

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**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)

**Copyright:** © 2020 Elsevier Inc. All rights reserved.

**Rights:** CC BY-NC-ND 4.0

**Rights url:** <https://creativecommons.org/licenses/by-nc-nd/4.0/>

**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>

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

3 Xueqiao Li<sup>\*a</sup>, Janne Pesonen<sup>a</sup>, Elina Haimi<sup>a</sup>, Huili Wang<sup>b</sup>, Piia Astikainen<sup>a</sup>

4 Affiliation:

5 <sup>a</sup>Department of Psychology  
6 University of Jyväskylä  
7 P. O. Box 35  
8 FIN-40100 Jyväskylä  
9 Finland

10 <sup>b</sup>School of Foreign Languages  
11 Dalian University of Technology  
12 Dalian  
13 China

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

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

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

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

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

24 <sup>\*</sup> *Corresponding author:* Xueqiao Li

25 **Abstract**

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

39

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

## 42 1 Introduction

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

49 The present study aimed to investigate irony comprehension and irony-related emotional  
50 reactions, and identify possible differences between dysphoric (individuals with elevated  
51 number of depressive symptoms) and non-dysphoric participants in these aspects. Facial  
52 electromyography (EMG) and event-related potentials (ERPs) were used to measure  
53 emotional facial expressions and cognitive processing related to irony, respectively.

54 Differences in the irony-related emotional reactions and irony comprehension were expected  
55 between dysphoric and control groups, because previous studies have found alterations in  
56 emotional reactivity (for reviews, see Bylsma et al., 2008; Cusi et al., 2012), and cognitive  
57 function (for a review, see Gotlib and Joormann, 2010), such as verbal fluency (Henry and  
58 Crawford, 2005) and mentalizing (Bora and Berg, 2016) in clinical and preclinical  
59 depression. Impairments in irony comprehension have been found in schizophrenia and  
60 schizotypy (e.g., Del Goleto et al., 2016; Herold et al., 2018; Varga et al., 2013); however,  
61 whether or how irony comprehension is altered in depressed individuals has not yet been  
62 explored.

### 63 1.1 Emotional reactions to irony

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

74 In addition to behavioral ratings, psychophysiological reactivity, such as facial muscle  
75 activity (i.e., facial EMG), is of interest when attempting to understand irony-evoked  
76 emotions (Thompson, Mackenzie, and Leuthold, 2016). Facial reactivity in the zygomaticus  
77 major muscle (cheek area) and in the corrugator supercilii muscle (brow region) is an index  
78 of emotional expressions corresponding to smiling and frowning, respectively (Van Boxtel,  
79 2010; Dimberg, 1990). Facial EMG has been measured in response to humor (e.g., Fiacconi  
80 and Owen, 2015), but only a few studies have investigated facial EMG responses to irony. In  
81 one study, facial EMG was measured in relation to written ironic praise and criticism, with  
82 and without emoticons (Thompson et al., 2016); it was found that ironic criticism reduced  
83 frowning and enhanced smiling in comparison to literal criticism, indicating weakening of the  
84 negative emotional response due to irony. However, no previous studies have investigated  
85 facial reactivity to spoken ironic statements.

86 1.2 Theoretical models and empirical evidence of irony comprehension

87 Different models have been used to explain irony comprehension (for a review, see Attardo,  
88 2000). The one-stage model (also called the direct access view, e.g., Gibbs, 2002) in general  
89 proposes that the figurative language (written or oral), such as irony, can be accessed directly  
90 as the literal language. In contrast, two-stage models, such as the traditional standard  
91 pragmatic view (e.g., Grice, 1975) and the graded salience hypothesis (e.g., Giora and Fein,  
92 1999; Giora, 2003), suggest that irony comprehension requires access to both literal and  
93 ironic meanings. The former expects that, in irony comprehension, one should first access the  
94 literal meaning, detect and distinguish the discrepancy in the semantic context, and then  
95 reconstruct the information to retrieve the intended ironic meaning. The graded salience  
96 hypothesis assumes that there are two stages in irony comprehension, but it makes no explicit  
97 hypothesis about their sequential order.

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

105 The cognitive processes in irony comprehension have been investigated using time-sensitive  
106 ERP method in healthy participants (e.g., Balconi and Amenta, 2008; Baptista et al., 2018;  
107 Caillies et al., 2019; Cornejo et al., 2007; Filik et al., 2014; Regel et al., 2011, 2014;  
108 Spotorno et al., 2013). These studies have shown that comprehending ironic sentences  
109 requires additional inferential processes in comparison to non-ironic sentences; this finding is  
110 consistent with the traditional standard pragmatic view (e.g., Regel et al., 2011) or the graded

111 salience hypothesis (e.g., Filik et al., 2014). In many of the studies (Cornejo et al., 2007;  
112 Filik et al., 2014; Caillies et al., 2019), two ERP components, N400 and P600, were elicited  
113 consecutively, and both were larger in amplitude for irony than non-irony, reflecting the two-  
114 stage processing of irony comprehension.

115 The N400 is an ERP component that is triggered by semantic violations and modulated by the  
116 context given to the statement (for reviews, see Kutas and Federmeier, 2011; Lau et al.,  
117 2008). It is observed as a shift towards negative polarity at 200–600 ms after a stimulus onset  
118 at the centro-parietal electrode sites (e.g., Kutas and Federmeier, 2000; Kutas and Hillyard,  
119 1980; Van Petten and Luka, 2006). N400 amplitude modulations have been interpreted to  
120 reflect either an additional effort in semantic integration (Kutas and Federmeier, 2011; Lau et  
121 al., 2008) or the difficulty of retrieving the stored knowledge associated with the stimulus  
122 (Brouwer et al., 2012). Larger N400 amplitude has been observed during the processing of  
123 nonliteral language, for instance, humor (e.g., Coulson and Kutas, 2001; Feng, Chan, and  
124 Chen, 2014; Marinkovic et al., 2011) and irony (Baptista et al., 2018; Caillies et al., 2019;  
125 Cornejo et al., 2007; Filik et al., 2014). Several studies (Cornejo et al., 2007; Filik et al.,  
126 2014; Caillies et al., 2019) have demonstrated that the modulation of N400 reflects  
127 difficulties in integrating the meaning of an irony-related incongruent word into the context;  
128 thus, N400 modulation has been found to be one of the indicators reflecting the two-stage  
129 processing of irony. However, the findings related to irony are inconsistent, as the N400  
130 effect is not always found in response to irony (e.g., Balconi and Amenta, 2008; Regel et al.,  
131 2011, 2014).

132 In contrast to the inconsistent findings on the N400 elicitation to irony, the amplitude of the  
133 P600 has been repeatedly found to be larger for ironic than non-ironic sentences (Filik et al.,  
134 2014; Regel et al., 2011, 2014; Spotorno et al., 2013). P600 is a shift towards positive

135 polarity, reaching its maximum amplitude approximately at 600 ms latency at the centro-  
136 parietal electrode sites (for a review, see Bornkessel-Schlesewsky and Schlewsky, 2008). It  
137 was initially observed when the participants encountered syntactic anomalies (e.g., “The  
138 woman encouraged to write a blog.”; “the fancy very car”) and it was interpreted as a  
139 reflection of syntactic reanalysis (e.g., Gouvea et al., 2010; Osterhout and Holcomb, 1992).  
140 The P600 effect has also been found in response to semantic violations (e.g., Nieuwland and  
141 Van Berkum, 2005; Sanford et al., 2011), which was interpreted as continued analysis to  
142 achieve conflict resolution or the updating of mental representations (for reviews, see  
143 Bornkessel-Schlesewsky and Schlewsky, 2008; Brouwer et al., 2012; Kuperberg, 2007) in  
144 language comprehension. Increased P600 amplitude is often observed in response to irony  
145 (Filik et al., 2014; Regel et al., 2011, 2014; Spotorno et al., 2013), humor (e.g., Feng et al.,  
146 2014; Canal, et al., 2019; Shibata, et al., 2017; Marinkovic, et al., 2011) and figurative  
147 language in general (e.g., Bambini, et al., 2016). It has been suggested that the modulation of  
148 the P600 reflects the cognitive effort made when interpreting the implied meaning of  
149 nonliteral language (for irony, see e.g., Regel et al., 2011, 2014).

### 150 1.3 Depression and possible emotional and cognitive alterations related to irony 151 comprehension

152 Depression is a severe mental health disorder characterized by both affective and cognitive  
153 symptoms, including sadness, tiredness, disturbances in sleep and appetite, and feelings of  
154 guilt or low self-worth (World Health Organization, 2010). Depressive individuals possess  
155 persistent mood-congruent rumination that negatively affects their ability to process  
156 emotional information (for reviews, see Gotlib and Joormann, 2010; Peckham, McHugh, and  
157 Otto, 2010). One of the most prominent features in depression is blunted reactivity to  
158 emotional stimuli (see a review, Bylsma et al., 2008), such as to pleasant pictures (e.g., Allen,



159 Trinder, and Brennen, 1999; Dunn et al., 2004; Sloan, Strauss, and Wisner, 2001), positive  
160 words (Canli et al., 2004), and affective audiovisual videos (Rottenberg et al., 2002; 2005).

161 Concerning nonliteral language processing, humor comprehension has been found to be  
162 altered in depression (Uekermann et al., 2008). Depressed patients performed worse than the  
163 controls in selecting correct punchlines to humorous discussions, and they rated the  
164 humorous punchlines as being less funny than the controls (Uekermann et al., 2008).

165 Alterations in humor processing have been associated with impairments in executive  
166 functions and mentalizing in depressed participants (Bora and Berg, 2016; Uekermann et al.,  
167 2008). Deficits in mentalizing have also been suggested to underlie the impairments in irony  
168 processing in individuals with schizophrenia and schizotypal personality disorder (e.g., Del  
169 Goleto et al., 2016; Herold et al., 2018; Varga et al., 2013), but no studies have investigated  
170 irony processing in people with depression. In addition to mentalizing, irony comprehension  
171 requires verbal fluency and verbal memory (e.g., Gaudreau et al., 2015; Spotorno et al.,  
172 2012), both of which are impaired in depression (Basso and Bornstein, 1999; Henry and  
173 Crawford, 2005). Based on the previous findings on the blunted emotional reactions and the  
174 cognitive deficits in depression, it is very likely that both emotional reactions to ironic stimuli  
175 and the cognitive processing aspect of irony are affected by depressive symptoms.

#### 176 1.4 The present study

177 In the current study, conversational irony was investigated by presenting the participants with  
178 spoken dyadic conversations that resembled the natural occurrence of irony in the daily life.  
179 The keywords in the conversations were allocated to different positions, which made the  
180 stimuli less predictable and more naturalistic in comparison to previous studies in which the  
181 keyword was always presented as the last word of the sentence (e.g., Filik et al., 2014; Regel  
182 et al., 2011, 2014). The conversations were combined with pictures that were either

183 congruent or incongruent with the statements; thus, the conversation was defined as either  
184 non-ironic or ironic, respectively. This approach allowed the responses to be contrasted to the  
185 same spoken sentences (non-ironic vs. ironic), avoiding the confounding caused by possible  
186 differences in the physical features of the stimuli, e.g., the position of the keywords.

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

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

206 the corrugator supercili, reflecting frowning (Van Boxtel, 2010; Dimberg, 1990), should be  
207 observed.

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

218 For the ERP responses, we hypothesized that an N400-like effect (Filik et al., 2014) would be  
219 similar in the dysphoric and control groups, as reported in previous studies where the  
220 depression group and the controls showed similar N400 amplitudes in semantic processing of  
221 congruent and incongruent sentence endings (Deldin et al., 2006; Klumpp et al., 2010).

222 However, it is possible that the N400-like activity would not be elicited, as the recognition of  
223 semantic incongruence may not be a necessary processing stage for irony comprehension  
224 (Balconi and Amenta, 2008; Regel et al., 2011, 2014).

225 In addition, we hypothesized that irony would elicit a larger P600 response than non-ironic  
226 stimuli, as has been consistently reported in previous irony studies (e.g., Filik et al., 2014;  
227 Regel et al., 2011, 2014), and it is possible that the P600 response would be decreased in  
228 amplitude in the dysphoric group. While no previous studies have investigated the P600  
229 response to ironic stimuli in depressed participants, one study has shown that P600 elicited in

230 a working memory task can be used to distinguish between depressed and control participants  
231 through using a machine learning approach (Kalatzis et al, 2004). In schizotypal personality  
232 disorder, a decrease in P600 modulation in relation to irony has been associated mainly with  
233 difficulties in mentalizing (Del Goleto et al., 2016). Since there is also a deficit in mentalizing  
234 (Bora and Berk, 2016; Inoue et al., 2004; Lee et al., 2005) and in other cognitive functions in  
235 depression, such as verbal fluency and verbal memory (Basso and Bornstein, 1999; Henry  
236 and Crawford, 2005), it is possible that a decrease in P600 will be found in the dysphoric  
237 group.

## 238 **2 Materials and methods**

### 239 2.1 Participants

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

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

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

258 In total, twenty-four dysphoric participants and twenty-eight control participants (age range  
259 18–40 years) volunteered for the study. Seven participants were excluded from the dysphoric  
260 group and another seven participants were excluded from the control group according to the  
261 exclusion criteria or because of extensive ocular artefacts in their EEG data. Therefore, 17  
262 dysphoric participants (3 male, 14 female) and 21 control participants (6 male, 15 female)  
263 were included in the final sample. Sixteen out of 17 dysphoric participants reported having an  
264 existing diagnosis of depression. One of them had been diagnosed with a mixed anxiety and  
265 depressive disorder. Demographics and clinical information of the dysphoric and control  
266 group are presented in Table 1.

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

## 272 2.2 Procedure

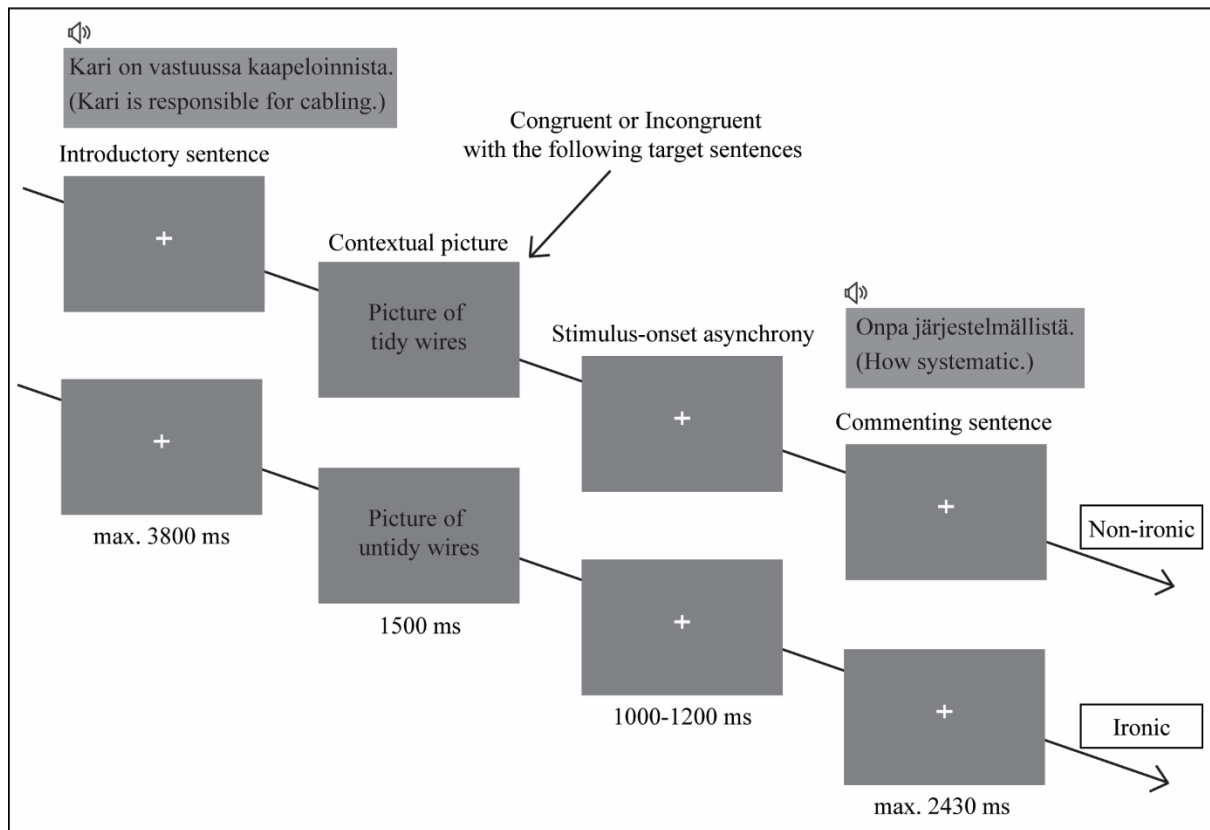
273 Psychophysiological and behavioral measurements for each participant were conducted on  
274 two separate days.

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

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

### 289 2.3 Stimuli

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



299

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

303 The presentation of each experimental trial is illustrated in Figure 1. First, an introductory  
 304 sentence was presented as a setup for the conversation. Next, a contextual picture was  
 305 displayed on the center of the screen. Then, a fixation mark appeared to the screen, after  
 306 which a spoken commenting sentence was presented together with the fixation mark. The  
 307 commenting sentence was spoken by a different person than the introductory sentence  
 308 providing a comment to the introductory sentence. The same conversations were presented  
 309 twice during the experiment: with a picture which provided a congruent context to the  
 310 commenting sentence (non-ironic condition) and with a picture which provided an  
 311 incongruent context to the commenting sentence (ironic condition). The inter-trial interval  
 312 (from the offset of the last word in the commenting sentence to the onset of the first word in  
 313 the introductory sentence) was randomized between 4000–4200 ms. Identities of the four

314 speakers in the audio files changed trial-by-trial, and the keywords were allocated at different  
315 positions of the commenting sentences making the sentences more variable and less  
316 predictable.

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

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

327 Therefore, 90 conversations with two contrasting pictures (90 trials for ironic and non-ironic  
328 conditions, respectively) were applied in the EMG and EEG measurement. They were  
329 divided into four blocks of 45 trials. Stimulus presentation order was pseudorandom: the  
330 ironic and non-ironic stimuli were presented in a random order, but to avoid immediate  
331 repetitions, the same conversations were separated with at least 45 other conversations. In  
332 order to keep the measurement time reasonable and participants better focused on the task at  
333 hand, for the measurement of the behavioral responses participants were presented with a half  
334 of the same stimuli they were presented in the EMG and EEG measurement: a half of the  
335 participants rated the first two stimulus blocks, and the other half rated the last two stimulus  
336 blocks applied in the EMG and EEG measurement. For all the measurements, the



337 presentation of the stimuli was controlled by E-Prime 2.0 software (Psychology Software  
338 Tools, Inc., Sharpsburg, PA, USA).

## 339 2.4 Facial EMG and EEG data

### 340 2.4.1 Measurement of facial EMG and EEG data

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

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

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

### 352 2.4.2 Analysis and statistics for facial EMG data

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

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

364 In order to reduce the levels of the time windows for analysis of variance (ANOVA; Akechi  
365 et al., 2013), facial EMG amplitude was averaged over a 1200-ms time window. The  
366 percentage of mean amplitude relative to the baseline was calculated for three intervals: 0–  
367 1200 ms (Time 1), 1200–2400 ms (Time 2) and 2400–3600 ms (Time 3) after the onset of the  
368 keyword. The values for facial EMG were analyzed separately for facial reactivity in the  
369 zygomaticus major muscle and in the corrugator supercilii muscle by using three-way  
370 repeated measures ANOVAs with Stimulus type (non-ironic vs. ironic) and Time window (1  
371 vs. 2 vs. 3) as within-subject factors and Group (control vs. dysphoric) as a between-subject  
372 factor.

### 373 2.4.3 Analysis and statistics for EEG data

374 EEG data were analyzed by using Brain Vision Analyzer 2.1 (Brain Products GmbH,  
375 Munich, Germany). A new reference was calculated offline as an average over all channels.  
376 The Gratton and Coles method (Gratton et al., 1983) was applied to detect and correct the  
377 interference of eye-blinks, and Channel 8, which is above the midpoint of the right eye, was  
378 chosen as the vertical electro-oculogram (VEOG) channel. Channels with an extensive  
379 amount of noise were interpolated using a spherical spline model. The EEG signal was  
380 filtered with a low cut-off of 0.1 Hz and a high cut-off of 20 Hz, accompanied by a 24  
381 dB/octave roll-off. A 50 Hz notch filter was also applied. After this, EEG data were extracted  
382 into 1100-ms long segments relative to the onset of keywords, starting from 100 ms before

383 the presentation of keywords. To exclude noisy segments, gradient criterion and max–min  
384 amplitude criterion were both applied. Specifically, the maximum allowed difference in  
385 amplitude between two consecutive time points was 75  $\mu\text{V}$ , and the maximum allowed  
386 voltage was  $-150 \mu\text{V}$  and  $150 \mu\text{V}$  at an interval of 200 ms for all the channels. Segments  
387 exceeding the specified values were omitted from the averages. The mean voltage during a  
388 period of 100 ms prior to the onset of the keyword was used as a baseline for each segment.  
389 Data were averaged separately for the two stimulus types (ironic vs. non-ironic) and for each  
390 participant.

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

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

399 The selection of the electrodes for the analysis was based on cluster-based permutation  
400 statistics (Maris and Oostenveld, 2007). This data-driven method was used as a tool to restrict  
401 the number of electrodes for the subsequent analyses conducted with repeated measures of  
402 ANOVAs (see also, e.g., Hämäläinen et al., 2018; Strömmer et al., 2017). This procedure is  
403 described in Supplementary materials. According to the results of the permutation statistics,  
404 the N400-like response was analyzed from one region of interest (ROI) (the electrodes 5, 6, 7,  
405 12, 31, 80, 106 and 112), and the P600 response was analyzed from two ROIs (Left ROI: the

406 electrodes 50, 51, 58 and 59; Right ROI: the electrodes 91, 96, 97 and 102). Figure 3 and  
407 Supplementary Figure 1 depict the electrode locations.

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

## 420 2.5 Behavioral responses

### 421 2.5.1 Measurement of behavioral responses

422 During the behavioral testing, participants were instructed to keep both hands on the  
423 keyboard (standard Finnish QWERTY keyboard), and press buttons with their index fingers.  
424 Response buttons (C, V, B, N, M in the keyboard) were labelled and covered by stickers with  
425 numbers (-2, -1, 0, 1, 2). Participants were first asked to evaluate as quickly as possible  
426 whether the picture and the spoken commenting sentence after it were congruent or  
427 incongruent. Then participants were asked to evaluate the valence of the conversations by  
428 using a scaling of -2 to 2 by pressing buttons that were labelled: hyvin ikävä (very

429 unpleasant) = -2; ikävä (somewhat unpleasant) = -1; siltä väliltä (between unpleasant and  
430 funny) = 0; jokseenkin hauska (somewhat funny) = 1; hyvin hauska (very funny) = 2. The  
431 valence rating was thus designed to measure especially perceived funniness of the stimuli and  
432 the opposite negative aspect. “Ikävä”, the word chosen to be the opposite (the other end of the  
433 rating scale) for funny can imply that there is a certain dislikeable aspect. Even if the  
434 promptness of the responding was requested, there was no time limit for responding. Before  
435 the actual evaluation task, the participants practiced with six rehearsal stimuli, and the  
436 experimenter confirmed that participants understood the task correctly.

#### 437 2.5.1 Analysis and statistics for behavioral data

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

446 The response times for congruency detection were calculated for trials with correct responses  
447 from the onset of the keywords. In the given example (Figure 1), “systemaattista”  
448 (“systematic”) was defined as the keyword of the commenting sentence. For each participant,  
449 response times above 2.5 standard deviations from the mean response time were excluded  
450 from the analyses. As a result of the trimming procedures, approximately 2.7% of all trials for  
451 all participants were lost for the analysis of response times. Each participant had more than  
452 32 trials for each stimulus type (ironic or non-ironic) for the reaction time measure.

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

## 458 2.6 General information on statistics

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

471 Pearson's correlation coefficients (two-tailed, computed at subject level) were applied to  
472 examine the relationships between the comprehension-related variables (the N400-like  
473 activity and the P600, and accuracy of congruency detection, respectively), and between the  
474 emotion-related variables (facial EMG corrugator and zygomaticus, and valence ratings of  
475 conversation, respectively). The correlations were calculated for ironic and non-ironic

476 stimulus conditions separately. Multiple correlations were controlled by applying false  
477 discovery rate (Benjamini and Yekutieli, 2001) at 0.050.

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

497 In addition to ANOVA/ANCOVA analyses, effect of depressive symptoms on responses  
498 were investigated with simple linear regression analyses. A simple linear regression was  
499 calculated for the whole sample with the amount of depressive symptoms (BDI-II scores) as a

500 predictor of ERPs, facial EMG amplitudes, and behavioral evaluations separately. Results of  
501 these analyses are reported in Supplementary materials.

## 502 **3 Results**

### 503 3.1 Behavioral results

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

#### 506 3.1.1 Accuracy of congruency detection

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

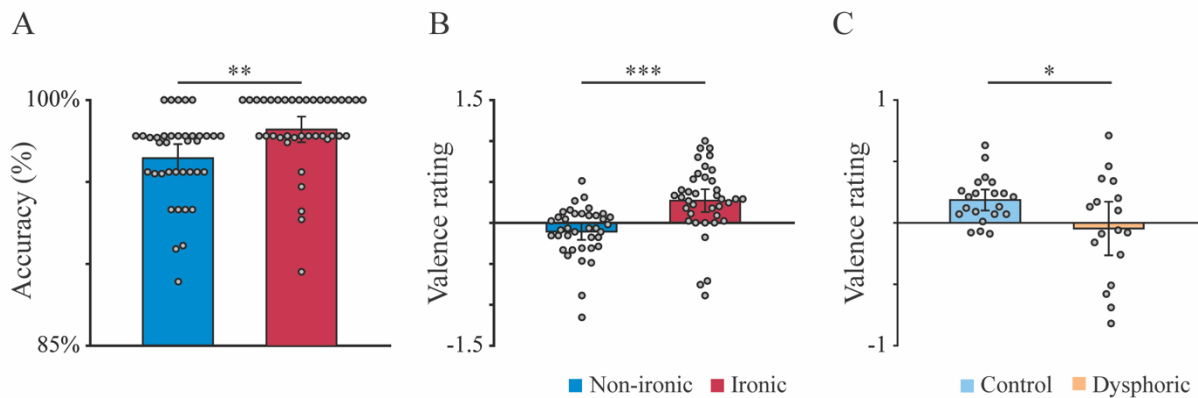
#### 512 3.1.2 Response time of congruency detection

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



518 3.1.3 Valence ratings of conversations

519 For the valence ratings of ironic and non-ironic conversations, repeated measures ANOVA  
520 revealed a significant main effect of Stimulus type,  $F(1,36) = 49.404, p < .001, \eta_p^2 = .578$ .  
521 Ironic conversations ( $M = 0.27, SD = 0.42$ ) were rated as funnier than non-ironic  
522 conversations ( $M = -0.11, SD = 0.31$ ). The main effect of group was also significant,  $F(1,36)$   
523  $= 5.098, p = .030, \eta_p^2 = .124$ . Valence ratings were more negative in the dysphoric group ( $M =$   
524  $-0.05, SD = 0.42$ ) than in the control group ( $M = 0.19, SD = 0.19$ ) over the stimulus types.  
525 Interaction effect of Stimulus type  $\times$  Group was non-significant,  $p = .112$ .



526

527 **Figure 2.** Behavioral responses on the second day's measurement. (A) Mean accuracy of congruency detection  
528 and 95% confidence intervals to non-ironic (blue) and ironic (red) conversations. Values of individual responses  
529 are presented as a scatterplot.  $**p < .01$ . (B) Mean values of valence ratings and 95% confidence intervals to  
530 non-ironic (blue) and ironic (red) conversations. Individual responses are presented as a scatterplot.  $***p < .001$ .  
531 (C) Mean values of valence ratings and 95% confidence intervals in control (light blue) and dysphoric (light  
532 orange) groups (averaged over non-ironic and ironic). Individual responses are presented as a scatterplot.  $*p$   
533  $< .05$ . The scale for valence ratings ranged from -2 (very unpleasant) to 2 (very funny), with 0 meaning neutral.  
534 Please note that the scale is different for B and C.

535 3.2 Facial EMG responses results

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

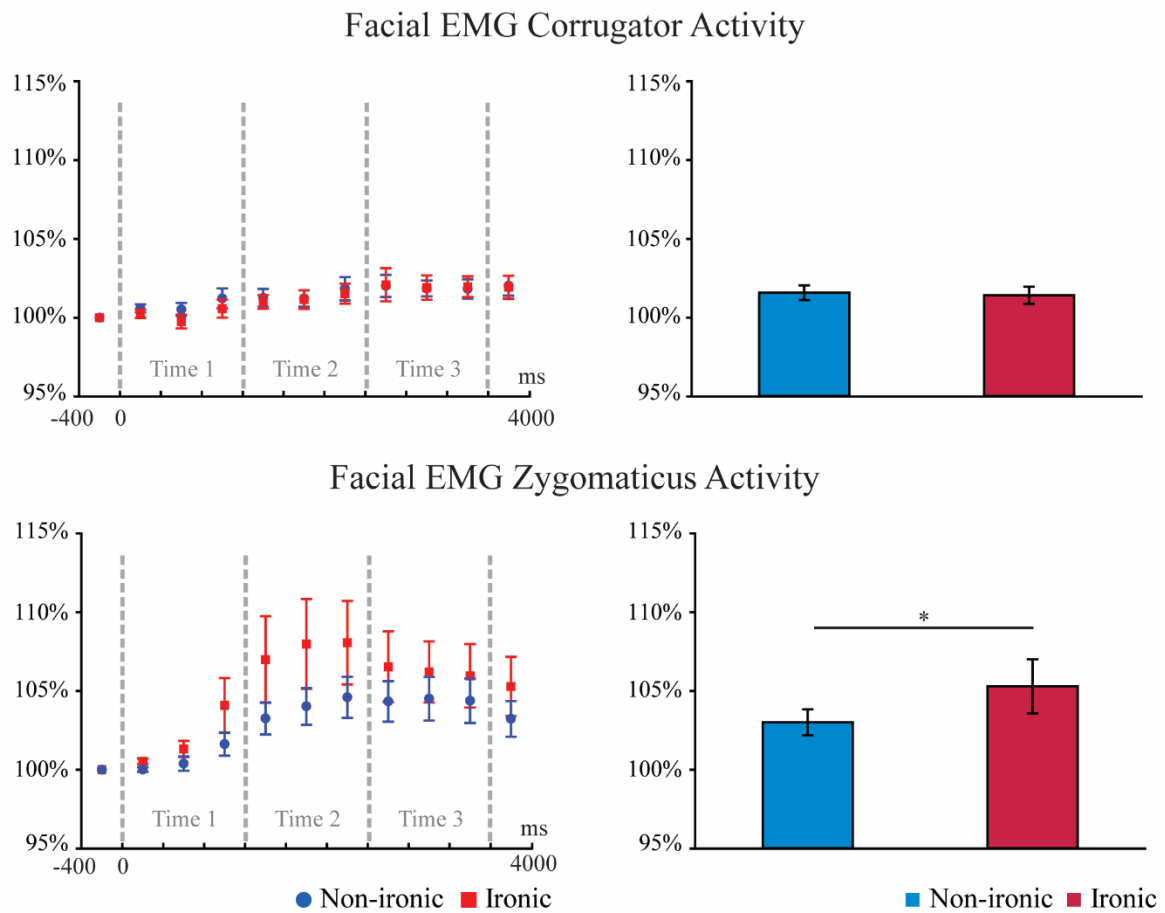
### 539 3.2.1 Facial EMG corrugator activity

540 For facial EMG corrugator activity, three-way repeated measures ANOVA with within-  
541 subject variables Stimulus type (non-ironic vs. ironic) and Time window (1 vs. 2 vs. 3), and  
542 between-subject variable Group (control vs. dysphoric) revealed that the main effect of Time  
543 window was significant,  $p = .009$ . The corrugator activity was smaller between 0–1200 ms  
544 ( $M = 100.4\%$ ,  $SD = .022$ ), compared with the activity in the period of 1200–2400 ms ( $M =$   
545  $101.3\%$ ,  $SD = .032$ ),  $t(37) = 2.599$ ,  $p = .013$ ,  $d = 0.327$ , and with the activity during 2400–  
546 3600 ms ( $M = 102\%$ ,  $SD = .040$ ),  $t(37) = 2.738$ ,  $p = .009$ ,  $d = 0.495$ . The difference between  
547 the activity in the period of 1200–2400 ms and 2400–3600 ms was not found,  $p = .051$ . Other  
548 main effects or interactions were not found for corrugator responses (all  $p$ -values  $> .312$ ).

### 549 3.2.2 Facial EMG zygomaticus activity

550 For facial EMG zygomaticus activity, three-way repeated measures ANOVA showed that the  
551 main effect of Time window was significant,  $p = .002$ . The zygomaticus activity was larger in  
552 the time window of 1200–2400 ms ( $M = 105.8\%$ ,  $SD = .115$ ) and 2400–3600 ms ( $M =$   
553  $105.3\%$ ,  $SD = .098$ ), compared with the activity between 0–1200 ms ( $M = 101.3\%$ ,  $SD$   
554  $= .033$ ),  $t(37) = 3.082$ ,  $p = .004$ ,  $d = 0.532$ ;  $t(37) = 2.739$ ,  $p = .009$ ,  $d = 0.547$ , respectively.  
555 There was no significant difference between the activity in the period of 1200–2400 ms and  
556 2400–3600 ms,  $p = .674$ . The main effect of Stimulus type was significant,  $p = .039$ . The  
557 facial EMG zygomaticus amplitude was larger after the ironic stimuli ( $M = 105.3\%$ ,  $SD$

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



560

561 **Figure 3.** The grand-averages of the mean EMG amplitude percentage relative to the baseline for the corrugator  
 562 (upper) and for the zygomaticus (lower) to non-ironic (blue) and ironic (red) stimulus. Left: Average percentage  
 563 values and standard error of mean (SE) are presented for descriptive purposes in 400-ms segments, and the grey  
 564 dotted lines show the segments applied in the ANOVA (Time 1: 0–1200 ms; Time 2: 1200–2400 ms; Time 3:  
 565 2400–3600 ms). Right: The grand-averages of the mean values and SD for non-ironic and ironic responses  
 566 averaged over the time segments reflect the main effect of Stimulus type ( $*p < .05$ ) for the zygomaticus. The  
 567 responses are averaged over the two groups, since there were no group differences (see Supplementary Figure 3  
 568 for the responses in each group).

569 3.3 ERPs results

570 ERPs results are presented in Figure 4 (N400-like activity) and Figure 5 (P600). Grand-

571 averaged responses showed larger responses for the ironic than the non-ironic stimuli at 300–

572 500 ms and 500–800 ms after the onset of the keyword reflecting N400-like activity and  
573 P600, respectively. Results of repeated measures ANOVAs for N400-like activity and P600  
574 activity are reported in Table 3.

### 575 3.3.1 N400-like activity

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

### 580 3.3.2 P600

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

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

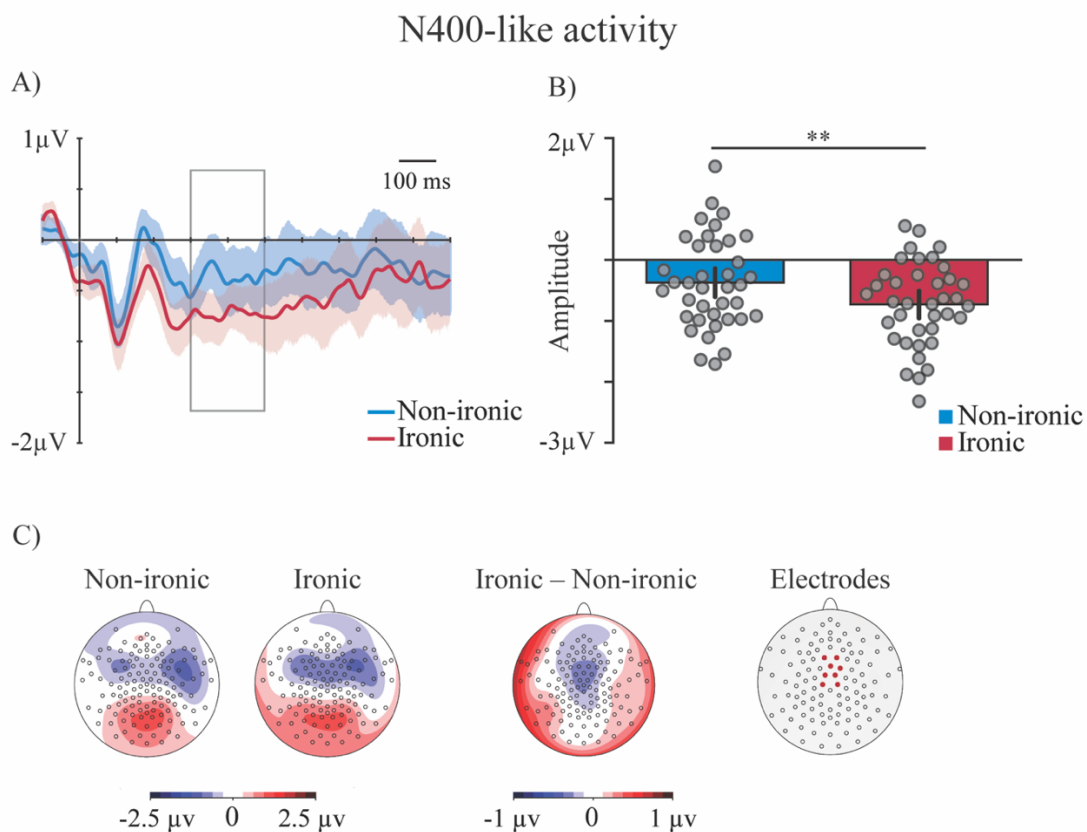
589 First, follow-up two-way repeated measures ANOVAs with within-subjects variables  
590 Stimulus type (non-ironic vs. ironic) and ROI (left vs. right) were conducted separately for  
591 the control and the dysphoric group. In the control group, the analysis showed a significant  
592 main effect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

594 0.65) stimuli. There were no other main or interaction effects in the control group, all  $p$ -  
595 values  $> .257$ . In the dysphoric group, a main effect of Stimulus type was observed,  $F(1,16) =$   
596  $22.864$ ,  $p < .001$ ,  $\eta_p^2 = .588$ . The amplitude of the P600 response was larger for ironic ( $M =$   
597  $1.17 \mu\text{V}$ ,  $SD = 0.65$ ) than non-ironic ( $M = 0.61 \mu\text{V}$ ,  $SD = 0.57$ ) stimuli. The main effect of  
598 ROI was not significant,  $p > .297$ . A marginally significant Stimulus type  $\times$  ROI interaction  
599 effect,  $F(1,16) = 3.912$ ,  $p = .065$ ,  $\eta_p^2 = .196$ , was found in the dysphoric group. Paired  $t$ -tests  
600 revealed that, in the right ROI, P600 amplitude was larger to ironic than to non-ironic stimuli  
601 in the dysphoric group,  $t(16) = 4.453$ ,  $p < .001$ ,  $d = 0.832$ . In the left ROI, there was no  
602 significant difference between the responses to ironic and non-ironic stimuli,  $p = .139$ .

603 Second, follow-up two-way repeated measures ANOVAs with a within-subjects variable ROI  
604 (left vs. right) and a between-subjects variable Group (control vs. dysphoric) were conducted  
605 for the ironic and non-ironic stimuli separately. For the non-ironic stimuli, there were neither  
606 a main effect of ROI, a main effect of Group nor their interaction effect, all  $p$ -values  $> .340$ .  
607 For the ironic stimuli, the main effect of ROI and the main effect of Group were non-  
608 significant, both  $p$ -values  $> .776$ . There was a significant interaction effect of ROI  $\times$  Group,  
609  $F(1,36) = 6.978$ ,  $p = .012$ ,  $\eta_p^2 = .162$ , however. Post hoc tests based on independent samples  
610  $t$ -tests comparing the groups in P600 amplitude to ironic stimuli separately at the left and at  
611 the right ROI indicated no group differences, both  $p$ -values  $> .135$ . Paired samples  $t$ -tests  
612 exploring the amplitude difference to ironic stimuli between the left and the right ROIs within  
613 each group were implemented. The analysis showed that the P600 amplitude to ironic stimuli  
614 was larger on the right ROI than on the left ROI in the dysphoric group,  $t(16) = 2.399$ ,  $p$   
615  $= .029$ ,  $d = 0.674$ , but there was no difference in the control group,  $p = .132$ .

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

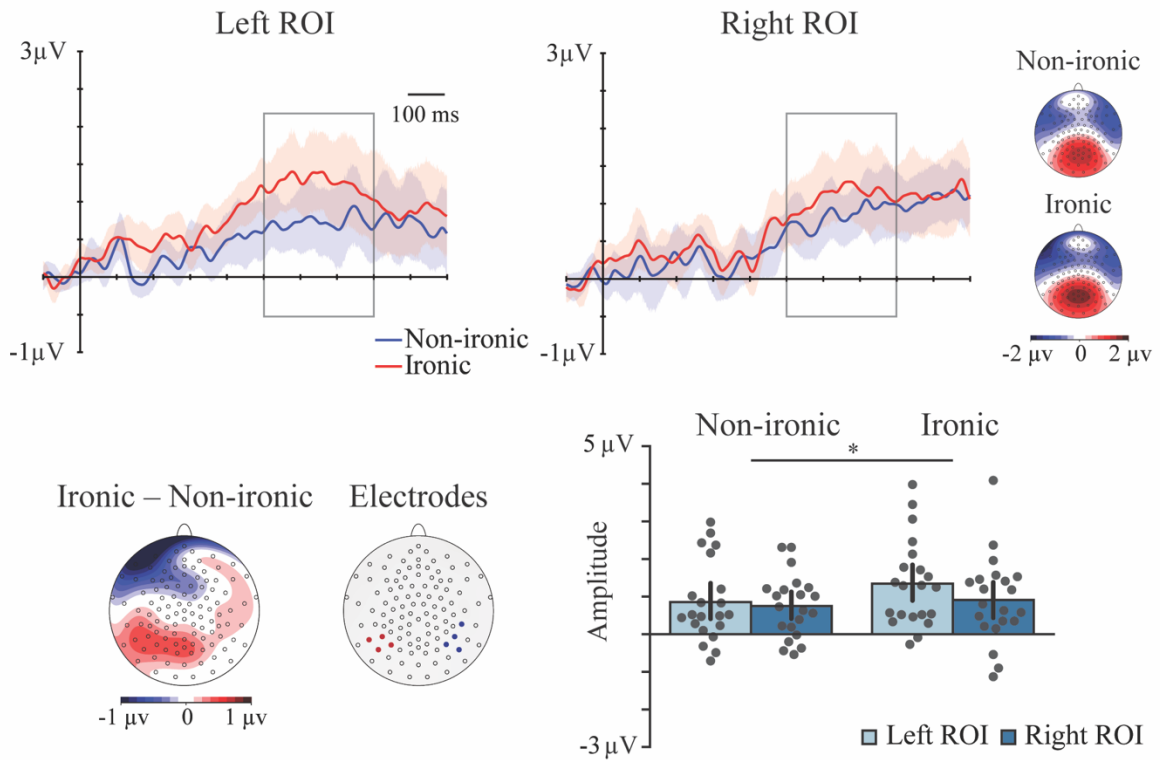
618 Values of three factors of the cognitive tests (executive function, list memory, semantic  
 619 processing) were added as covariates independently in the ANCOVA model for P600  
 620 responses. Only list memory as a covariate in the ANCOVA changed the original results: the  
 621 three-way interaction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



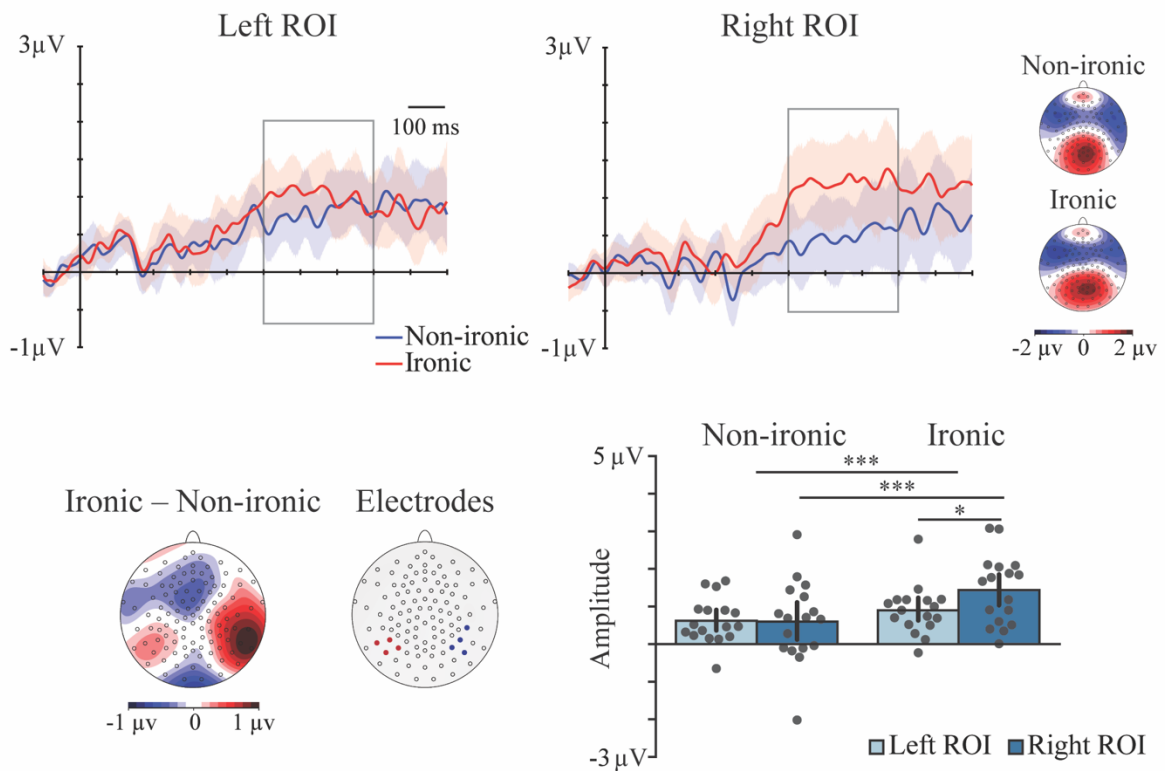
622  
 623 **Figure 4.** Grand-averaged waveforms, mean amplitude values, topographical maps, and selections of electrodes  
 624 for N400-like activity. A) Grand-averaged waveforms to non-ironic and ironic stimuli are presented. The blue  
 625 and red shadows in the waveforms represent 95% confidence intervals. The grey rectangle shows the time  
 626 window applied in the analysis for N400-like activity (mean amplitude values between 300–500 ms). B) Mean  
 627 amplitudes for N400-like activity to non-ironic (blue) and ironic (red) stimulus are illustrated. The error bars  
 628 represent 95% confidence intervals.  $**p < .01$ , a main effect of Stimulus type. The scatterplots overlaid with the  
 629 histograms represent individual participants' amplitudes in each condition. C) Grand-averaged topographical  
 630 maps for non-ironic and ironic stimuli, and a difference between the two stimulus types in topographies and  
 631 selections of electrodes (marked with red) for N400-like activity are shown. The responses are averaged over the  
 632 two groups, since there were no group differences (see Supplementary Figure 2 for the responses in each group).

# P600 responses

## CONTROL



## DYSPHORIC



634 **Figure 5.** Grand-averaged ERP waveforms, mean amplitude values, topographical maps, and selections of  
635 electrodes for P600 on left and right ROIs separately for the control and the dysphoric group. The blue and red  
636 shadows in the waveforms represent 95% confidence intervals. The grey rectangles in waveform figures show  
637 the time window applied in the analysis for P600 (mean amplitude values between 500–800 ms). The red dots  
638 and the blue dots represent the electrode cluster on the left and right ROI, respectively. The histograms represent  
639 mean amplitudes of P600 to non-ironic and ironic stimulus on the left and right ROIs. The error bars represent  
640 95% confidence intervals ( $*p < .05$ ;  $***p < .001$ , post hoc paired t-tests investigating Stimulus type  $\times$  ROI  $\times$   
641 Group interaction). The scatterplots overlaid with the histograms represent individual participants' amplitudes in  
642 each condition.

### 643 3.4 Correlations

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

## 649 4 Discussion

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

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



659 cognitive processing of irony. A difference in the P600 responses was found between the  
660 dysphoric group and the control group. Next, the results are discussed in details.

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

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

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

687 There were no differences in the facial EMG responses between the dysphoric group and the  
688 control group. This finding was surprising, taking into account the well-documented  
689 alterations in emotional information processing in depression (e.g., Gotlib and Joormann,  
690 2010; Leppänen, 2006) and the reduced tendency to react with exhilaration to funny stimuli  
691 (Uekermann et al., 2008; Falkenberg et. al., 2011). It is possible that the two groups did not  
692 differ in their facial EMG responses because similar to the dysphoric group, the ironic stimuli  
693 were not found to be very funny in the control group, although they were found to be  
694 relatively more funny than the non-ironic conversations at the whole sample level. Another  
695 possible reason for the lack of group difference could be the large variance we observed  
696 especially in the valence ratings and zygomaticus facial EMG for the dysphoric group. The  
697 variance can be expected within dysphoric participants due to the heterogeneity of depression  
698 related to different symptom profiles (e.g., more or less vegetative vs. affective symptoms),  
699 differences in amount of symptoms or diagnosis type (e.g. atypical depression vs.  
700 melancholic depression).

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

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

712 The N400-like activity was more negative in polarity in response to the ironic stimuli than in  
713 response to the non-ironic stimuli. Unlike traditional N400 (e.g., Kutas and Federmeier,  
714 2011), the response did not seem to have a clear peak (see also, Filik et al., 2014) and was  
715 slightly frontally-distributed. Here, we applied naturally spoken sentences in which the  
716 position of the keyword varied between the sentences. This may have affected the  
717 morphology and the topography of the N400 response (Kutas and Federmeier, 2011), which  
718 is usually measured in a condition where the keyword is in the end of the sentence (e.g., Filik  
719 et al., 2014; Regel et al., 2011, 2014).

720 The observed N400-like activity to irony is in line with previous findings of N400 elicited to  
721 irony (Cornejo et al., 2007; Filik et al., 2014; Caillies et al., 2019), humor (Coulson and  
722 Kutas, 2001; Coulson and Wu, 2005), and other nonliteral language (metaphor: Coulson and  
723 Van Petten, 2002) in showing enhanced response amplitude to nonliteral stimuli (here ironic)  
724 in comparison to literal stimuli (here non-ironic). This finding may suggest that the  
725 participants had difficulty processing the meaning of the commenting sentence in the context  
726 of the opposing contextual picture. This difficulty may stem from the holistic strategy the  
727 participants applied during comprehension (Cornejo et al., 2007) or the stimuli, which were  
728 unfamiliar to the participants, because a previous study only found N400-like modulation for  
729 unfamiliar ironic utterances (Filik et al., 2014). Some studies have not found N400

730 modulation for irony, however (Balconi and Amenta, 2008; Regel et al., 2011, 2014). Regel  
731 et al. (2011, 2014) reported that the lack of N400 modulation to ironic stimuli means that the  
732 participants were capable of comprehending irony based on the supportive context, with no  
733 difficulty in semantic interpretation. In our study, even though we did not measure the  
734 difficulty of semantic interpretation directly with behavioral tests, the results of congruence  
735 detection showed that the participants were very accurate in categorizing both congruent (i.e.,  
736 non-ironic) and incongruent (i.e., ironic) picture-sentence pairs, and we found the amplitude  
737 of the N400 was still modulated by the irony. However, the results regarding the N400 effect  
738 should be interpreted with caution, because the effect size of the irony effect is small and as  
739 mentioned earlier, it is different in morphology from traditional N400 effects.

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

745 As expected, P600, which displayed a centro-parietal distribution, was also modulated by  
746 irony in our study. The amplitude of P600 was more positive for the ironic punchlines than  
747 the non-ironic punchlines; this result has also been reported in previous studies on irony  
748 comprehension (Baptista et al., 2018; Filik et al., 2014; Regel et al., 2011, 2014; Caillies et  
749 al., 2019). Visual observation of the grand-averaged difference topographies (ironic minus  
750 non-ironic) for the P600 responses in Filik et al. (2014) and those in the healthy controls in  
751 our study shows a remarkable similarity: they both show irony-related activity in the central  
752 and in the left parietal electrode sites. Moreover, the results in our study and in previous

753 studies are similar in that no differences were found in irony modulation between the left and  
754 right ROIs in the healthy controls (Filik et al., 2014; Regel et al., 2011, 2014).

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

768 Here, the participants did not engage in any task related to the stimuli during the EEG/EMG-  
769 measurement, and the behavioral evaluations were conducted on separate days. Therefore, it  
770 is unlikely that the present study's finding of enhanced P600 in relation to irony is elicited by  
771 the requirements of the comprehension tasks or categorization tasks, which were possible  
772 reasons for the P600 modulation in studies by Regel et al. (2011) and Filik et al. (2014),  
773 respectively. Regel et al. (2011) also suggested that an increased P600 amplitude to irony  
774 may reflect the processing of emotional information expressed by ironic utterances. Partly  
775 supporting this idea, a recent study found that greater P600 modulation was observed in  
776 relation to ironic criticism than ironic praise (Caillies et al., 2019). In the present study, we

777 had positive keywords in half of the sentences and negative keywords in another half of the  
778 sentences. Due to the limited number of trials in each sub-type (ironic criticism or ironic  
779 praise), we were unable to analyze the possible effects related to the different sub-types.  
780 However, we found no correlations between the amplitude of P600 and valence ratings of the  
781 stimuli. Thus, it seems an unlikely explanation that, in the present study, the modulation of  
782 P600 is due to the processing of emotional information. Consequently, we considered that the  
783 larger amplitude of P600 likely reflects the greater inferential effort required for the  
784 resolution of the ironic punchlines than for the non-ironic punchlines arising from the conflict  
785 between the meaning of the keyword and the contextual picture in the irony trials.

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

797 In addition to the analysis investigating group differences categorically, we also calculated  
798 simple linear regressions for P600 (separately for the left and right ROI), facial EMG  
799 activities, and valence ratings with BDI-II scores as a predictor. However, the regression  
800 analyses did not show any significant effects (see Supplementary materials). One reason for

801 this result could be that self-assessment questionnaires, such as the BDI-II, are not the most  
802 accurate measures of depressive symptoms because some depressed individuals do not have a  
803 clear awareness of their symptoms; thus, they can inaccurately estimate their symptoms,  
804 which in turn can lead to non-significant linear regressions between amount of symptoms and  
805 responses to irony.

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

822 Several limitations in the present study are worth noting. The dysphoric participants self-  
823 reported their diagnostic status; not all of them had been recently diagnosed with depression.  
824 However, BDI-II scores were used to measure the depressive symptoms at the time of the

825 measurement, and all the participants in the dysphoric group had scores of 14 or more,  
826 reflecting at least mild depression (Beck et al., 1996). Still, our results may not be generalized  
827 to clinical depression. Moreover, there were more female participants than male participants  
828 in both the dysphoric and the control groups. Therefore, our results cannot unconditionally be  
829 generalized to both genders. However, the proportion of gender distribution in the dysphoric  
830 and control groups was similar. Furthermore, ten dysphoric participants were taking  
831 antidepressant medication while participating in the study, but analyses using the medication  
832 status as a covariate did not reveal any interactions between the medication status and the  
833 dependent variables. It is also notable that our stimuli included both ironic praise and ironic  
834 criticism, and as mentioned before, we were not able to analyze the effects separately for  
835 each sub-type. It is possible that valence ratings could have been more positive, at least in the  
836 control group, if only ironic criticism had been used (Dews et al., 1995; Thompson et al.,  
837 2016), and this could have led to group differences also in facial EMG responses. Last, the  
838 ERP analysis was based on sensor-level analysis, and we did not utilize any source  
839 localization methods. Therefore, we cannot accurately estimate the sources of the activity in  
840 the two groups. Future studies should confirm our findings with a larger sample, and also  
841 investigate the sources of the brain activity related to irony comprehension.

842 One advantage of the present study is its design in which the non-ironic and ironic conditions  
843 were defined by the previous contextual picture, which provided either a non-ironic or ironic  
844 context for the commenting sentence. This arrangement allows for a valid comparison of the  
845 non-ironic and ironic conversations irrespective of the low-level stimulus features, the  
846 position of the keywords, or other potentially confounding factors because the comparison  
847 was always made between sentences that were physically identical. A definite strength of this  
848 study is also its use of multimodal recordings and two measurement sessions conducted on  
849 separate days. Namely, the ratings of congruency recognition were collected separately after



850 the EEG and EMG measurements. This allowed the participants to focus on the stimuli  
851 during the EEG/EMG recordings without responding.

## 852 **5 Conclusions**

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

## 867 **6 Acknowledgments**

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

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

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

876 7 References

- 877 Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013).  
878 Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative  
879 Ratings. *PLoS ONE*, 8(3), e59312. <https://doi.org/10.1371/journal.pone.0059312>
- 880 Akimoto, Y., Sugiura, M., Yomogida, Y., Miyauchi, C. M., Miyazawa, S., & Kawashima, R.  
881 (2014). Irony comprehension: Social conceptual knowledge and emotional response.  
882 *Human Brain Mapping*, 35(4), 1167–1178. <https://doi.org/10.1002/hbm.22242>
- 883 Allen, N. B., Trinder, J., & Brennan, C. (1999). Affective startle modulation in clinical  
884 depression: Preliminary findings. *Biological Psychiatry*, 46(4), 542–550.  
885 [https://doi.org/10.1016/S0006-3223\(99\)00025-6](https://doi.org/10.1016/S0006-3223(99)00025-6)
- 886 Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, 32(6), 793–  
887 826. [https://doi.org/10.1016/S0378-2166\(99\)00070-3](https://doi.org/10.1016/S0378-2166(99)00070-3)
- 888 Balconi, M., & Amenta, S. (2008). Isn't it Ironic? An Analysis on the Elaboration of Ironic  
889 Sentences with ERPs. *The Open Applied Linguistics Journal*, 1(1), 9–17.  
890 <https://doi.org/10.2174/1874913500801010009>
- 891 Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling  
892 metaphor from context: An ERP study. *Frontiers in Psychology*, 7, 559.  
893 <https://doi.org/10.3389/fpsyg.2016.00559>
- 894 Baptista, N. I., Manfredi, M., & Boggio, P. S. (2018). Medial prefrontal cortex stimulation  
895 modulates irony processing as indexed by the N400. *Social Neuroscience*, 13(4), 495–  
896 510. <https://doi.org/10.1080/17470919.2017.1356744>
- 897 Basso, M. R., & Bornstein, R. A. (1999). Relative memory deficits in recurrent versus first-  
898 episode major depression on a word-list learning task. *Neuropsychology*, 13(4), 557.  
899 <https://doi.org/10.1037/0894-4105.13.4.557>
- 900 Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. Retrieved  
901 from <https://books.google.fi/books?id=6rigrtdo0u2UC>
- 902 Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Manual for the Beck depression inventory-  
903 II. *San Antonio, TX: Psychological Corporation*, 1–82.
- 904 Beck, A. T. (2008). The evolution of the cognitive model of depression and its  
905 neurobiological correlates. *American Journal of Psychiatry*, 165(8), 969–977.  
906 <https://doi.org/10.1176/appi.ajp.2008.08050721>
- 907 Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple  
908 testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188.  
909 <https://doi.org/10.1214/aos/1013699998>
- 910 Bora, E., & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis.  
911 *Journal of Affective Disorders*, 191, 49–55. <https://doi.org/10.1016/j.jad.2015.11.023>

- 912 Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on  
 913 “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1),  
 914 55–73. <https://doi.org/10.1016/J.BRAINRESREV.2008.05.003>
- 915 Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking  
 916 the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–  
 917 143. <https://doi.org/10.1016/j.brainres.2012.01.055>
- 918 Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional  
 919 reactivity in major depressive disorder. *Clinical Psychology Review*, *28*(4), 676–691.  
 920 <https://doi.org/10.1016/J.CPR.2007.10.001>
- 921 Caillies, S., Gobin, P., Obert, A., Terrien, S., Coutté, A., Iakimova, G., & Besche-Richard, C.  
 922 (2019). Asymmetry of affect in verbal irony understanding: What about the N400 and  
 923 P600 components? *Journal of Neurolinguistics*, *51*, 268–277.  
 924 <https://doi.org/10.1016/j.jneuroling.2019.04.004>
- 925 Calmus, A., & Caillies, S. (2014). Verbal irony processing: How do contrast and humour  
 926 correlate? *International Journal of Psychology*, *49*(1), 46–50.  
 927 <https://doi.org/10.1002/ijop.12003>
- 928 Canal, P., Bischetti, L., Di Paola, S., Bertini, C., Ricci, I., & Bambini, V. (2019). ‘Honey,  
 929 shall I change the baby? – Well done, choose another one’: ERP and time-frequency  
 930 correlates of humor processing. *Brain and Cognition*, *132*, 41–55.  
 931 <https://doi.org/10.1016/j.bandc.2019.02.001>
- 932 Canli, T., Sivers, H., Thomason, M. E., Whitfield-Gabrieli, S., Gabrieli, J. D., & Gotlib, I. H.  
 933 (2004). Brain activation to emotional words in depressed vs healthy subjects.  
 934 *Neuroreport*, *15*(17), 2585–2588.
- 935 Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York:  
 936 Routledge. <https://doi.org/10.4324/9780203771587>
- 937 Cornejo, C., Simonetti, F., Aldunate, N., Ibáñez, A., López, V., & Melloni, L. (2007).  
 938 Electrophysiological Evidence of Different Interpretative Strategies in Irony  
 939 Comprehension. *Journal of Psycholinguistic Research*, *36*(6), 411–430.  
 940 <https://doi.org/10.1007/s10936-007-9052-0>
- 941 Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in  
 942 good and poor comprehenders. *Neuroscience Letters*, *316*(2), 71–74.  
 943 [https://doi.org/10.1016/S0304-3940\(01\)02387-4](https://doi.org/10.1016/S0304-3940(01)02387-4)
- 944 Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related  
 945 potential study. *Memory & Cognition*, *30*(6), 958–968.  
 946 <https://doi.org/10.3758/BF03195780>
- 947 Coulson, S., & Wu, Y. C. (2005). *Right Hemisphere Activation of Joke-related Information:  
 948 An Event-related Brain Potential Study*. *Journal of cognitive neuroscience*, *17*(3), 494–  
 949 506.

- 950 Cusi, A. M., Nazarov, A., Holshausen, K., MacQueen, G. M., & McKinnon, M. C. (2012).  
 951 Systematic review of the neural basis of social cognition in patients with mood  
 952 disorders. *Journal of Psychiatry and Neuroscience*, 37(3), 154–169.  
 953 <https://doi.org/10.1503/jpn.100179>
- 954 Del Goletto, S., Kostova, M., & Blanchet, A. (2016). Impaired context processing during  
 955 irony comprehension in schizotypy: An ERPs study. *International Journal of*  
 956 *Psychophysiology*, 105, 17–25. <https://doi.org/10.1016/j.ijpsycho.2016.04.009>
- 957 Deldin, P., Keller, J., Casas, B. R., Best, J., Gergen, J., & Miller, G. A. (2006). Normal N400  
 958 in mood disorders. *Biological Psychology*, 71(1), 74–79.  
 959 <https://doi.org/10.1016/J.BIOPSYCHO.2005.02.005>
- 960 Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of  
 961 irony. *Discourse Processes*, 19(3), 347–367.  
 962 <https://doi.org/10.1080/01638539509544922>
- 963 Dimberg, U. (1990). Facial Electromyography and Emotional Reactions. *Psychophysiology*,  
 964 27(5), 481–494. <https://doi.org/10.1111/j.1469-8986.1990.tb01962.x>
- 965 Dunn, B. D., Dalgleish, T., Lawrence, A. D., Cusack, R., & Ogilvie, A. D. (2004).  
 966 Categorical and Dimensional Reports of Experienced Affect to Emotion-Inducing  
 967 Pictures in Depression. *Journal of Abnormal Psychology*, 113(4), 654–  
 968 660. <https://doi.org/10.1037/0021-843X.113.4.654>
- 969 Dynel, M. (2009). Beyond a Joke: Types of Conversational Humour. *Language and*  
 970 *Linguistics Compass*, 3(5), 1284–1299. <https://doi.org/10.1111/j.1749-818X.2009.00152.x>
- 972 Fiacconi, C. M., & Owen, A. M. (2015). Using Psychophysiological Measures to Examine  
 973 the Temporal Profile of Verbal Humor Elicitation. *Plos One*, 10(9), e0135902.  
 974 <https://doi.org/10.1371/journal.pone.0135902>
- 975 Falkenberg, I., Jarmuzek, J., Bartels, M., & Wild, B. (2011). Do depressed patients lose their  
 976 sense of humor? *Psychopathology*, 44(2), 98–105. <https://doi.org/10.1159/000317778>
- 977 Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical  
 978 power analysis program for the social, behavioral, and biomedical sciences. *Behavior*  
 979 *Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- 980 Feng, Y.-J., Chan, Y.-C., & Chen, H.-C. (2014). Specialization of neural mechanisms  
 981 underlying the three-stage model in humor processing: An ERP study. *Journal of*  
 982 *Neurolinguistics*, 32, 59–70. <https://doi.org/10.1016/j.jneuroling.2014.08.007>
- 983 Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony  
 984 processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning*  
 985 *Memory and Cognition*, 40(3), 811–828. <https://doi.org/10.1037/a0035658>
- 986 Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, 116(3),  
 987 421–436. <https://doi.org/10.1016/j.cognition.2010.06.005>

- 988 Fossati, P., Harvey, P. O., Le Bastard, G., Ergis, A. M., Jouvent, R., & Allilaire, J. F. (2004).  
989 Verbal memory performance of patients with a first depressive episode and patients with  
990 unipolar and bipolar recurrent depression. *Journal of Psychiatric Research*, 38(2), 137–  
991 144. <https://doi.org/10.1016/j.jpsychires.2003.08.002>
- 992 Gaudreau, G., Monetta, L., Macoir, J., Poulin, S., Laforce, R. J., & Hudon, C. (2015). Mental  
993 State Inferences Abilities Contribution to Verbal Irony Comprehension in Older Adults  
994 with Mild Cognitive Impairment. *Behavioural Neurology*, 2015, 1–9.  
995 <https://doi.org/10.1155/2015/685613>
- 996 Gibbs Jr., R. W. (2002). A new look at literal meaning in understanding what is said and  
997 implicated. *Journal of Pragmatics*, 34, 457–486.
- 998 Giora, R. (2003). On Our Mind Saliency, Context, and Figurative Language. In *On Our Mind:*  
999 *Saliency, Context, and Figurative Language*.  
1000 <https://doi.org/10.1093/acprof:oso/9780195136166.001.0001>
- 1001 Giora, R., & Fein, O. (1999). Irony comprehension: The graded saliency hypothesis. *Humor*,  
1002 12(4), 425–436. <https://doi.org/10.1515/humr.1999.12.4.425>
- 1003 Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Grade Saliency and Indirect Negation.  
1004 *Metaphor and Symbol*, 13(2), 83–101. [https://doi.org/10.1207/s15327868ms1302\\_1](https://doi.org/10.1207/s15327868ms1302_1)
- 1005 Good, P. I. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses* (Third). New  
1006 York: Springer.
- 1007 Gotlib, I. H., & Joormann, J. (2010). Cognition and Depression: Current Status and Future  
1008 Directions. *Annual Review of Clinical Psychology*, 6(1), 285–312.  
1009 <https://doi.org/10.1146/annurev.clinpsy.121208.131305>
- 1010 Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes  
1011 underlying the P600. *Language and Cognitive Processes*, 25(2), 149–188.  
1012 <https://doi.org/10.1080/01690960902965951>
- 1013 Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of  
1014 ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55(4), 468–484.  
1015 [https://doi.org/10.1016/0013-4694\(83\)90135-9](https://doi.org/10.1016/0013-4694(83)90135-9)
- 1016 Grice, H. P. (1975). Logic and conversation. In *Syntax and semantics 3: Speech acts* (pp. 41–  
1017 58). Brill.
- 1018 Hämäläinen, J., Landi, N., Loberg, O., Lohvansuu, K., Pugh, K., & Leppänen, P. H. T.  
1019 (2018). Brain event-related potentials to phoneme contrasts and their correlation to  
1020 reading skills in school-age children. *International Journal of Behavioral Development*,  
1021 42(3), 357–372. <https://doi.org/10.1177/0165025417728582>
- 1022 Henry, J. D., & Crawford, J. R. (2005). A meta-analytic review of verbal fluency deficits in  
1023 depression. *Journal of Clinical and Experimental Neuropsychology*, 27(1), 78–101.  
1024 <https://doi.org/10.1080/138033990513654>

- 1025 Herold, R., Varga, E., Hajnal, A., Hamvas, E., Berecz, H., Tóth, B., & Tényi, T. (2018).  
1026 Altered neural activity during irony comprehension in unaffected first-degree relatives  
1027 of schizophrenia patients - An fMRI study. *Frontiers in Psychology*, 8, 1–15.  
1028 <https://doi.org/10.3389/fpsyg.2017.02309>
- 1029 Hutcherson, C. A., Goldin, P. R., Ochsner, K. N., Gabrieli, J. D., Feldman Barrett, L., &  
1030 Gross, J. J. (2005). Attention and emotion: Does rating emotion alter neural responses to  
1031 amusing and sad films? *NeuroImage*, 27(3), 656–668.  
1032 <https://doi.org/10.1016/j.neuroimage.2005.04.028>
- 1033 Inoue, Y., Tonooka, Y., Yamada, K., & Kanba, S. (2004). Deficiency of theory of mind in  
1034 patients with remitted mood disorder. *Journal of Affective Disorders*, 82(3), 403–409.  
1035 <https://doi.org/10.1016/j.jad.2004.04.004>
- 1036 Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not)  
1037 and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.  
1038 <https://doi.org/10.1016/j.jml.2007.11.007>
- 1039 Kaakinen, J. K., Olkonemi, H., Kinnari, T., & Hyönä, J. (2014). Processing of Written Irony:  
1040 An Eye Movement Study. *Discourse Processes*, 51(4), 287–311.  
1041 <https://doi.org/10.1080/0163853X.2013.870024>
- 1042 Kalatzis, I., Piliouras, N., Ventouras, E., Papageorgiou, C. C., Rabavilas, A. D., & Cavouras,  
1043 D. (2004). Design and implementation of an SVM-based computer classification system  
1044 for discriminating depressive patients from healthy controls using the P600 component  
1045 of ERP signals. *Computer Methods and Programs in Biomedicine*, 75(1), 11–22.  
1046 <https://doi.org/10.1016/j.cmpb.2003.09.003>
- 1047 Klumpp, H., Keller, J., Miller, G. A., Casas, B. R., Best, J. L., & Deldin, P. J. (2010).  
1048 Semantic processing of emotional words in depression and schizophrenia. *International*  
1049 *Journal of Psychophysiology*, 75(2), 211–215.  
1050 <https://doi.org/10.1016/j.ijpsycho.2009.12.004>
- 1051 Kreuz, R. J., & Glucksberg, S. (1989). How to Be Sarcastic: The Echoic Reminder Theory of  
1052 Verbal Irony. *Journal of Experimental Psychology: General*, 118(4), 374–386.  
1053 <https://doi.org/10.1037/0096-3445.118.4.374>
- 1054 Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to  
1055 syntax. *Brain Research*, 1146, 23–49. <https://doi.org/10.1016/j.brainres.2006.12.063>
- 1056 Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the  
1057 N400 component of the event-related brain potential (ERP). *Annual Review of*  
1058 *Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- 1059 Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in  
1060 language comprehension. *Trends in Cognitive Sciences*, 4(12), 463–470.  
1061 [https://doi.org/10.1016/S1364-6613\(00\)01560-6](https://doi.org/10.1016/S1364-6613(00)01560-6)
- 1062 Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect  
1063 semantic incongruity. *Science*, 207(4427), 203–205.

- 1064 Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). International affective picture system  
1065 (IAPS): Affective ratings of pictures and instruction manual. *University of Florida,*  
1066 *Gainesville.* Tech Rep A-8.
- 1067 Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:  
1068 (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12),920–933.  
1069 <https://doi.org/10.1038/nrn2532>
- 1070 Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding  
1071 abilities in clinical depression. *Journal of Affective Disorders*, *86*(2–3), 247–258.  
1072 <https://doi.org/10.1016/j.jad.2005.02.007>
- 1073 Leggitt, J. S., & Gibbs, R. W. (2000). Emotional reactions to verbal irony. *Discourse*  
1074 *Processes*, *29*(1), 1–24. [https://doi.org/10.1207/S15326950dp2901\\_1](https://doi.org/10.1207/S15326950dp2901_1)
- 1075 Leppänen, J. M. (2006). Emotional information processing in mood disorders: a review of  
1076 behavioral and neuroimaging findings. *Current Opinion in Psychiatry*, *19*(1), 34–39.  
1077 <https://doi.org/10.1097/01.yco.0000191500.46411.00>
- 1078 Marinkovic, K., Baldwin, S., Courtney, M. G., Witzel, T., Dale, A. M., & Halgren, E. (2011).  
1079 Right hemisphere has the last laugh: neural dynamics of joke appreciation. *Cognitive,*  
1080 *Affective, & Behavioral Neuroscience*, *11*(1), 113–130. [https://doi.org/10.3758/s13415-](https://doi.org/10.3758/s13415-010-0017-7)  
1081 [010-0017-7](https://doi.org/10.3758/s13415-010-0017-7)
- 1082 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.  
1083 *Journal of Neuroscience Methods*, *164*(1), 177–190.  
1084 <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- 1085 Martin, R. a. (2010). *The Psychology of Humor: An Integrative Approach (Google eBook).*  
1086 <https://doi.org/10.1037/h0075400>
- 1087 Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual  
1088 differences in uses of humor and their relation to psychological well-being:  
1089 Development of the humor styles questionnaire. *Journal of Research in Personality*,  
1090 *37*(1), 48–75. [https://doi.org/10.1016/S0092-6566\(02\)00534-2](https://doi.org/10.1016/S0092-6566(02)00534-2)
- 1091 Matthews, J. K., Hancock, J. T., & Dunham, P. J. (2006). The Roles of Politeness and Humor  
1092 in the Asymmetry of Affect in Verbal Irony. *Discourse Processes*, *41*(1), 3–24.  
1093 [https://doi.org/10.1207/s15326950dp4101\\_2](https://doi.org/10.1207/s15326950dp4101_2)
- 1094 Moran, E. K., Mehta, N., & Kring, A. M. (2012). Emotional responding in depression:  
1095 Distinctions in the time course of emotion. *Cognition & Emotion*, *26*(7), 1153–1175.  
1096 <https://doi.org/10.1080/02699931.2011.638909>
- 1097 Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion  
1098 phenomenon: ERPs reveal temporary semantic change deafness in discourse  
1099 comprehension. *Cognitive Brain Research*, *24*(3), 691–701.  
1100 <https://doi.org/10.1016/j.cogbrainres.2005.04.003>



- 1101 Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic  
 1102 anomaly. *Journal of Memory and Language*, *31*(6), 785–806.  
 1103 [https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
- 1104 Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of  
 1105 biased attention in depression. *Depression and Anxiety*, *27*(12), 1135–1142.  
 1106 <https://doi.org/10.1002/da.20755>
- 1107 Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological  
 1108 exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2),  
 1109 277–293. <https://doi.org/10.1162/jocn.2010.21411>
- 1110 Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes  
 1111 reflected by P600 effects: Evidence from ERPs and neural oscillations. *PLoS ONE*, *9*(5),  
 1112 e96840. <https://doi.org/10.1371/journal.pone.0096840>
- 1113 Rottenberg, J., Gross, J. J., & Gotlib, I. H. (2005). Emotion context insensitivity in major  
 1114 depressive disorder. *Journal of Abnormal Psychology*, *114*(4), 627–639.  
 1115 <https://doi.org/10.1037/0021-843X.114.4.627>
- 1116 Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement  
 1117 reactivity differentially predict concurrent and prospective functioning in major  
 1118 depressive disorder. *Emotion*, *2*(2), 135–146. <https://doi.org/10.1037/1528-3542.2.2.135>
- 1119 Sanford, A. J., Leuthold, H., Bohan, J., & Sanford, A. J. S. (2011). Anomalies at the  
 1120 borderline of awareness: An ERP study. *Journal of Cognitive Neuroscience*, *23*(3), 514–  
 1121 523. <https://doi.org/10.1162/jocn.2009.21370>
- 1122 Schwoebel, J., Dews, S., Winner, E., & Srinivas, K. (2000). Obligatory Processing of the  
 1123 Literal Meaning of Ironic Utterances: Further Evidence. *Metaphor and Symbol*, *15*(1–2),  
 1124 47–61. [https://doi.org/10.1207/s15327868ms151&2\\_4](https://doi.org/10.1207/s15327868ms151&2_4)
- 1125 Shibata, M., Terasawa, Y., Osumi, T., Masui, K., Ito, Y., Sato, A., & Umeda, S. (2017). Time  
 1126 course and localization of brain activity in humor comprehension: An ERP / sLORETA  
 1127 study. *Brain Research*, *1657*, 215–222. <https://doi.org/10.1016/j.brainres.2016.12.010>
- 1128 Sloan, D. M., Strauss, M. E., & Wisner, K. L. (2001). Diminished response to pleasant  
 1129 stimuli by depressed women. *Journal of Abnormal Psychology*, *110*(3), 488–  
 1130 493. <https://doi.org/10.1037/0021-843X.110.3.488>
- 1131 Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What's behind a  
 1132 P600? Integration Operations during Irony Processing. *PLoS ONE*, *8*(6), 1–10.  
 1133 <https://doi.org/10.1371/journal.pone.0066839>
- 1134 Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., & Noveck, I. A. (2012). Neural  
 1135 evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*,  
 1136 *63*(1), 25–39. <https://doi.org/10.1016/j.neuroimage.2012.06.046>
- 1137 Strömmer, J. M., Pöldver, N., Waselius, T., Kirjavainen, V., Järveläinen, S., Björkstén, S., ...  
 1138 Astikainen, P. (2017). Automatic auditory and somatosensory brain responses in relation

- 1139 to cognitive abilities and physical fitness in older adults. *Scientific reports*, 7(1), 1–12.  
1140 <https://doi.org/10.1038/s41598-017-14139-9>
- 1141 Talarowska, M., Zajączkowska, M., & Gałęcki, P. (2015). Cognitive functions in first-  
1142 episode depression and recurrent depressive disorder. *Psychiatria Danubina*, 27(1), 38–  
1143 43.
- 1144 Thompson, D., Mackenzie, I. A. N. G., & Leuthold, H. (2016). Emotional responses to irony  
1145 and emoticons in written language: Evidence from EDA and facial EMG.  
1146 *Psychophysiology*, 53(7), 1054–1062. <https://doi.org/10.1111/psyp.12642>
- 1147 Toplak, M., & Katz, A. N. (2000). On the uses of sarcastic irony. *Journal of pragmatics*,  
1148 32(10), 1467–1488.
- 1149 Uekermann, J., Channon, S., Lehmkämpfer, C., Abdel-Hamid, M., Vollmoeller, W., & Daum,  
1150 I. (2008). Executive function, mentalizing and humor in major depression. *Journal of the*  
1151 *International Neuropsychological Society*, 14(1), 55–62.  
1152 <https://doi.org/10.1017/S1355617708080016>
- 1153 Varga, E., Simon, M., Tényi, T., Schnell, Z., Hajnal, A., Orsi, G., ... Herold, R. (2013). Irony  
1154 comprehension and context processing in schizophrenia during remission - A functional  
1155 MRI study. *Brain and Language*, 126(3), 231–242.  
1156 <https://doi.org/10.1016/j.bandl.2013.05.017>
- 1157 Van Boxtel, A. (2010, August). Facial EMG as a tool for inferring affective states.  
1158 In *Proceedings of measuring behavior* (pp. 104-108). Wageningen: Noldus Information  
1159 Technology.
- 1160 Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in  
1161 electromagnetic and hemodynamic studies. *Brain and Language*, 97(3), 279–293.  
1162 <https://doi.org/10.1016/j.bandl.2005.11.003>
- 1163 World Health Organization. (2010). ICD-10 Version: 2010. Retrieved from  
1164 <http://apps.who.int/classifications/icd10/browse/2010/en>
- 1165 Xu, Q., Ruohonen, E. M., Ye, C., Li, X., Kreegipuu, K., Stefanics, G., ... Astikainen, P.  
1166 (2018). Automatic Processing of Changes in Facial Emotions in Dysphoria: A  
1167 Magnetoencephalography Study. *Frontiers in Human Neuroscience*, 12, 186.  
1168 <https://doi.org/10.3389/fnhum.2018.00186>

1169 Table 1. Demographics and clinical measures. Statistics show independent t-test or  $\chi^2$  test  
 1170 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 <sup>a</sup>	Medium	9	9	$\chi^2(1) = .38, p = .536$
	High	8	12	
Gender	Male	3	6	$\chi^2(1) = .62, p = .431$
	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 <sup>b</sup>	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 <sup>c</sup>	SSRI	5	Na	Na
	SNRI	1	Na	Na
	Other	4	Na	Na

1171 *Note.* DYS = dysphoric group, CTRL = control group, SD = standard deviation, BDI-II = Beck's Depression  
 1172 Inventory, Second Edition.

1173 <sup>a</sup>Medium = high school or vocational school, High = university.

1174 <sup>b</sup>Depression severity based on participants self-report on their diagnosis. The diagnosis of depression was based  
 1175 on the International Classification of Diseases and Related Health Problems, 10th Revision (ICD-10; World  
 1176 Health Organization, 2010) criteria. There is missing information related to disease severity from three  
 1177 participants, and one participant did not have diagnosis.

1178 <sup>c</sup>SSRI = selective serotonin reuptake inhibitor, SNRI = serotonin and norepinephrine reuptake inhibitor, Other =  
 1179 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,  $\eta_p^2$  = partial eta squared for effect size.

Facial EMG activity	Effect	F	df	p	$\eta_p^2$
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
	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
Zygomaticus activity	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,  $\eta_p^2$  = partial eta squared for effect size.

ERP	Effect	F	df	p	$\eta_p^2$
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
	Stimulus type (non-ironic vs. ironic)	23.243	1, 36	< .001	.392
P600	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

1183 Table 4. Mean amplitude values ( $\mu\text{V}$ ) for P600 response for each ROI and each stimulus type  
 1184 in the control and dysphoric groups. SD = standard deviation.

<b>Group</b>	<b>Stimulus type</b>	<b>Group</b>	<b>Mean (SD)</b>
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)

## 1185 **Supplementary Material**

### 1186 1 Humor Styles Questionnaire

1187 Humor Styles Questionnaire (HSQ; Martin et al., 2003) was filled out on the same day with  
1188 the second day's behavioral measurement. The HSQ, translated into Finnish based on the  
1189 original version, consists of a 32-item measure with a Likert-scale from 1 to 7 (totally  
1190 disagree = 1, totally agree = 7). The HSQ evaluates individual preference for the use of  
1191 humor in four dimensions (affiliative, self-enhancing, aggressive, self-defeating) and has  
1192 been applied in previous studies to investigate the relationship between the use of humor  
1193 styles and depressive symptoms, loneliness, or suicidal ideation (e.g., Frewen et al., 2008;  
1194 Schermer et al., 2017; Tucker et al., 2013). Here, HSQ was applied to explore if there were  
1195 differences in humor style preference between the dysphoric and control groups, which could  
1196 explain possible group differences in the responsiveness to irony, since irony is related to  
1197 sarcasm (Kreuz and Glucksberg, 1989) that can be categorized as aggressive humor (Martin  
1198 et al., 2003).

#### 1199 1.1 Analysis and statistics for the Humor Style Questionnaire and behavioral data

1200 The reliability of the HSQ was first estimated based on internal consistency values.  
1201 Cronbach's alphas for each dimension were: affiliative humor = .853; self-enhancing = .829;  
1202 aggressive humor = .684; self-defeating = .756. The averaged HSQ scores were calculated  
1203 separately for each participant and dimension. For the HSQ scores, multivariate analysis of  
1204 variance (MANOVA) was applied with a within-subject factor Humor style (affiliative, self-  
1205 enhancing, aggressive, self-defeating) and a between-subject factor Group (control,  
1206 dysphoric). An additional MANOVA analysis, similar to the previously mentioned but with  
1207 current medication status (whether the participant currently had depression medication or not)

1208 serving as a covariate, was conducted. Post-hoc tests for within-subject comparisons were  
1209 implemented by applying repeated measures of analysis of variance (ANOVAs) with  
1210 Bonferroni correction in the control and dysphoric groups separately. Post-hoc tests  
1211 comparing each humor style dimension between the control and dysphoric groups were  
1212 performed by using two-tailed independent samples *t*-tests with Bootstrap statistics based on  
1213 1000 permutations as implemented in IBM SPSS Statistics 24.0 (Armonk, NY: IBM  
1214 incorporated). *P*-values in multiple comparisons for independent sample *t*-tests were adjusted  
1215 by false discovery rate (Benjamini and Hochberg, 1995).

1216 Pearson's correlation coefficients (two-tailed) were used to evaluate the relationship between  
1217 depressive symptoms (BDI-II scores) and humor styles. Multiple correlations were controlled  
1218 by applying false discovery rate (Benjamini and Yekutieli, 2001) at 0.05.

## 1219 1.2 Results for Humor Style Questionnaire

1220 The MANOVA indicated no main effect of Group ( $p = .929$ ). A main effect of Humor style  
1221 was found,  $F(3,34) = 63.356, p < .001, \eta_p^2 = .848$ . The main effect was modified by an  
1222 interaction effect of Humor style  $\times$  Group,  $F(3,34) = 3.827, p = .018, \eta_p^2 = .252$ . MANOVA  
1223 with current medication status as a covariate variable showed no Stimulus condition  $\times$   
1224 Current medication status interaction,  $p = .587$ .

1225 Separate ANOVAs for the two groups showed a significant main effect of Humor style in  
1226 each group (dysphoric:  $F(3,48) = 22.616, p < .001, \eta_p^2 = .586$ ; control:  $F(3,60) = 47.029, p$   
1227  $< .001, \eta_p^2 = .702$ ). For the dysphoric group, pairwise comparisons with Bonferroni correction  
1228 showed that the tendency to use an affiliative humor style in the dysphoric participants was  
1229 greater compared with the tendency to use a self-enhancing, an aggressive, or a self-defeating  
1230 humor style, all *p*-values  $< .001$ . The comparisons also showed that the dysphoric participants



1231 were more likely to use a self-defeating humor style than an aggressive humor style,  $p = .005$ .  
1232 There were no differences between using a self-enhancing humor style and an aggressive or a  
1233 self-defeating humor style. Among the controls, pairwise comparisons with Bonferroni  
1234 correction showed that the control group had the greatest preference for an affiliative humor  
1235 style compared with a self-enhancing ( $p = .009$ ), an aggressive ( $p < .001$ ), or a self-defeating  
1236 ( $p < .001$ ) humor style. Furthermore, the control group were most unlikely to use an  
1237 aggressive humor style compared with an affiliative, a self-enhancing ( $p < .001$ ), or a self-  
1238 defeating ( $p = .005$ ) humor style. The pairwise comparisons also revealed that the control  
1239 participants prefer to use a self-enhancing humor style rather than a self-defeating humor  
1240 style,  $p = .003$ . Results of post-hoc tests comparing each humor style dimension between the  
1241 control and dysphoric groups are shown in Supplementary Table 3. Mean values, standard  
1242 deviation (SD) and range of Humor Style Questionnaire (HSQ) scores separately for each  
1243 humor style dimension and group are also presented in Supplementary Table 3.

1244 No correlations were found between the amount of depressive symptoms (measured with the  
1245 BDI-II scores) and scores for each HSQ dimension when the whole sample was included or  
1246 when only the dysphoric group was included in the analysis (all  $p$ -values  $> .100$ ).

## 1247 2 Linear regression analyses

1248 In addition to the ANOVAs reported in the main text, simple linear regression analyses were  
1249 conducted for the whole sample to investigate the relationship between the amount of  
1250 depressive symptoms (BDI-II scores) and ERP responses (P600 only, because we  
1251 hypothesized a group difference specifically for it), facial EMG amplitudes and behavioral  
1252 evaluations (group difference hypothesized for valence ratings of conversation) separately.  
1253 BDI-II scores were applied as a predictor of P600 amplitude, facial EMG amplitude and

1254 behavioral ratings for valence. Results of simple linear regression model with BDI-II scores  
1255 as a predictor are presented in Supplementary Table 4.

1256 In summary, the results of simple linear regression analyses showed that there were no  
1257 significant relationships between BDI-II scores and P600 amplitudes, or facial EMG  
1258 amplitudes, or behavioral valence ratings.

1259 3 Cognitive tests

1260 3.1 Principal component analysis on cognitive tests

1261 In order to reduce variables for analysis of covariance investigating effect of cognitive  
1262 abilities for P600 responses, principal component analysis (PCA) using an Oblimin rotation  
1263 with Kaiser Normalization was conducted to extract factors of the cognitive tests. First, we  
1264 selected the cognitive tests that can be expected to associate with irony processing or the  
1265 cognitive deficits in depression (Supplementary Table 5). Thus, eight tests were applied as  
1266 variables in the PCA to investigate factor solutions. With a criteria of eigenvalue larger than  
1267 1.0, the PCA showed that three factors which could explain 71.5% of the variance, were  
1268 extracted (Supplementary Table 5). Finally, these three factors, named executive function, list  
1269 memory, and semantic processing, were utilized as covariates independently in ANCOVA  
1270 models of P600. In addition, a two-way repeated measures ANOVA with a within-subject  
1271 variable Cognitive factor (executive function vs. list memory vs. semantic processing) and a  
1272 between-subject variable Group (control vs. dysphoric) was conducted to examine possible  
1273 group difference in cognitive factors.

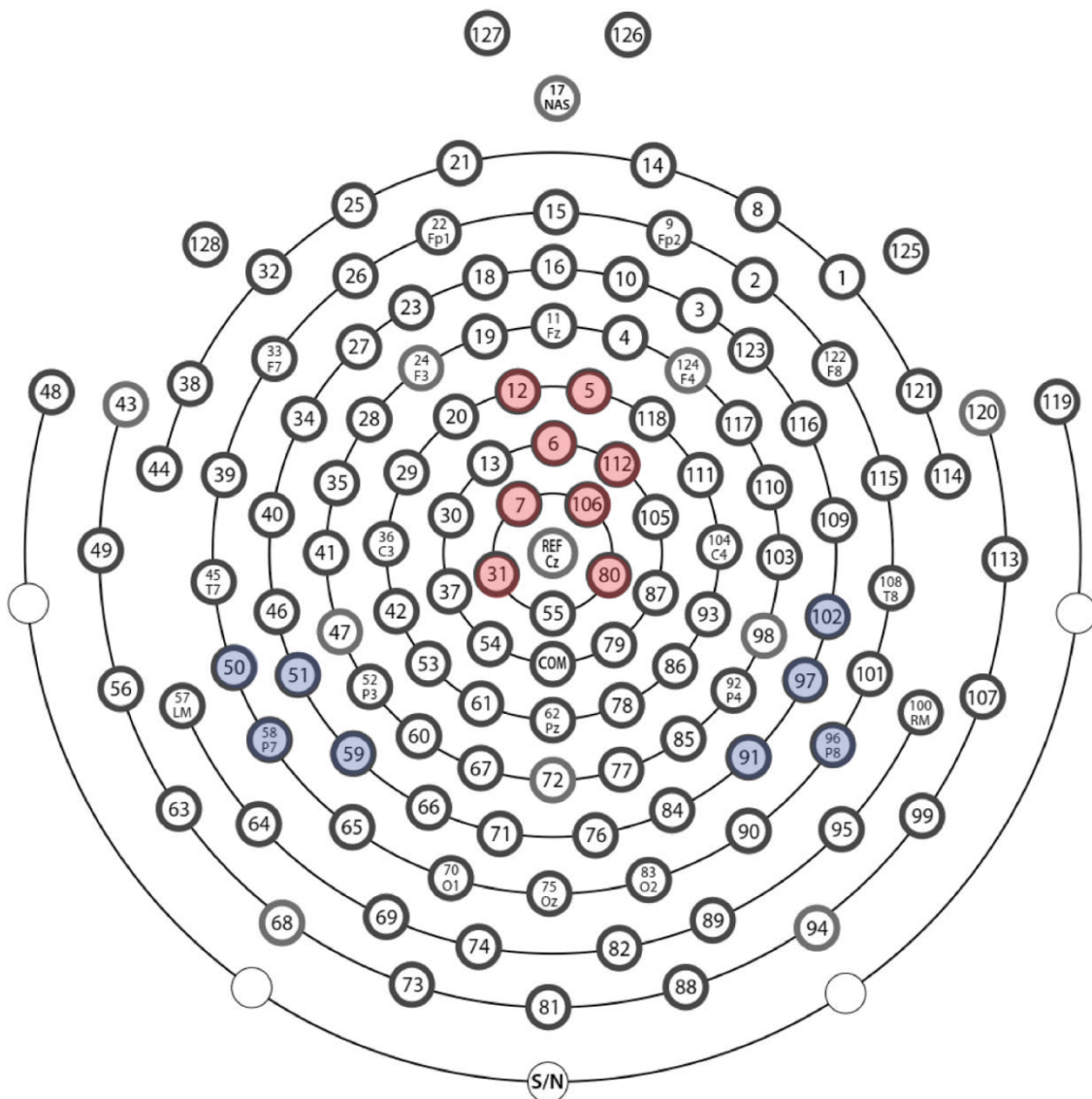
## 1274 3.2 Results of group comparison in cognitive factors

1275 The two-way repeated measures ANOVA showed that there was neither main effect of  
1276 Cognitive factor ( $F(2, 72) = 0.015, p = .985, \eta_p^2 < .001$ ) nor main effect of Group ( $F(1, 36) =$   
1277  $0.917, p = .345, \eta_p^2 = .025$ ). The interaction of Cognitive factor and Group were non-  
1278 significant,  $F(2, 72) = 1.368, p = .261, \eta_p^2 = .037$ .

## 1279 4 Permutation tests for electrode selection

1280 For the N400-like activity and P600, the selection of the electrodes for the further statistical  
1281 analyses was based on a data-driven method (see also, e.g., Hämäläinen et al., 2018;  
1282 Strömmer et al., 2017). BESA Statistics 2.0 software (BESA GmbH, Graefelfing, Germany)  
1283 was applied to perform cluster-based permutation tests (Maris and Oostenveld, 2007)  
1284 between the amplitude values of responses to ironic and non-ironic trials with 3000 iterations.  
1285 In order to enhance the sensitivity of the test (Groppe, Urbach, and Kutas, 2011; Maris and  
1286 Oostenveld, 2007), time points where stimulus type effect (non-ironic vs. ironic) related  
1287 N400-like activity and P600 is unlikely to be observed were excluded. That is, we performed  
1288 the permutation tests including each time point from two separate time windows: 300 to 500  
1289 ms after stimulus onset for the N400-like activity (Kutas & Federmeier, 2011), and 500 to  
1290 800 ms after stimulus onset for P600 (Bornkessel-Schlesewsky & Schlewsky, 2008). For  
1291 both tests, the channel (electrode) cluster distance was set to 3.5 cm, and the alpha level for  
1292 significant cluster was 0.05. In time window of 300-500 ms (N400-like activity), the cluster-  
1293 based test showed one electrode cluster ( $p = .025$ ), where the responses to ironic and non-  
1294 ironic sentences differed. In time window of 500-800 ms (P600), two clusters were observed:  
1295 one on the left ( $p = .013$ ) and one on the right ( $p = .014$ ) hemisphere. Then, we located the  
1296 highest t-value ( $t\text{-value}_{\max}$ ) within each electrode cluster: in the electrode cluster for N400-

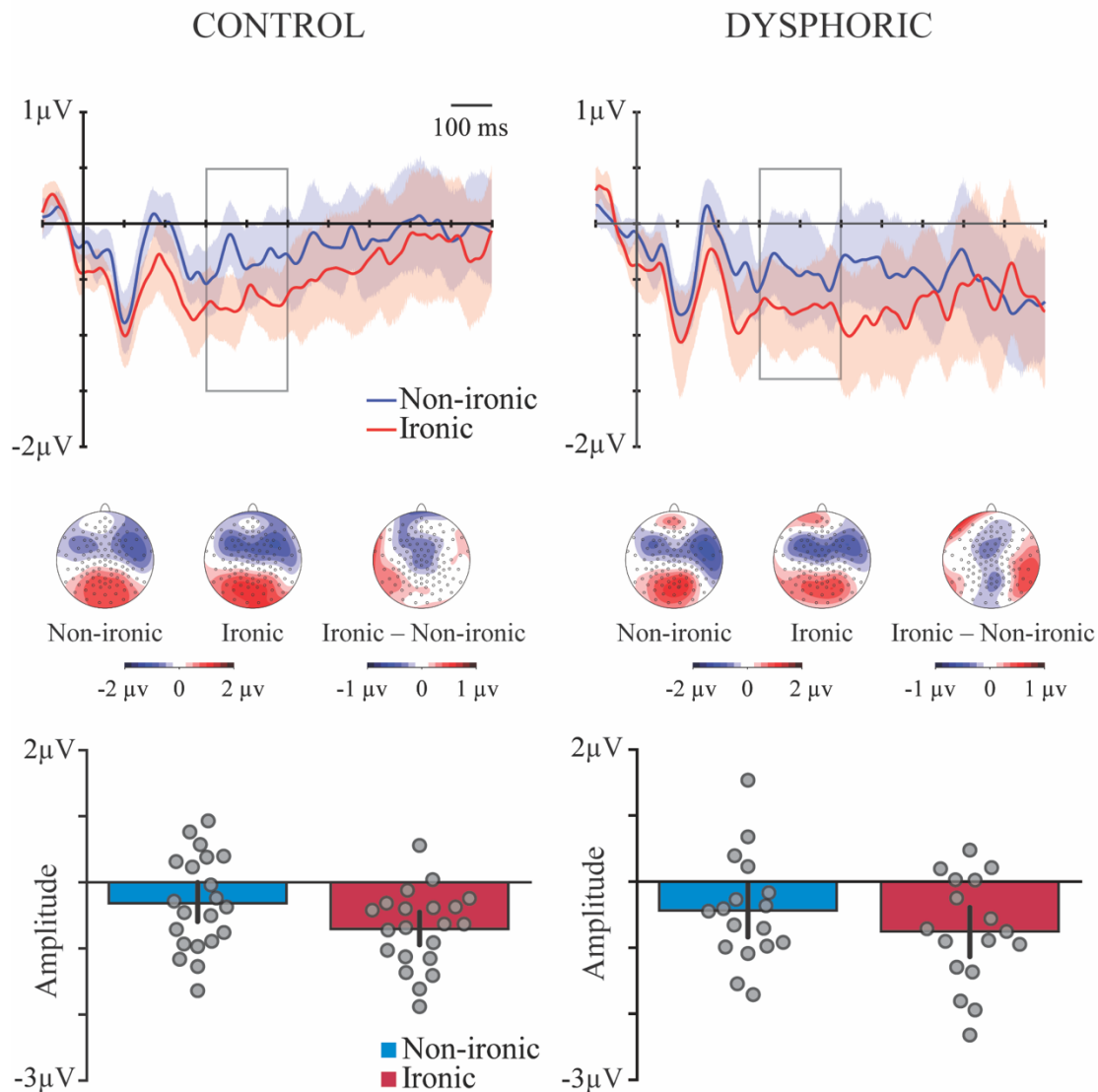
1297 like activity:  $t\text{-value}_{\max} = 4.0$ ; in the electrode clusters for P600, for the left cluster:  $t\text{-value}_{\max}$   
 1298  $= 4.5$ , and for the right cluster:  $t\text{-value}_{\max} = 4.0$ . The electrode with the highest  $t$ -value and its  
 1299 surrounding electrodes within the cluster were selected to the further analyses if the highest  $t$ -  
 1300 value of the electrode was larger than the 75% of  $t\text{-value}_{\max}$  (for N400-like activity: the  
 1301 highest  $t$ -values for all selected electrodes  $> 3.0$ ; for P600: the highest  $t$ -values for selected  
 1302 electrodes on the left  $> 3.4$ , for selected electrodes on the right  $> 3.0$ ). The electrodes applied  
 1303 in the statistical analysis, i.e., the results of this procedure, are depicted in Supplementary  
 1304 Figure 1.



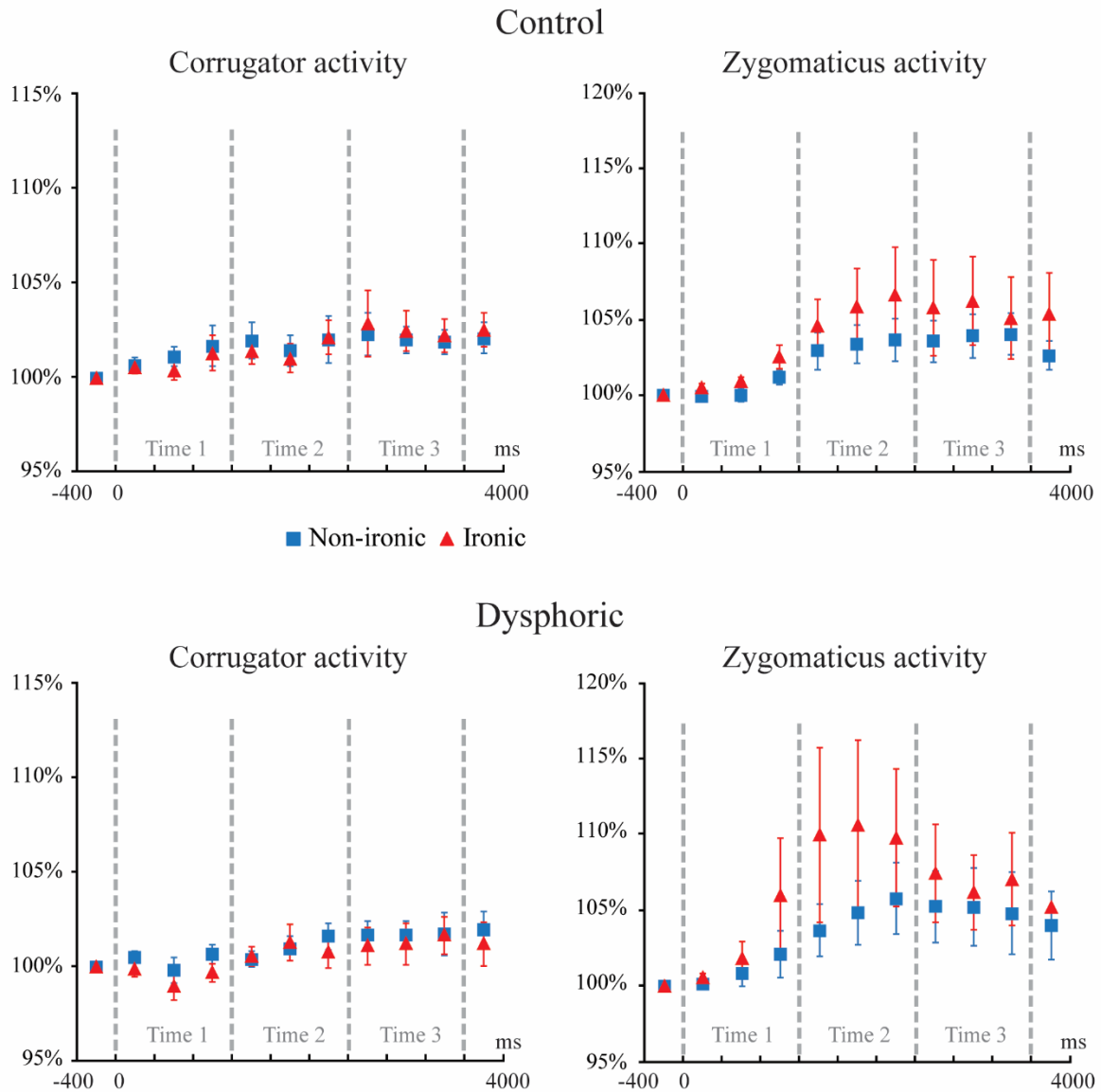
1305

1306 **Supplementary Figure 1.** Map of EGI 128-Channel Net (HydroCel Geodesic Sensor Net)  
1307 and the electrodes applied in the analyses. For N400-like activity, electrodes 5, 6, 7, 12, 31,  
1308 80, 106, 112 (red fillings) were applied in the analyses. For P600, electrodes 50, 51, 58, 59,  
1309 91, 96, 97, 102 (blue fillings) were applied in the analyses.

## N400-like activity



1310  
 1311 Supplementary Figure 2. Grand-averaged waveforms, topographical maps and histograms  
 1312 depicting mean amplitudes for N400-like activity to non-ironic and ironic stimuli in control  
 1313 (left) and dysphoric (right) group. Please refer to Supplementary Figure 1 for the electrode  
 1314 sites applied in the waveforms and histograms (averaged activity of the electrodes within the  
 1315 ROI presented here). In the waveform figure, the blue and red shadows represent 95%  
 1316 confidence intervals, and the grey rectangles the time window applied in the analysis for  
 1317 N400-like activity (mean amplitude values between 300–500 ms). The histograms shows the  
 1318 mean amplitudes in the analysis window and the error bars represent 95% confidence  
 1319 intervals. The scatterplots represent individual participants' amplitudes.



1320  
 1321 **Supplementary Figure 3.** The grand-averages of the mean EMG amplitude percentage  
 1322 relative to the baseline for the corrugator (left) and for the zygomaticus (right) to non-ironic  
 1323 (blue) and ironic (red) stimulus in control and dysphoric group. Average percentage values  
 1324 and standard error of mean (SE) are presented for descriptive purposes in 400-ms segments,  
 1325 and the grey dotted lines show the segments applied in the ANOVA (Time 1: 0–1200 ms;  
 1326 Time 2: 1200–2400 ms; Time 3: 2400–3600 ms).

1327 **Supplementary Table 1.** Cognitive test scores in the control and dysphoric group. The  
 1328 statistics show the results based on two-tailed independent samples t-tests comparing the  
 1329 groups in the scores.

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

1330 *Note.* Differences between two groups were tested by using two-tailed independent-samples t-tests.  $d$  = Cohen's  
1331  $d$ , AVLT = Auditory Verbal Learning Test, TMT = Trail Making Test, # = scores measured in the number of  
1332 items, p = point, more means better, s = second, less means better.

1333 <sup>a</sup>Uncorrected  $p$ -values.  $P$ -values smaller than .05 are in bold, and for them corrected  $p$ -values based on FDR  
1334 (false discovery rate) are presented in parentheses.

1335 **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. <i>J. Affect. Disord.</i> 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). <i>Neuropsychol. Rev.</i> 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? <i>Assessment</i> 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. <i>Neuropsychol. Rev.</i> 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. <i>Arch. Clin. Neuropsychol.</i> 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. <i>Nat. Protoc.</i> 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. <i>Neuropsychol. Rev.</i> 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. <i>Clin. Neuropsychol.</i> 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.), <i>Studies in languages</i> (Vol. 12). Joensuu, Finland: University of Joensuu.

1336 Note. WMS-R = Wechsler Memory Scale-Revised, WMS-III = Wechsler Memory Scale-Third Edition, WAIS-

1337 III = Wechsler Adult Intelligence Scale-Third Edition, WAIS-IV = Wechsler Adult Intelligence Scale-Fourth

1338 Edition.

1339 **Supplementary Table 3.** Mean values, standard deviation (SD) and range of Humor Style  
1340 Questionnaire (HSQ) scores separately for each humor style dimension and group reflecting  
1341 significant interaction of humor style × group. Statistics show post hoc tests (two-tailed

1342 independent samples *t*-tests, bootstrapping with 1000 permutations) comparing the humor style  
 1343 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]	

1344 *Note.* Statistics show the results comparing the two groups in the preference for each humor style. Reported *p*-  
 1345 values are adjusted by false discovery rate.

1346 **Supplementary Table 4.** Results of simple linear regression model with BDI-II scores as a  
 1347 predictor.

<b>Dependent variables</b>	<b>Coefficients (Beta)</b>	<b>R Square</b>	<b>P-value</b>
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

1348

1349 **Supplementary Table 5.** List and description of selected cognitive tests for principal  
 1350 component analysis, and factor loadings using an oblimin rotation with Kaiser Normalization  
 1351 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.	<b>-0.678</b>	-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.	<b>-0.811</b>	0.068	-0.03
Similarities (p)	Identifying the qualitative relationship between two words. This test measures abstract thinking, verbal reasoning, and concept formatting.	<b>0.727</b>	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	<b>0.958</b>	-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	<b>0.916</b>	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	<b>0.870</b>
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	<b>0.818</b>
Fluency semantic-animals(#)	Naming words as fast as possible that belong to the animal category. This test measures semantic fluency.	-0.255	0.06	<b>0.566</b>

1352 *Note.* TMT = Trail Making Test, AVLT = Auditory Verbal Learning Test, s = second, less means better, p =  
 1353 point, more means better, # = scores measured in the number of items. Factor loadings, for the component that  
 1354 the tests contribute most, are marked in bold.

1355 **References**

- 1356 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and  
1357 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series*  
1358 *B (Methodological)*, 57(1), 289–300. [https://doi.org/10.1111/j.2517-](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)  
1359 [6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)
- 1360 Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple  
1361 testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188.  
1362 <https://doi.org/10.1214/aos/1013699998>
- 1363 Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on  
1364 “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59(1),  
1365 55–73. <https://doi.org/10.1016/J.BRAINRESREV.2008.05.003>
- 1366 Frewen, P. A., Brinker, J., Martin, R. O. D. A., and Dozois, D. J. A. (2008). Humor styles and  
1367 personality- vulnerability to depression. *Humor*, 21(2), 179–195.  
1368 [doi:10.1515/HUMOR.2008.009](https://doi.org/10.1515/HUMOR.2008.009).
- 1369 Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related  
1370 brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711–  
1371 1725. <https://doi.org/10.1111/j.1469-8986.2011.01273.x>
- 1372 Hämäläinen, J., Landi, N., Loberg, O., Lohvansuu, K., Pugh, K., & Leppänen, P. H. T.  
1373 (2018). Brain event-related potentials to phoneme contrasts and their correlation to  
1374 reading skills in school-age children. *International Journal of Behavioral Development*,  
1375 42(3), 357–372. <https://doi.org/10.1177/0165025417728582>
- 1376 Kreuz, R. J., and Glucksberg, S. (1989). How to Be Sarcastic: The Echoic Reminder Theory  
1377 of Verbal Irony. *Journal of experimental psychology*. 118(4), 374–386.  
1378 [doi:10.1037/0096-3445.118.4.374](https://doi.org/10.1037/0096-3445.118.4.374).
- 1379 Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the  
1380 N400 component of the event-related brain potential (ERP). *Annual Review of*  
1381 *Psychology*, 62, 621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- 1382 Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data.  
1383 *Journal of Neuroscience Methods*, 164(1), 177–190.  
1384 <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- 1385 Martin, R. A., Puhlik-Doris, P., Larsen, G., Gray, J., and Weir, K. (2003). Individual  
1386 differences in uses of humor and their relation to psychological well-being:  
1387 Development of the humor styles questionnaire. *Journal of research in personality*.  
1388 37(1), 48–75. [doi:10.1016/S0092-6566\(02\)00534-2](https://doi.org/10.1016/S0092-6566(02)00534-2).
- 1389 Schermer, J. A., Martin, R. A., Vernon, P. A., Martin, N. G., Conde, L. C., Statham, D., et al.  
1390 (2017). Lonely people tend to make fun of themselves: A behavior genetic analysis of  
1391 humor styles and loneliness. *Personality and Individual Differences*. 117, 71–73.  
1392 [doi:10.1016/j.paid.2017.05.042](https://doi.org/10.1016/j.paid.2017.05.042).

- 1393 Strömmer, J. M., Pöldver, N., Waselius, T., Kirjavainen, V., Järveläinen, S., Björkstén, S., ...  
1394 Astikainen, P. (2017). Automatic auditory and somatosensory brain responses in relation  
1395 to cognitive abilities and physical fitness in older adults. *Scientific reports*, 7(1), 1–12.  
1396 <https://doi.org/10.1038/s41598-017-14139-9>
- 1397 Tucker, R. P., Wingate, L. R. R., O’Keefe, V. M., Sligh, M. L., Judah, M. R., and Rhoades-  
1398 Kerswill, S. (2013). The moderating effect of humor style on the relationship between  
1399 interpersonal predictors of suicide and suicidal ideation. *Personality and Individual*  
1400 *Differences*. 54(5), 610–615. doi:10.1016/j.paid.2012.11.023.