difference in lateralization was evident in the control group. More research is needed to confirm this finding and to define the cortical sources of the activity related to irony comprehension in healthy and depressed brains.

6 Acknowledgments

The authors thank Kaisa Pentikäinen, Dr. Jari Kurkela and Dr. Juho Strömmer for their help in producing the spoken sentences that were applied as stimulus materials; Dr. Elisa Ruohonen for the training for cognitive tests; Iris Juvenius, Emilia Tuhkanen and Emmi Susi for their help in data acquisition; Petri Kinnunen and Lauri Viljanto for technical assistance;

Dr. Joona Muotka for advice in statistical analysis; and Professor Fengyu Cong, Yongjie Zhu and Dr. Weiyong Xu for their advice in the data analysis phase.

Funding: The work was supported by the National Social Science Foundation of China [Grant 14ZDB155] to HW.

7 References

Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013). Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative Ratings. *PLoS ONE*, *8*(3), e59312. https://doi.org/10.1371/journal.pone.0059312

Akimoto, Y., Sugiura, M., Yomogida, Y., Miyauchi, C. M., Miyazawa, S., & Kawashima, R. (2014). Irony comprehension: Social conceptual knowledge and emotional response. *Human Brain Mapping*, *35*(4), 1167–1178. https://doi.org/10.1002/hbm.22242

Allen, N. B., Trinder, J., & Brennan, C. (1999). Affective startle modulation in clinical depression: Preliminary findings. *Biological Psychiatry*, *46*(4), 542–550. https://doi.org/10.1016/S0006-3223(99)00025-6

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, *32*(6), 793–826. https://doi.org/10.1016/S0378-2166(99)00070-3

Balconi, M., & Amenta, S. (2008). Isn't it Ironic? An Analysis on the Elaboration of Ironic Sentences with ERPs. *The Open Applied Linguistics Journal*, *1*(1), 9–17. https://doi.org/10.2174/1874913500801010009

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling metaphor from context: An ERP study. *Frontiers in Psychology*, *7*, 559. https://doi.org/10.3389/fpsyg.2016.00559

Baptista, N. I., Manfredi, M., & Boggio, P. S. (2018). Medial prefrontal cortex stimulation modulates irony processing as indexed by the N400. *Social Neuroscience*, *13*(4), 495–510. https://doi.org/10.1080/17470919.2017.1356744

Basso, M. R., & Bornstein, R. A. (1999). Relative memory deficits in recurrent versus first-episode major depression on a word-list learning task. *Neuropsychology*, *13*(4), 557. https://doi.org/10.1037/0894-4105.13.4.557

Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. Retrieved from https://books.google.fi/books?id=6rigtdo0u2UC

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the Beck depression inventory-II. *San Antonio, TX: Psychological Corporation*, 1–82.

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry*, *165*(8), 969–977. https://doi.org/10.1176/appi.ajp.2008.08050721

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bora, E., & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *191*, 49–55. https://doi.org/10.1016/j.jad.2015.11.023

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/J.BRAINRESREV.2008.05.003

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, *28*(4), 676–691. https://doi.org/10.1016/J.CPR.2007.10.001

Caillies, S., Gobin, P., Obert, A., Terrien, S., Coutté, A., Iakimova, G., & Besche-Richard, C. (2019). Asymmetry of affect in verbal irony understanding: What about the N400 and P600 components? *Journal of Neurolinguistics*, *51*, 268–277. https://doi.org/10.1016/j.jneuroling.2019.04.004

Calmus, A., & Caillies, S. (2014). Verbal irony processing: How do contrast and humour correlate? *International Journal of Psychology*, *49*(1), 46–50. https://doi.org/10.1002/ijop.12003

Canal, P., Bischetti, L., Di Paola, S., Bertini, C., Ricci, I., & Bambini, V. (2019). ‘Honey, shall I change the baby? – Well done, choose another one’: ERP and time-frequency correlates of humor processing. *Brain and Cognition*, *132*, 41–55. https://doi.org/10.1016/j.bandc.2019.02.001

Canli, T., Sivers, H., Thomason, M. E., Whitfield-Gabrieli, S., Gabrieli, J. D., & Gotlib, I. H. (2004). Brain activation to emotional words in depressed vs healthy subjects. *Neuroreport*, *15*(17), 2585−2588.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New York: Routledge. https://doi.org/10.4324/9780203771587

Cornejol, C., Simonetti, F., Aldunate, N., Ibáñez, A., López, V., & Melloni, L. (2007). Electrophysiological Evidence of Different Interpretative Strategies in Irony Comprehension. *Journal of Psycholinguistic Research*, *36*(6), 411–430. https://doi.org/10.1007/s10936-007-9052-0

Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*(2), 71–74. https://doi.org/10.1016/S0304-3940(01)02387-4

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. https://doi.org/10.3758/BF03195780

Coulson, S., & Wu, Y. C. (2005). *Right Hemisphere Activation of Joke-related Information: An Event-related Brain Potential Study*. *Journal of cognitive neuroscience*, *17*(3), 494-506.

Cusi, A. M., Nazarov, A., Holshausen, K., MacQueen, G. M., & McKinnon, M. C. (2012). Systematic review of the neural basis of social cognition in patients with mood disorders. *Journal of Psychiatry and Neuroscience*, *37*(3), 154–169. https://doi.org/10.1503/jpn.100179

Del Goleto, S., Kostova, M., & Blanchet, A. (2016). Impaired context processing during irony comprehension in schizotypy: An ERPs study. *International Journal of Psychophysiology*, *105*, 17–25. https://doi.org/10.1016/j.ijpsycho.2016.04.009

Deldin, P., Keller, J., Casas, B. R., Best, J., Gergen, J., & Miller, G. A. (2006). Normal N400 in mood disorders. *Biological Psychology*, *71*(1), 74–79. https://doi.org/10.1016/J.BIOPSYCHO.2005.02.005

Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Processes*, *19*(3), 347–367. https://doi.org/10.1080/01638539509544922

Dimberg, U. (1990). Facial Electromyography and Emotional Reactions. *Psychophysiology*, *27*(5), 481–494. https://doi.org/10.1111/j.1469-8986.1990.tb01962.x

Dunn, B. D., Dalgleish, T., Lawrence, A. D., Cusack, R., & Ogilvie, A. D. (2004). Categorical and Dimensional Reports of Experienced Affect to Emotion-Inducing Pictures in Depression. *Journal of Abnormal Psychology, 113*(4), 654–660. https://doi.org/10.1037/0021-843X.113.4.654

Dynel, M. (2009). Beyond a Joke: Types of Conversational Humour. *Language and Linguistics Compass*, *3*(5), 1284–1299. https://doi.org/10.1111/j.1749-818X.2009.00152.x

Fiacconi, C. M., & Owen, A. M. (2015). Using Psychophysiological Measures to Examine the Temporal Profile of Verbal Humor Elicitation. *Plos One*, *10*(9), e0135902. https://doi.org/10.1371/journal.pone.0135902

Falkenberg, I., Jarmuzek, J., Bartels, M., & Wild, B. (2011). Do depressed patients lose their sense of humor? *Psychopathology*, *44*(2), 98–105. https://doi.org/10.1159/000317778

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feng, Y.-J., Chan, Y.-C., & Chen, H.-C. (2014). Specialization of neural mechanisms underlying the three-stage model in humor processing: An ERP study. *Journal of Neurolinguistics*, *32*, 59–70. https://doi.org/10.1016/j.jneuroling.2014.08.007

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(3), 811–828. https://doi.org/10.1037/a0035658

Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, *116*(3), 421–436. https://doi.org/10.1016/j.cognition.2010.06.005

Fossati, P., Harvey, P. O., Le Bastard, G., Ergis, A. M., Jouvent, R., & Allilaire, J. F. (2004). Verbal memory performance of patients with a first depressive episode and patients with unipolar and bipolar recurrent depression. *Journal of Psychiatric Research*, *38*(2), 137–144. https://doi.org/10.1016/j.jpsychires.2003.08.002

Gaudreau, G., Monetta, L., Macoir, J., Poulin, S., Laforce, R. J., & Hudon, C. (2015). Mental State Inferences Abilities Contribution to Verbal Irony Comprehension in Older Adults with Mild Cognitive Impairment. *Behavioural Neurology*, *2015*, 1–9. https://doi.org/10.1155/2015/685613

Gibbs Jr., R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*, 457–486.

Giora, R. (2003). On Our MindSalience, Context, and Figurative Language. In *On Our Mind: Salience, Context, and Figurative Language*. https://doi.org/10.1093/acprof:oso/9780195136166.001.0001

Giora, R., & Fein, O. (1999). Irony comprehension: The graded salience hypothesis. *Humor*, *12*(4), 425–436. https://doi.org/10.1515/humr.1999.12.4.425

Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Grade Salience and Indirect Negation. *Metaphor and Symbol*, *13*(2), 83–101. https://doi.org/10.1207/s15327868ms1302_1

Good, P. I. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses* (Third). New York: Springer.

Gotlib, I. H., & Joormann, J. (2010). Cognition and Depression: Current Status and Future Directions. *Annual Review of Clinical Psychology*, *6*(1), 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 149–188. https://doi.org/10.1080/01690960902965951

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Grice, H. P. (1975). Logic and conversation. In *Syntax and semantics 3: Speech arts* (pp. 41–58). Brill.

Hämäläinen, J., Landi, N., Loberg, O., Lohvansuu, K., Pugh, K., & Leppänen, P. H. T. (2018). Brain event-related potentials to phoneme contrasts and their correlation to reading skills in school-age children. *International Journal of Behavioral Development*, *42*(3), 357–372. https://doi.org/10.1177/0165025417728582

Henry, J. D., & Crawford, J. R. (2005). A meta-analytic review of verbal fluency deficits in depression. *Journal of Clinical and Experimental Neuropsychology*, *27*(1), 78–101. https://doi.org/10.1080/138033990513654

Herold, R., Varga, E., Hajnal, A., Hamvas, E., Berecz, H., Tóth, B., & Tényi, T. (2018). Altered neural activity during irony comprehension in unaffected first-degree relatives of schizophrenia patients - An fMRI study. *Frontiers in Psychology*, *8*, 1–15. https://doi.org/10.3389/fpsyg.2017.02309

Hutcherson, C. A., Goldin, P. R., Ochsner, K. N., Gabrieli, J. D., Feldman Barrett, L., & Gross, J. J. (2005). Attention and emotion: Does rating emotion alter neural responses to amusing and sad films? *NeuroImage*, *27*(3), 656–668. https://doi.org/10.1016/j.neuroimage.2005.04.028

Inoue, Y., Tonooka, Y., Yamada, K., & Kanba, S. (2004). Deficiency of theory of mind in patients with remitted mood disorder. *Journal of Affective Disorders*, *82*(3), 403–409. https://doi.org/10.1016/j.jad.2004.04.004

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kaakinen, J. K., Olkoniemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of Written Irony: An Eye Movement Study. *Discourse Processes*, *51*(4), 287–311. https://doi.org/10.1080/0163853X.2013.870024

Kalatzis, I., Piliouras, N., Ventouras, E., Papageorgiou, C. C., Rabavilas, A. D., & Cavouras, D. (2004). Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals. *Computer Methods and Programs in Biomedicine*, *75*(1), 11–22. https://doi.org/10.1016/j.cmpb.2003.09.003

Klumpp, H., Keller, J., Miller, G. A., Casas, B. R., Best, J. L., & Deldin, P. J. (2010). Semantic processing of emotional words in depression and schizophrenia. *International Journal of Psychophysiology*, *75*(2), 211–215. https://doi.org/10.1016/j.ijpsycho.2009.12.004

Kreuz, R. J., & Glucksberg, S. (1989). How to Be Sarcastic: The Echoic Reminder Theory of Verbal Irony. *Journal of Experimental Psychology: General*, *118*(4), 374–386. https://doi.org/10.1037/0096-3445.118.4.374

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. *University of Florida*, *Gainesville*. Tech Rep A-8.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12),920–933. https://doi.org/10.1038/nrn2532

Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. *Journal of Affective Disorders*, *86*(2–3), 247–258. https://doi.org/10.1016/j.jad.2005.02.007

Leggitt, J. S., & Gibbs, R. W. (2000). Emotional reactions to verbal irony. *Discourse Processes*, *29*(1), 1–24. https://doi.org/10.1207/S15326950dp2901_1

Leppänen, J. M. (2006). Emotional information processing in mood disorders: a review of behavioral and neuroimaging findings. *Current Opinion in Psychiatry*, *19*(1), 34–39. https://doi.org/10.1097/01.yco.0000191500.46411.00

Marinkovic, K., Baldwin, S., Courtney, M. G., Witzel, T., Dale, A. M., & Halgren, E. (2011). Right hemisphere has the last laugh: neural dynamics of joke appreciation. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(1), 113–130. https://doi.org/10.3758/s13415-010-0017-7

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, R. a. (2010). *The Psychology of Humor: An Integrative Approach (Google eBook)*. https://doi.org/10.1037/h0075400

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the humor styles questionnaire. *Journal of Research in Personality*, *37*(1), 48–75. https://doi.org/10.1016/S0092-6566(02)00534-2

Matthews, J. K., Hancock, J. T., & Dunham, P. J. (2006). The Roles of Politeness and Humor in the Asymmetry of Affect in Verbal Irony. *Discourse Processes*, *41*(1), 3–24. https://doi.org/10.1207/s15326950dp4101_2

Moran, E. K., Mehta, N., & Kring, A. M. (2012). Emotional responding in depression: Distinctions in the time course of emotion. *Cognition & Emotion*, *26*(7), 1153–1175. https://doi.org/10.1080/02699931.2011.638909

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, *24*(3), 691–701. https://doi.org/10.1016/j.cogbrainres.2005.04.003

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety*, *27*(12), 1135–1142. https://doi.org/10.1002/da.20755

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293. https://doi.org/10.1162/jocn.2010.21411

Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes reflected by P600 effects: Evidence from ERPs and neural oscillations. *PLoS ONE*, *9*(5), e96840. https://doi.org/10.1371/journal.pone.0096840

Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2005). Emotion context insensitivity in major depressive disorder. *Journal of Abnormal Psychology*, *114*(4), 627–639. https://doi.org/10.1037/0021-843X.114.4.627

Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion, 2*(2), 135–146. https://doi.org/10.1037/1528-3542.2.2.135

Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*(3), 514–523. https://doi.org/10.1162/jocn.2009.21370

Schwoebel, J., Dews, S., Winner, E., & Srinivas, K. (2000). Obligatory Processing of the Literal Meaning of Ironic Utterances: Further Evidence. *Metaphor and Symbol*, *15*(1–2), 47–61. https://doi.org/10.1207/s15327868ms151&2_4

Shibata, M., Terasawa, Y., Osumi, T., Masui, K., Ito, Y., Sato, A., & Umeda, S. (2017). Time course and localization of brain activity in humor comprehension: An ERP / sLORETA study. *Brain Research*, *1657*, 215–222. https://doi.org/10.1016/j.brainres.2016.12.010

Sloan, D. M., Strauss, M. E., & Wisner, K. L. (2001). Diminished response to pleasant stimuli by depressed women. *Journal of Abnormal Psychology, 110*(3), 488–493. https://doi.org/10.1037/0021-843X.110.3.488

Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What's behind a P600? Integration Operations during Irony Processing. *PLoS ONE*, *8*(6), 1–10. https://doi.org/10.1371/journal.pone.0066839

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Strömmer, J. M., Põldver, N., Waselius, T., Kirjavainen, V., Järveläinen, S., Björksten, S., … Astikainen, P. (2017). Automatic auditory and somatosensory brain responses in relation to cognitive abilities and physical fitness in older adults. *Scientific reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-14139-9

Talarowska, M., Zajączkowska, M., & Gałecki, P. (2015). Cognitive functions in first-episode depression and recurrent depressive disorder. *Psychiatria Danubina*, *27*(1), 38–43.

Thompson, D., Mackenzie, I. A. N. G., & Leuthold, H. (2016). Emotional responses to irony and emoticons in written language: Evidence from EDA and facial EMG. *Psychophysiology*, *53*(7), 1054–1062. https://doi.org/10.1111/psyp.12642

Toplak, M., & Katz, A. N. (2000). On the uses of sarcastic irony. *Journal of pragmatics*, *32*(10), 1467–1488.

Uekermann, J., Channon, S., Lehmkämper, C., Abdel-Hamid, M., Vollmoeller, W., & Daum, I. (2008). Executive function, mentalizing and humor in major depression. *Journal of the International Neuropsychological Society*, *14*(1), 55–62. https://doi.org/10.1017/S1355617708080016

Varga, E., Simon, M., Tényi, T., Schnell, Z., Hajnal, A., Orsi, G., … Herold, R. (2013). Irony comprehension and context processing in schizophrenia during remission - A functional MRI study. *Brain and Language*, *126*(3), 231–242. https://doi.org/10.1016/j.bandl.2013.05.017

Van Boxtel, A. (2010, August). Facial EMG as a tool for inferring affective states. In *Proceedings of measuring behavior* (pp. 104-108). Wageningen: Noldus Information Technology.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*(3), 279–293. https://doi.org/10.1016/j.bandl.2005.11.003

World Health Organization. (2010). ICD-10 Version: 2010. Retrieved from http://apps.who.int/classifications/icd10/browse/2010/en

Xu, Q., Ruohonen, E. M., Ye, C., Li, X., Kreegipuu, K., Stefanics, G., … Astikainen, P. (2018). Automatic Processing of Changes in Facial Emotions in Dysphoria: A Magnetoencephalography Study. *Frontiers in Human Neuroscience*, *12*, 186. https://doi.org/10.3389/fnhum.2018.00186

Table 1. Demographics and clinical measures. Statistics show independent t-test or x^2 test values investigating the group differences.

Characteristics		DYS (n = 17)	CTRL (n = 21)	Statistics
Age (year)	Mean	24.82	23.52	$t(36) = 1.15, p = .255$
	SD[range]	3.73 [21-33]	3.20 [19-32]	
Level of Education[a]	Medium	9	9	$x^2(1) = .38, p = .536$
	High	8	12	
Gender	Male	3	6	$x^2(1) = .62, p = .431$
	Female	14	15	
BDI-II	Mean	23.71	2.52	$t(36) = 13.48, p < .001$
	SD[range]	6.59 [14-39]	2.64 [0-9]	
Severity of Depression[b]	Moderate (F32.1)	12	Na	Na
	Severe (F32.2)	1	Na	Na
Time of Diagnosis	Within six months	3	Na	Na
	Within one year	4	Na	Na
	Over one year	9	Na	Na
Depression Medication[c]	SSRI	5	Na	Na
	SNRI	1	Na	Na
	Other	4	Na	Na

Note. DYS = dysphoric group, CTRL = control group, SD = standard deviation, BDI-II = Beck's Depression Inventory, Second Edition.

[a]Medium = high school or vocational school, High = university.
[b]Depression severity based on participants self-report on their diagnosis. The diagnosis of depression was based on the International Classification of Diseases and Related Health Problems, 10th Revision (ICD-10; World Health Organization, 2010) criteria. There is missing information related to disease severity from three participants, and one participant did not have diagnosis.
[c]SSRI = selective serotonin reuptake inhibitor, SNRI = serotonin and norepinephrine reuptake inhibitor, Other = other antidepressant medication.

Table 2. Results of repeated measures ANOVAs for facial EMG corrugator and zygomaticus activity. F = F-values, df = degrees of freedom, p = p-values, η_p^2 = partial eta squared for effect size.

Facial EMG activity	**Effect**	**F**	**df**	**p**	**η_p^2**
Corrugator activity	Stimulus type (non-ironic vs. ironic)	0.863	1, 36	.359	.023
	Time window (1 vs. 2 vs. 3)	6.565	2, 72	.009	.154
	Group (control vs. dysphoric)	0.743	1, 36	.394	.020
	Stimulus type × Time window	1.142	2, 72	.313	.031
	Stimulus type × Group	0.427	1, 36	.518	.012
	Time window × Group	0.082	2, 72	.922	.002
	Stimulus type × Time window × Group	0.680	2, 72	.510	.019
Zygomaticus activity	Stimulus type (non-ironic vs. ironic)	4.579	1, 36	.039	.113
	Time window (1 vs. 2 vs. 3)	6.642	2, 72	.002	.156
	Group (control vs. dysphoric)	0.561	1, 36	.459	.015
	Stimulus type × Time window	2.436	2, 72	.092	.064
	Stimulus type × Group	0.347	1, 36	.560	.010
	Time window × Group	0.293	2, 72	.747	.008
	Stimulus type × Time window × Group	0.964	2, 72	.386	.026

Table 3. Results of repeated measures ANOVAs for amplitudes of the N400-like activity and P600. F = F-values, df = degrees of freedom, p = p-values, η_p^2 = partial eta squared for effect size.

ERP	Effect	F	df	p	η_p^2
N400	Stimulus type (non-ironic vs. ironic)	8.279	1, 36	.007	.187
	Group (control vs. dysphoric)	0.170	1, 36	.682	.005
	Stimulus type × Group	0.077	1, 36	.783	.002
P600	Stimulus type (non-ironic vs. ironic)	23.243	1, 36	< .001	.392
	ROI (left vs. right)	0.001	1, 36	.973	< .001
	Group (control vs. dysphoric)	0.118	1, 36	.734	.003
	Stimulus type × ROI	0.337	1, 36	.565	.009
	Stimulus type × Group	1.656	1, 36	.206	.044
	ROI × Group	2.157	1, 36	.151	.057
	Stimulus type × ROI × Group	4.844	1, 36	.034	.119

Table 4. Mean amplitude values (μV) for P600 response for each ROI and each stimulus type in the control and dysphoric groups. SD = standard deviation.

Group	**Stimulus type**	**Group**	**Mean (SD)**
Control	Non-ironic	Left	0.86 (1.07)
		Right	0.75 (0.84)
	Ironic	Left	1.34 (1.16)
		Right	0.91(1.16)
Dysphoria	Non-ironic	Left	0.62 (0.61)
		Right	0.60 (1.09)
	Ironic	Left	0.90 (0.66)
		Right	1.44 (0.92)

Supplementary Material

1 Humor Styles Questionnaire

Humor Styles Questionnaire (HSQ; Martin et al., 2003) was filled out on the same day with the second day's behavioral measurement. The HSQ, translated into Finnish based on the original version, consists of a 32-item measure with a Likert-scale from 1 to 7 (totally disagree = 1, totally agree = 7). The HSQ evaluates individual preference for the use of humor in four dimensions (affiliative, self-enhancing, aggressive, self-defeating) and has been applied in previous studies to investigate the relationship between the use of humor styles and depressive symptoms, loneliness, or suicidal ideation (e.g., Frewen et al., 2008; Schermer et al., 2017; Tucker et al., 2013). Here, HSQ was applied to explore if there were differences in humor style preference between the dysphoric and control groups, which could explain possible group differences in the responsiveness to irony, since irony is related to sarcasm (Kreuz and Glucksberg, 1989) that can be categorized as aggressive humor (Martin et al., 2003).

1.1 Analysis and statistics for the Humor Style Questionnaire and behavioral data

The reliability of the HSQ was first estimated based on internal consistency values. Cronbach's alphas for each dimension were: affiliative humor = .853; self-enhancing = .829; aggressive humor = .684; self-defeating = .756. The averaged HSQ scores were calculated separately for each participant and dimension. For the HSQ scores, multivariate analysis of variance (MANOVA) was applied with a within-subject factor Humor style (affiliative, self-enhancing, aggressive, self-defeating) and a between-subject factor Group (control, dysphoric). An additional MANOVA analysis, similar to the previously mentioned but with current medication status (whether the participant currently had depression medication or not) serving as a covariate, was conducted. Post-hoc tests for within-subject comparisons were implemented by applying repeated measures of analysis of variance (ANOVAs) with Bonferroni correction in the control and dysphoric groups separately. Post-hoc tests comparing each humor style dimension between the control and dysphoric groups were performed by using two-tailed independent samples *t*-tests with Bootstrap statistics based on 1000 permutations as implemented in IBM SPSS Statistics 24.0 (Armonk, NY: IBM corporated). *P*-values in multiple comparisons for independent sample *t*-tests were adjusted by false discovery rate (Benjamini and Hochberg, 1995).

Pearson's correlation coefficients (two-tailed) were used to evaluate the relationship between depressive symptoms (BDI-II scores) and humor styles. Multiple correlations were controlled by applying false discovery rate (Benjamini and Yekutieli, 2001) at 0.05.

1.2 Results for Humor Style Questionnaire

The MANOVA indicated no main effect of Group ($p = .929$). A main effect of Humor style was found, $F(3,34) = 63.356$, $p < .001$, $\eta_p^2 = .848$. The main effect was modified by an interaction effect of Humor style × Group, $F(3,34) = 3.827$, $p = .018$, $\eta_p^2 = .252$. MANOVA with current medication status as a covariate variable showed no Stimulus condition × Current medication status interaction, $p = .587$.

Separate ANOVAs for the two groups showed a significant main effect of Humor style in each group (dysphoric: $F(3,48) = 22.616$, $p < .001$, $\eta_p^2 = .586$; control: $F(3,60) = 47.029$, $p < .001$, $\eta_p^2 = .702$). For the dysphoric group, pairwise comparisons with Bonferroni correction showed that the tendency to use an affiliative humor style in the dysphoric participants was greater compared with the tendency to use a self-enhancing, an aggressive, or a self-defeating humor style, all *p*-values < .001. The comparisons also showed that the dysphoric participants were more likely to use a self-defeating humor style than an aggressive humor style, p = .005. There were no differences between using a self-enhancing humor style and an aggressive or a self-defeating humor style. Among the controls, pairwise comparisons with Bonferroni correction showed that the control group had the greatest preference for an affiliative humor style compared with a self-enhancing (p = .009), an aggressive (p < .001), or a self-defeating (p < .001) humor style. Furthermore, the control group were most unlikely to use an aggressive humor style compared with an affiliative, a self-enhancing (p < .001), or a self-defeating (p = .005) humor style. The pairwise comparisons also revealed that the control participants prefer to use a self-enhancing humor style rather than a self-defeating humor style, p = .003. Results of post-hoc tests comparing each humor style dimension between the control and dysphoric groups are shown in Supplementary Table 3. Mean values, standard deviation (SD) and range of Humor Style Questionnaire (HSQ) scores separately for each humor style dimension and group are also presented in Supplementary Table 3.

No correlations were found between the amount of depressive symptoms (measured with the BDI-II scores) and scores for each HSQ dimension when the whole sample was included or when only the dysphoric group was included in the analysis (all p-values > .100).

2 Linear regression analyses

In addition to the ANOVAs reported in the main text, simple linear regression analyses were conducted for the whole sample to investigate the relationship between the amount of depressive symptoms (BDI-II scores) and ERP responses (P600 only, because we hypothesized a group difference specifically for it), facial EMG amplitudes and behavioral evaluations (group difference hypothesized for valence ratings of conversation) separately. BDI-II scores were applied as a predictor of P600 amplitude, facial EMG amplitude and behavioral ratings for valence. Results of simple linear regression model with BDI-II scores as a predictor are presented in Supplementary Table 4.

In summary, the results of simple linear regression analyses showed that there were no significant relationships between BDI-II scores and P600 amplitudes, or facial EMG amplitudes, or behavioral valence ratings.

3 Cognitive tests

3.1 Principal component analysis on cognitive tests

In order to reduce variables for analysis of covariance investigating effect of cognitive abilities for P600 responses, principal component analysis (PCA) using an Oblimin rotation with Kaiser Normalization was conducted to extract factors of the cognitive tests. First, we selected the cognitive tests that can be expected to associate with irony processing or the cognitive deficits in depression (Supplementary Table 5). Thus, eight tests were applied as variables in the PCA to investigate factor solutions. With a criteria of eigenvalue larger than 1.0, the PCA showed that three factors which could explain 71.5% of the variance, were extracted (Supplementary Table 5). Finally, these three factors, named executive function, list memory, and semantic processing, were utilized as covariates independently in ANCOVA models of P600. In addition, a two-way repeated measures ANOVA with a within-subject variable Cognitive factor (executive function vs. list memory vs. semantic processing) and a between-subject variable Group (control vs. dysphoric) was conducted to examine possible group difference in cognitive factors.

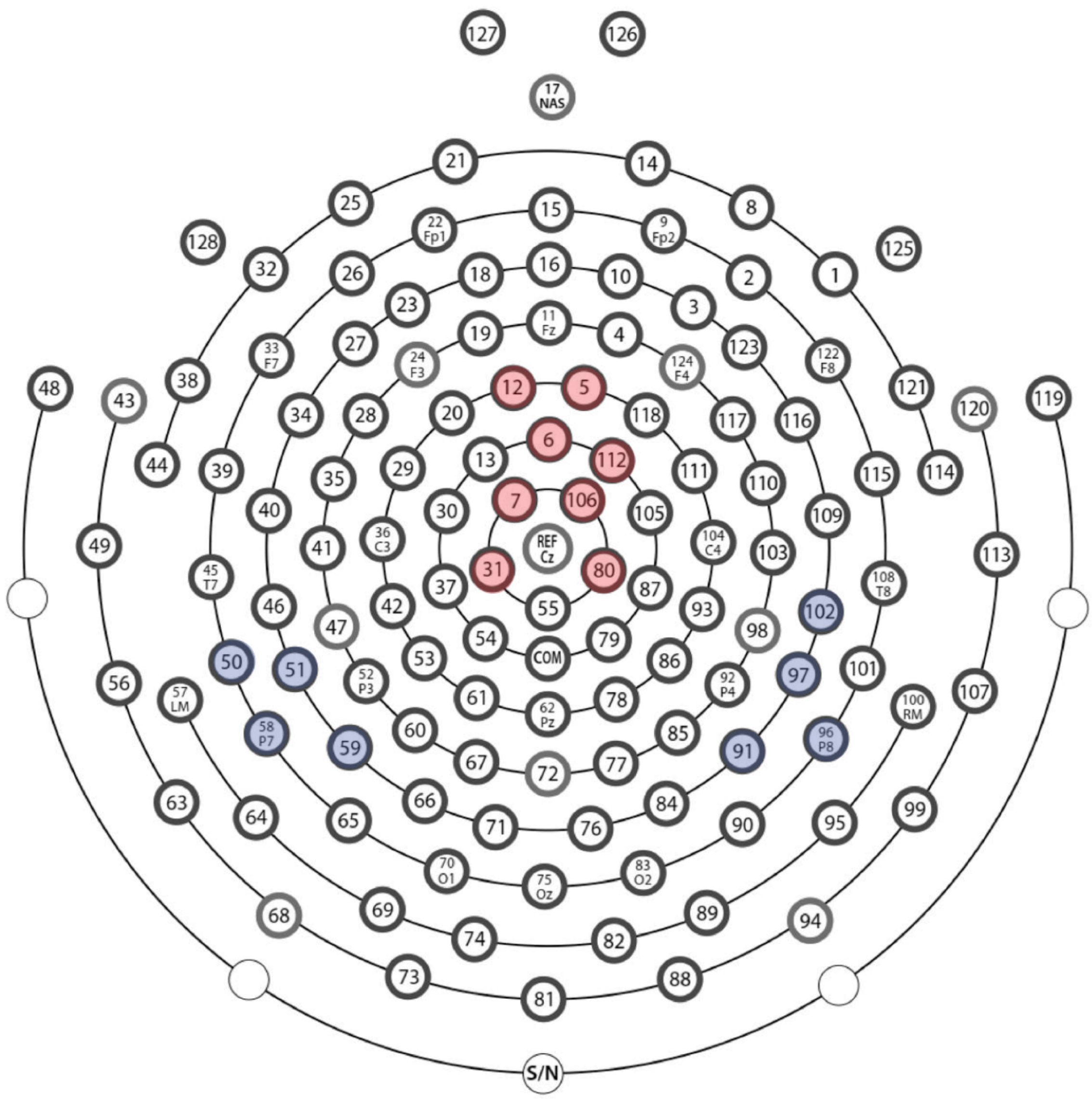
3.2 Results of group comparison in cognitive factors

The two-way repeated measures ANOVA showed that there was neither main effect of Cognitive factor ($F(2, 72) = 0.015$, $p = .985$, $\eta_p^2 < .001$) nor main effect of Group ($F(1, 36) = 0.917$, $p = .345$, $\eta_p^2 = .025$). The interaction of Cognitive factor and Group were non-significant, $F(2, 72) = 1.368$, $p = .261$, $\eta_p^2 = .037$.

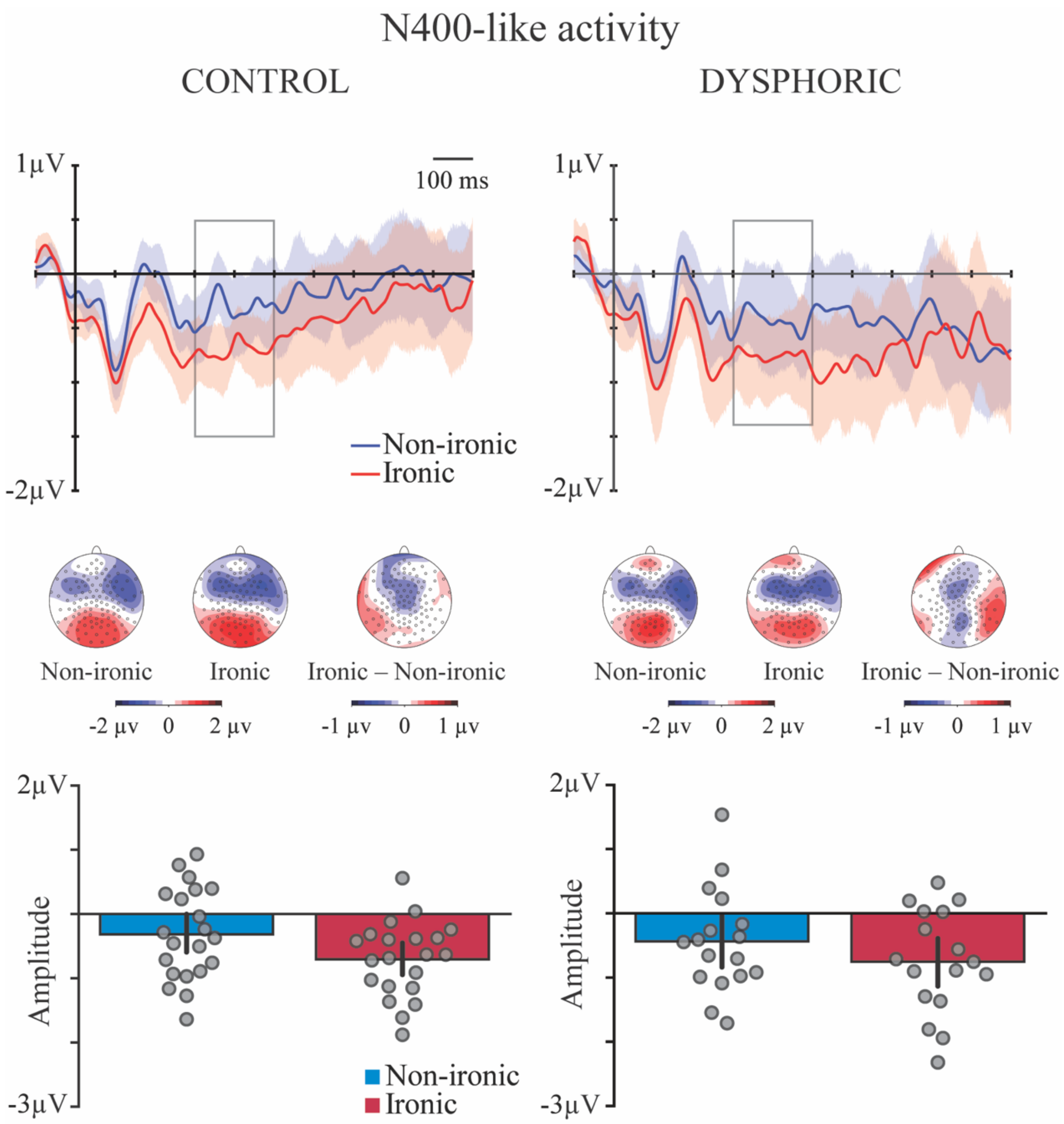
4 Permutation tests for electrode selection

For the N400-like activity and P600, the selection of the electrodes for the further statistical analyses was based on a data-driven method (see also, e.g., Hämäläinen et al., 2018; Strömmer et al., 2017). BESA Statistics 2.0 software (BESA GmbH, Graefelfing, Germany) was applied to perform cluster-based permutation tests (Maris and Oostenveld, 2007) between the amplitude values of responses to ironic and non-ironic trials with 3000 iterations. In order to enhance the sensitivity of the test (Groppe, Urbach, and Kutas, 2011; Maris and Oostenveld, 2007), time points where stimulus type effect (non-ironic vs. ironic) related N400-like activity and P600 is unlikely to be observed were excluded. That is, we performed the permutation tests including each time point from two separate time windows: 300 to 500 ms after stimulus onset for the N400-like activity (Kutas & Federmeier, 2011), and 500 to 800 ms after stimulus onset for P600 (Bornkessel-Schlesewsky & Schlesewsky, 2008). For both tests, the channel (electrode) cluster distance was set to 3.5 cm, and the alpha level for significant cluster was 0.05. In time window of 300-500 ms (N400-like activity), the cluster-based test showed one electrode cluster ($p = .025$), where the responses to ironic and non-ironic sentences differed. In time window of 500-800 ms (P600), two clusters were observed: one on the left ($p = .013$) and one on the right ($p = .014$) hemisphere. Then, we located the highest t-value (t-value$_{max}$) within each electrode cluster: in the electrode cluster for N400-

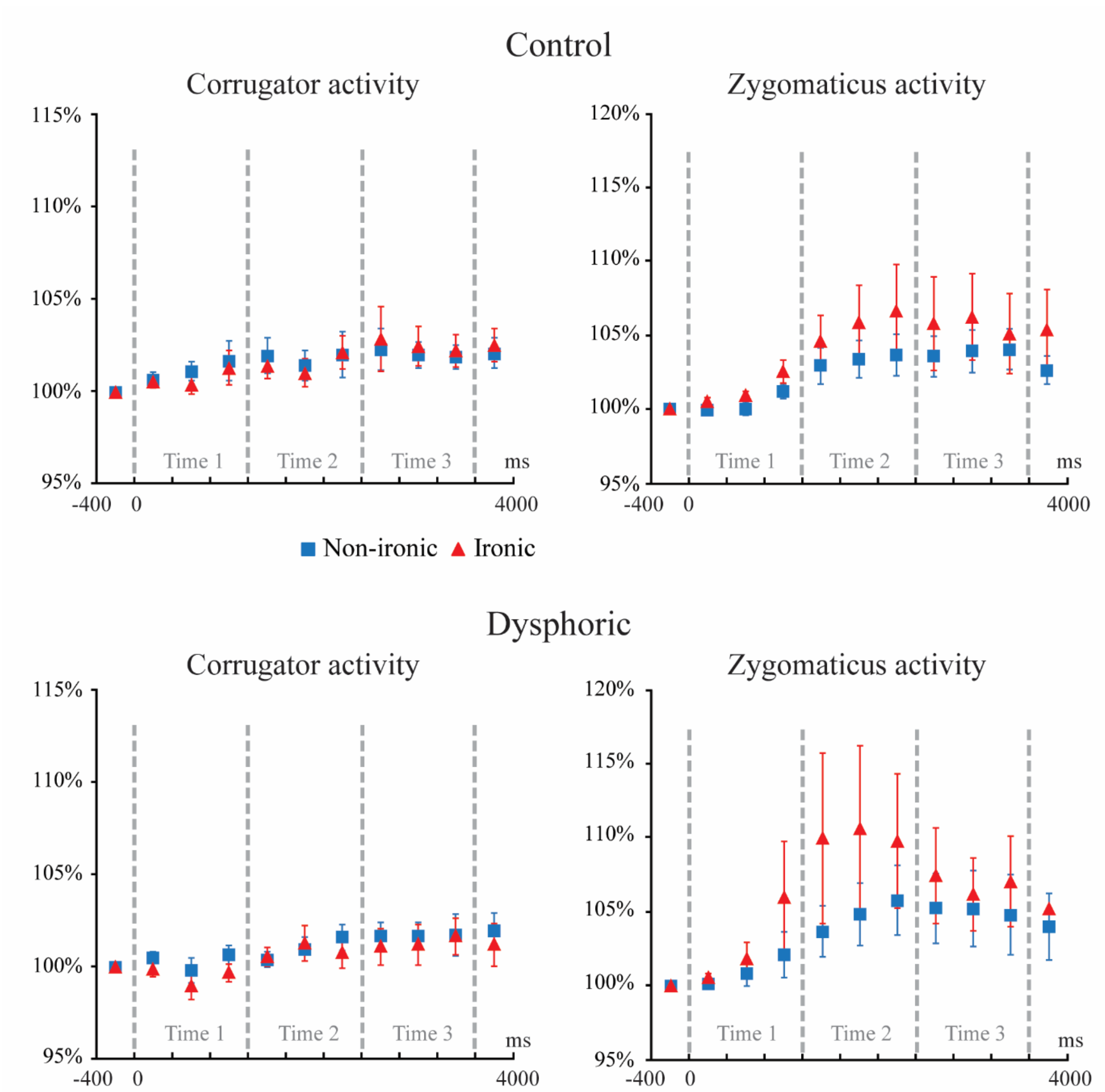
like activity: t-value$_{max}$ = 4.0; in the electrode clusters for P600, for the left cluster: t-value$_{max}$ = 4.5, and for the right cluster: t-value$_{max}$ = 4.0. The electrode with the highest t-value and its surrounding electrodes within the cluster were selected to the further analyses if the highest t-value of the electrode was larger than the 75% of t-value$_{max}$ (for N400-like activity: the highest *t*-values for all selected electrodes > 3.0; for P600: the highest *t*-values for selected electrodes on the left > 3.4, for selected electrodes on the right > 3.0). The electrodes applied in the statistical analysis, i.e., the results of this procedure, are depicted in Supplementary Figure 1.



Supplementary Figure 1. Map of EGI 128-Channel Net (HydroCel Geodesic Sensor Net) and the electrodes applied in the analyses. For N400-like activity, electrodes 5, 6, 7, 12, 31, 80, 106, 112 (red fillings) were applied in the analyses. For P600, electrodes 50, 51, 58, 59, 91, 96, 97, 102 (blue fillings) were applied in the analyses.



Supplementary Figure 2. Grand-averaged waveforms, topographical maps and histograms depicting mean amplitudes for N400-like activity to non-ironic and ironic stimuli in control (left) and dysphoric (right) group. Please refer to Supplementary Figure 1 for the electrode sites applied in the waveforms and histograms (averaged activity of the electrodes within the ROI presented here). In the waveform figure, the blue and red shadows represent 95% confidence intervals, and the grey rectangles the time window applied in the analysis for N400-like activity (mean amplitude values between 300–500 ms). The histograms shows the mean amplitudes in the analysis window and the error bars represent 95% confidence intervals. The scatterplots represent individual participants' amplitudes.



Supplementary Figure 3. The grand-averages of the mean EMG amplitude percentage relative to the baseline for the corrugator (left) and for the zygomaticus (right) to non-ironic (blue) and ironic (red) stimulus in control and dysphoric group. Average percentage values and standard error of mean (SE) are presented for descriptive purposes in 400-ms segments, and the grey dotted lines show the segments applied in the ANOVA (Time 1: 0–1200 ms; Time 2: 1200–2400 ms; Time 3: 2400–3600 ms).

Supplementary Table 1. Cognitive test scores in the control and dysphoric group. The statistics show the results based on two-tailed independent samples t-tests comparing the groups in the scores.

Cognitive test	Mean ± Standard deviation		Mean difference	p[a]	d
	Control (n = 21)	Dysphoric (n = 17)			
AVLT immediate (#)	13.33 ± 2.03	12.00 ± 2.72	1.33	.092	0.554
AVLT delayed (#)	13.10 ± 1.92	11.94 ± 2.11	1.15	.086	0.576
Digit span (p)	9.81 ± 1.60	10.41 ± 2.65	0.60	.418	0.183
Digit-letter (p)	12.57 ± 2.58	10.94 ± 3.34	1.63	.098	0.546
Logical memory immediate (p)	29.05 ± 5.39	29.71 ± 9.35	0.66	.799	0.143
Logical memory delayed (p)	27.81 ± 5.75	27.47 ± 9.04	0.89	.339	0.271
Symbol search (p)	12.52 ± 3.40	13.59 ± 3.47	1.06	.348	0.311
Digit symbol (p)	13.05 ± 2.38	13.59 ± 2.90	0.54	.531	0.204
TMT-A (s)	25.90 ± 8.78	25.53 ± 8.24	0.37	.894	0.044
TMT-B (s)	55.32 ± 23.71	55.38 ± 23.77	0.06	.994	0.003
Stroop1 reading (s)	47.16 ± 7.25	44.69 ± 6.37	2.46	.279	0.361
Stroop2 color labelling (s)	60.43 ± 9.81	62.35 ± 13.69	1.92	.618	0.161
Stroop3 inhibition (s)	87.37 ± 14.00	85.74 ± 17.76	1.63	.754	0.101
Similarities (p)	13.76 ± 2.14	13.12 ± 2.50	0.64	.398	0.275
Fluency phonemic (#)	23.93 ± 4.29	20.50 ± 3.93	3.45	**.016 (0.256)**	0.834
Fluency semantic (#)	32.05 ± 5.32	30.88 ± 5.01	1.17	.495	0.226

Note. Differences between two groups were tested by using two-tailed independent-samples t-tests. d = Cohen's d, AVLT = Auditory Verbal Learning Test, TMT = Trail Making Test, # = scores measured in the number of items, p = point, more means better, s = second, less means better.

[a]Uncorrected p-values. P-values smaller than .05 are in bold, and for them corrected p-values based on FDR (false discovery rate) are presented in parentheses.

Supplementary Table 2. List of cognitive tests applied and references to them.

Cognitive test (Version)	References
Memory	

Auditory Verbal Learning Test	Günther, T., Holtkamp, K., Jolles, J., Herpertz-Dahlmann, B., and Konrad, K. (2004). Verbal memory and aspects of attentional control in children and adolescents with anxiety disorders or depressive disorders. *J. Affect. Disord.* 82, 265–269. doi:10.1016/j.jad.2003.11.004.
Digit span (WAIS-III)	Ramsay, M. C., and Reynolds, C. R. (1995). Separate Digits tests: A brief history, a literature review, and a reexamination of the factor structure of the test of memory and learning (TOMAL). *Neuropsychol. Rev.* 5, 151–171. doi:10.1007/BF02214760.
Letter-number sequencing task (WMS-III)	Crowe, S. F. (2000). Does the Letter Number Sequencing task measure anything more than digit span? *Assessment* 7, 113–117. doi:10.1177/107319110000700202.
Logical memory task (WMS-R)	Elwood, R. W. (1991). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. *Neuropsychol. Rev.* 2, 179–201. doi:10.1007/BF01109053.
Processing speed	
Symbol search (WAIS-IV)	Wisdom, N. M., Mignogna, J., and Collins, R. L. (2012). Variability in wechsler adult intelligence scale-IV subtest performance across age. *Arch. Clin. Neuropsychol.* 27, 389–397. doi:10.1093/arclin/acs041.
Digit symbol substitution test/Coding (WAIS-IV)	
Executive function	
Trail Making Test A & Trail Making Test B	Bowie, C. R., and Harvey, P. D. (2006). Administration and interpretation of the Trail Making Test. *Nat. Protoc.* 1, 2277–2281. doi:10.1038/nprot.2006.390.
The Stroop Color-Word Test	Alvarez, J. A., and Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychol. Rev.* 16, 17–42. doi:10.1007/s11065-006-9002-x.
Linguistic ability	
Similarities (WAIS-IV)	Davies, G., and Piovesana, A. (2015). Adult Verbal Abstract Reasoning Assessment Instruments and their Clinimetric Properties. *Clin. Neuropsychol.* 29, 1010–1033. doi:10.1080/13854046.2015.1119889.
Fluency phonemic & Fluency semantic (Verbal fluency tests)	Laine, M. (1988). Correlates of word fluency performance. In P. Koivuselka-Sallinen & L. Sarajarvi (Eds.), Studies in languages (Vol. 12). Joensuu, Finland: University of Joensuu.

Note. WMS-R = Wechsler Memory Scale-Revised, WMS-III = Wechsler Memory Scale-Third Edition, WAIS-III = Wechsler Adult Intelligence Scale-Third Edition, WAIS-IV = Wechsler Adult Intelligence Scale-Fourth Edition.

Supplementary Table 3. Mean values, standard deviation (SD) and range of Humor Style Questionnaire (HSQ) scores separately for each humor style dimension and group reflecting significant interaction of humor style × group. Statistics show post hoc tests (two-tailed independent samples *t*-tests, bootstrapping with 1000 permutations) comparing the humor style scores between the groups in each dimension.

HSQ		**Control**	**Dysphoric**	**Statistics**
Affiliative	Mean	5.857	5.974	t (36) = 0.215, p = .831, d = 0.071
	SD[range]	0.937 [2.88-6.88]	0.852[4.13-7.00]	
Self-enhancing	Mean	5.095	4.346	t (36) = 2.14, p = .104, d = 0.713
	SD[range]	1.019[3.13-6.63]	1.138[2.13-6.13]	
Aggressive	Mean	3.250	3.450	t (36) = 0.671, p = .676, d = 0.223
	SD[range]	0.879[1.75-5.00]	1.012[1.88-5.38]	
Self-defeating	Mean	3.988	4.677	t (36) = 2.106, p = .104, d = 0.702
	SD[range]	0.826[2.25-5.38]	1.185[3.00-6.63]	

Note. Statistics show the results comparing the two groups in the preference for each humor style. Reported *p*-values are adjusted by false discovery rate.

Supplementary Table 4. Results of simple linear regression model with BDI-II scores as a predictor.

Dependent variables	**Coefficients (Beta)**	**R Square**	***P*-value**
P600 (non-ironic, left ROI)	-.178	.032	.258
P600 (non-ironic, right ROI)	-.044	.002	.794
P600 (ironic, left ROI)	-.254	.064	.124
P600 (ironic, right ROI)	.174	.030	.296
Zygomaticus activity (non-ironic)	.094	.009	.574
Zygomaticus activity (ironic)	.097	.009	.561
Corrugator activity (non-ironic)	-.051	.003	.762
Corrugator activity (non-ironic)	-.079	.006	.636
Valence ratings (non-ironic)	-.211	.045	.203
Valence ratings (ironic)	-.295	.087	.073

Supplementary Table 5. List and description of selected cognitive tests for principal component analysis, and factor loadings using an oblimin rotation with Kaiser Normalization on cognitive tests scores.

Cognitive test	Description	Principal component		
		Executive function	List memory	Semantic processing
TMT-A (s)	Connecting numbers by drawing straight lines between them in ascending order as fast and accurately as possible. This test measures sustained attention, cognitive flexibility, and hand motor speed.	**-0.678**	-0.003	-0.237
TMT-B (s)	Connecting numbers and letters by drawing straight lines between them in alternating order as fast and accurately as possible. Numbers are to be advanced in ascending order, while letters are to be connected in alphabetic order. This test measures divided attention, cognitive flexibility, and hand motor speed.	**-0.811**	0.068	-0.03
Similarities (p)	Identifying the qualitative relationship between two words. This test measures abstract thinking, verbal reasoning, and concept formatting.	**0.727**	0.167	-0.142
AVLT immediate (#)	Reading a list of 15 words repeatedly for five times. Reading another list of 15 words which serves as distractors. Recalling the words in the first list. This test measures immediate memory recall.	0.054	**0.958**	-0.097
AVLT delayed (#)	Recalling the words in the first list again after about one hour later. This test measures delayed memory recall.	-0.032	**0.916**	0.151
Logical memory immediate (p)	Listening to stories and repeating the stories immediately as accurately as possible. This test measures immediate auditory and declarative memory.	0.235	-0.029	**0.870**
Logical memory delayed (p)	Repeating the stories as accurately as possible after about one hour later. This test measures delayed auditory and declarative memory.	0.329	0.072	**0.818**
Fluency semantic-animals(#)	Naming words as fast as possible that belong to the animal category. This test measures semantic fluency.	-0.255	0.06	**0.566**

Note. TMT = Trail Making Test, AVLT = Auditory Verbal Learning Test, s = second, less means better, p = point, more means better, # = scores measured in the number of items. Factor loadings, for the component that the tests contribute most, are marked in bold.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/J.BRAINRESREV.2008.05.003

Frewen, P. A., Brinker, J., Martin, R. O. D. A., and Dozois, D. J. A. (2008). Humor styles and personality- vulnerability to depression. *Humor*, *21*(2), 179–195. doi:10.1515/HUMOR.2008.009.

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), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Hämäläinen, J., Landi, N., Loberg, O., Lohvansuu, K., Pugh, K., & Leppänen, P. H. T. (2018). Brain event-related potentials to phoneme contrasts and their correlation to reading skills in school-age children. *International Journal of Behavioral Development*, *42*(3), 357–372. https://doi.org/10.1177/0165025417728582

Kreuz, R. J., and Glucksberg, S. (1989). How to Be Sarcastic: The Echoic Reminder Theory of Verbal Irony. *Journal of experimental psychology. 118*(4), 374–386. doi:10.1037/0096-3445.118.4.374.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the humor styles questionnaire. *Journal of research in personality. 37*(1), 48–75. doi:10.1016/S0092-6566(02)00534-2.

Schermer, J. A., Martin, R. A., Vernon, P. A., Martin, N. G., Conde, L. C., Statham, D., et al. (2017). Lonely people tend to make fun of themselves: A behavior genetic analysis of humor styles and loneliness. *Personality and Individual Differences*. *117*, 71–73. doi:10.1016/j.paid.2017.05.042.

Strömmer, J. M., Põldver, N., Waselius, T., Kirjavainen, V., Järveläinen, S., Björksten, S., … Astikainen, P. (2017). Automatic auditory and somatosensory brain responses in relation to cognitive abilities and physical fitness in older adults. *Scientific reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-14139-9

Tucker, R. P., Wingate, L. R. R., O'Keefe, V. M., Slish, M. L., Judah, M. R., and Rhoades-Kerswill, S. (2013). The moderating effect of humor style on the relationship between interpersonal predictors of suicide and suicidal ideation. *Personality and Individual Differences*. *54(5)*, 610–615. doi:10.1016/j.paid.2012.11.023.