

Can Irony Regulate Negative Emotion?

Evidence from Behavior and ERPs

Valeria A. Pfeifer*, Jessica R. Andrews-Hanna, & Vicky T. Lai

University of Arizona, Department of Psychology & Cognitive Science Program

*Corresponding author

Email:

vpfeifer@arizona.edu

Mailing Address:

1503 E University Blvd

Tucson, AZ 85721

USA

Abstract

This study used ratings and event-related potentials (ERPs) to compare the mechanisms through which verbal irony and cognitive reappraisal mitigate negative emotion. Verbal irony is when the literal meaning of words contrasts with their intended meaning. Cognitive reappraisal is when we reconsider emotional stimuli to make them less intense. Our hypothesis was that cognitive reappraisal is a potential mechanism through which irony reduces negative emotion. Participants viewed mildly negative pictures first, then read an ironic or literal statement about it in one block, and used cognitive reappraisal of or attending to the picture in the other block. Participants then viewed the picture for a second time, before rating how negative they felt. Behaviorally, irony reduced negative feelings more than literal statements, and reappraisal reduced negative feelings more than attending, with a larger reduction from reappraisal than from irony. In ERPs, irony elicited a prolonged N400 compared to literal, indexing an initial contrast between picture and word affect and sustained processing of their combination. Cognitive reappraisal elicited a larger late positivity compared to attending at the instruction screen. No differences were found during second picture presentation. These findings suggest that irony and cognitive reappraisal can reduce negative affect in different ways.

Keywords: Irony, Cognitive Reappraisal, Emotion regulation, N400, LPC

How do you feel when something bad happens and someone makes an ironic comment about it? Verbal irony is when the literal meaning of words is in contrast with their intended meaning. For example, describing a dance performance as “elegant” after they stumbled. While people reported that they use irony “to express negative emotions” (Roberts & Kreuz, 1994), laboratory studies showed that irony can reduce negative feelings in certain scenarios (Dews & Winner, 1995; Pfeifer & Lai, 2021). Building on these observations, we take a novel approach to further explore the potential involvement of irony on the regulation of emotions. Specifically, using electrophysiology and ratings, we ask how effective irony is in reducing negative emotion in response to negative events, and if the underlying mechanisms are similar to those of cognitive reappraisal, a form of emotion regulation.

1.1. Verbal Irony and Emotion

The *tinge hypothesis* (Dews & Winner, 1995) posits that irony reduces negativity because the literal (positive) meaning tinges the interpretation of the message (negative), thereby muting it. Dews and Winner (1995) offered support for their *tinge hypothesis* through tactfulness ratings, wherein a speaker was rated as being less critical when using irony compared to literal language. Similarly, Ivanko and Pexman (2003) found that irony used in mildly negative situations is rated as more polite than literal language. More recently, Pfeifer and Lai (2021) found that a speaker is rated to be in a less negative and less aroused mental state when using irony compared to literal language, regardless of context emotion.

In recent years, researchers have used Event-related potentials (ERPs) which have a temporal resolution of milliseconds and can inform us about fine-grained aspects of emotion and of language. ERP studies of irony have typically used written or spoken context and focused on

two ERP components, the N400 and the P600 / Late Positivity Component (LPC). The N400 is a negative-going waveform that typically peaks around 300-500 ms after stimulus onset, with a centro-parietal distribution, and is sensitive to semantic processing (Kutas & Federmeier, 2011). The N400 has been associated with the ease of *retrieving* lexico-semantic information (Brouwer, Fitz, & Hoecks, 2012), as well as the ease of *integrating* the semantic information with prior context (Osterhout & Holcomb, 1992).

In ERP studies of irony (Lai et al., 2023), the most commonly found component is the P600 or LPC, a positive-going waveform after 500 ms post-stimulus with a centro-parietal distribution. The P600 in language has previously been associated with syntactic anomaly (Hagoort, Brown, & Groothusen, 1993), and more recently, with additional cognitive effort to establish a representation of the intended meaning (Brouwer et al., 2012). In irony studies, the P600 is typically larger for irony than for literal language, indexing reanalysis or inferential processing to arrive at the intended meaning.

The second component of interest in irony research is an increased N400 for irony compared to literal has been associated with unfamiliar forms of irony in written (Filik et al., 2014) and auditory tasks (Caillies et al., 2019). Additionally, Baptista and colleagues (2018) associated an irony N400 with “a difficulty to easily access the meaning”. They used transcranial direct current stimulation (tDCS) to either stimulate or inhibit the medial prefrontal cortex (mPFC), a structure involved in mentalizing/perspective taking. While behavioral accuracy of irony recognition was not impacted by tDCS, they found an increased N400 for irony compared to literal in the inhibition and sham (no stimulation) groups, but not in the stimulation group. Thus, they concluded that mentalizing capability impacts the difficulty of meaning retrieval for irony processing at N400. Note that this study used picture context, not linguistic context. The

N400 sometimes precedes the P600 effect, as was the case for unfamiliar ironic comments in Filik et al. (2014), suggesting that N400 and P600 index different aspects of ironic language processing.

Two recent studies investigated the relationship of irony and emotion. Thompson, Leuthold, and Filik (2021) auditorily presented short stories containing either literal or ironic criticism (e.g. *You're so strong!*) that ended with a hurt or amused reaction from the victim (e.g. *Charlie felt this was a very mean/amusing thing to say*), or speaker intention (e.g. *Ray had intended for this to be a very mean/amusing thing to say*). They found that ironically pronounced irony elicited a larger ERP effect than irony without ironic pronunciation, which the authors attributed to a faster detection of irony. Pfeifer and Lai (2021) investigated the processing of irony in highly negative and less negative situations using written contexts. They found a larger LPC (500-800 ms) for irony than literal in highly negative contexts, but not in mildly negative situations. They suggested that the irony-literal difference becomes more relevant in highly negative situations, which involves continued processing of the speaker's emotion or the context.

Past studies are limited in several ways. First, most situated the participant as an uninvolved bystander and asked them to rate fictional characters. Such tasks may be confounded by participants' perspective-taking skills. Second, researchers probed a variety of scales, including politeness, negativity, humor, and others. While these ratings provide insights on the outcomes of the meaning making process, online measures may enlighten us about the mechanism(s) by which irony affects negativity. Lastly, most studies conveyed context via narratives. While verbal irony has to be verbal by definition, the context can be multimodal, even in laboratory studies. A rich and detailed visual context is closer to real-life situation, increasing ecological validity and conveys more specific information than language alone.

1.2. Emotion Regulation via Cognitive Reappraisal

Emotion regulation is when we change which emotions we have, when we experience them, and how we experience and express them (Gross, 1998). Cognitive reappraisal is a form of emotion regulation (Gross, 1998), where a cognitive change alters the incoming emotional cues with the goal of changing one's emotional response. For example, to reduce negative emotion when seeing news coverage of an accident, one may use cognitive reappraisal by thinking that things are not as bad as they seem on screen (McRae, Ciesielski, & Gross, 2012).

ERP studies of emotion regulation have primarily focused on the P300 / Late Positivity Potential (LPP), which are typically reduced following emotion regulation compared to no instruction, or attending to emotions (Hajcak, MacNamara, & Olvet, 2010). Hajcak & Nieuwenhuis (2006) for example found a reduced LPP for cognitive reappraisal, compared to *attend*, a condition in which the participants do not alter their natural feelings about the pictures. In each trial, participants saw a negative picture, followed by either *attend* or *reinterpret* prompts and a second presentation of the same image. After image offset, participants rated their emotional response on a 1 (weak) – 4 (strong) scale. Behavioral responses showed a significant reduction in emotional intensity for *reinterpret* compared to *attend*. ERPs time-locked to the onset of the second presentation of the image showed a reduced LPP for *reinterpret* compared to *attend*. The LPP effect started at ~ 200 ms, was sustained until picture offset, and had a centro-parietal scalp distribution. According to Hajcak & Nieuwenhuis (2006), the early onset of the LPP indicates that the neural response to reappraisal is early, while the long-lasting LPP effect indexes a continued reduction in emotional experience. In our opinion, though, cognitive efforts

to reappraise a seen image would begin at the instruction prompt already, and not wait until the second presentation of the image 5 seconds after.

1.3. Emotion Regulation via Language

A number of studies have proposed that language can be used to regulate emotions via *affect labeling*. For example, Torre & Lieberman (2018) found that participants who label the emotional content of a stimulus report a reduction in their emotional response compared to those who don't label emotions. They proposed that affect labeling is a form of implicit emotion regulation that resembles cognitive reappraisal, and relies on several mechanisms, including symbolic conversion, where abstracting from an emotion to language reduces emotional experience. However, most *affect labeling* studies use single word labels for basic emotion (e.g., anger/angry, happy), disregarding more complex emotional and linguistic expressions. Verbal irony conveys mixed or ambiguous emotion (Pfeifer & Pexman, 2023), and could expand the traditional affect labeling paradigm by expressing more complex emotions, and doing so in more conversational, i.e. naturalistic manner. Irony induces contrast to a situation and provides a new perspective, which likely affects either *attention deployment* or *cognitive change*, both of which are antecedent focused strategies of emotion regulation (c.f. Gross, 1998). However, simply focusing attention away from a negative event (distraction) could potentially be accomplished by any verbal interaction (e.g., literal language). Thus, we think that irony induces a cognitive change, which then leads to a reappraisal of the situation.

1.3. Current study and predictions of ERP correlates

The current study took a novel approach to explore the potential involvement of irony in the regulation of emotions. We investigated the effects of irony and cognitive reappraisal behaviorally and neurally, to: (1) determine if irony effectively reduces negative feelings and (2) compare the neural signatures of irony-induced cognitive change and cognitive reappraisal. Behaviorally, we a priori hypothesized that *irony* would significantly reduce ratings of negativity compared to *literal* statements (Pfeifer & Lai, 2021), and that *cognitive reappraisal* would significantly reduce ratings of negativity compared to simply *attending* to the image (Hajcak & Nieuwenhuis, 2006). On a neural level, following previous literature, we predicted that *irony* would elicit a larger LPC than *literal* language at the irony/literal word-onset. We also predicted differences between *cognitive reappraisal* and *attend* at the second image presentation, with *reappraise* eliciting a reduced LPP compared to *attend*. If cognitive reappraisal was a key mechanism of irony, we expected to see a similar neural signature during the second image presentation for *irony* compared to *literal*, namely a reduced LPP. Regarding N400, we did not pose any strong predicted outcomes based on limited existing literature. But it would not be impossible to see an increased N400 for irony compared to literal. Post-hoc, we concluded that it was reasonable to expect a larger N400 for ironic compared than literal words, based on irony and picture combinations being unfamiliar (c.f. Filik et al., 2014), and the N400 being sensitive to irony in pictorial contexts (Baptista et al., 2018).

2. Methods

2.1. Participants.

Fifty-four right-handed native speakers of English (Mean Age: 19.07, $SD = 1.07$ years, 13 males) participated for course credit. None reported neurological or developmental

abnormalities, psychoactive medication use, and all had normal or corrected-to normal vision. While EEG data can be easily affected by noise, behavioral data remains unaffected, so we used all 54 participants in the behavioral analyses. Eleven participants were excluded from the EEG analysis due to excessive noise (>40% trial loss, or less than 10 trials in one condition). The study was approved by the local Institutional Review Board.

2.2. Materials and Procedure.

The study used a within-participants block design with a verbal (irony /literal) and a non-verbal (cognitive reappraisal/attend) block, counterbalanced with participant number.

Materials were 128 mildly negative pictures (e.g., flat tire, burnt food), each paired with literal/ironic statements (“Pretty flat/inflated!”, “Looks tasty/charred!”). A list of image descriptions, statements, and their norming results can be found in the study repository: https://osf.io/cb4k8/?view_only=99510c323a374dd89c9c9260e8c5c13b and a full set is available from the first author upon request. The ironic and literal words were matched for psycholinguistic properties (e.g., word frequencies and lengths) and pictures were normed for emotion to verify manipulation (see details and statistics in Appendix). Note that we did not match words in terms of valence and arousal, due to the inherently contrastive nature of irony (see Discussion). On the 9-point Likert-type scales in the English Lexicon Project database (Balota et al., 2007), literal words (mean rating=3.36) were significantly more negative than ironic ones (mean rating=6.62), and slightly more arousing (mean rating literal: 4.5; mean rating irony: 4.1). A paired t-test confirmed that the conditions further differed significantly in terms of irony ($p < .001$).

Participants gave written consent and filled out a language background questionnaire to confirm they met the inclusion criteria. They then were fitted the electrode cap and seated in a sound-attenuated booth, approximately 80 cm away from a 22-inch LCD screen. They were instructed to imagine that the situation in each picture was happening to them, to not look away, and to not generate another emotion to replace their emotional response. The experiment was deployed using EPrime 3.0 (Psychology Software Tools, Pittsburgh, PA, USA). Images did not exceed 500 pixels in width and 400 pixels in height, and text appeared centrally in white serif font (Consolas, size 18) on a black background.

Figure 1A shows the timing structure of the trials. In each trial, the image appeared for 2 seconds, followed by either the critical statement presented word-by-word (in the verbal block), or the instruction screen indicating they should either *attend* to or *reinterpret* the image (in the non-verbal block). Then, the image reappeared for 3 seconds, followed by the rating task of “How negative to you feel? 1 = weak, 4 = strong” to which participants responded via button presses

2.3. EEG Data processing and analyses.

Continuous EEG was recorded from 32 active electrodes (see Figure 1C). Offline, data were pre-processed using Brain Vision Analyzer 2.0, described in the Appendix. Briefly, data were band-pass filtered at 0.01 and 30 Hz, ocular correction was applied, and data were re-referenced to the average of the left and right mastoids. Then, data were segmented and those with a total Max-Min difference of 100 μ V were rejected. All remaining segments were normalized to a 200 ms pre-stimulus baseline before they were averaged by participant and condition.

Next, we extracted mean amplitudes for the following time windows in two regions of interest, frontal (Fz, FC1, FC2) and parietal (Pz, CP1, CP2) as shown in Figure 1C. In the verbal block, we analyzed N400 (300 – 500 ms) and LPC (600 – 900 ms) at the critical words, which is the word that rendered a statement ironic or literal. We examined ERPs at the second image onset, but did not observe visual components. Thus, we analyzed ERPs for the instructions screen in the non-verbal block: N400 (300 – 550 ms) and LPP (800 – 1500 ms).

3. Results

3.1. Behavioral results.

Ratings of participants' emotional state during the experiment were averaged across items per condition for each participant (Figure 1B). A RM-ANOVA of 2 condition (irony/cognitive reappraisal, literal/attend) X 2 modality (verbal, nonverbal) with block order as a between factor, investigated if differences in negativity between *literal* and *irony* were different from those between *reinterpret* and *attend*. We found a significant main effect of condition ($F(2,52) = 116.65, p < .001, \eta^2 = .69$), a main effect of modality ($F(2,52) = 90.92, p < .001, \eta^2 = .64$), and a significant interaction of condition and modality ($F(2,52) = 98.73, p < .001, \eta^2 = .66$). Paired-t-tests within each modality showed that *ironic* statements ($M = 2.58, SD = 0.44$) made participants feel significantly less negative compared to *literal* statements ($M = 2.68, SD = 0.4$), ($t(53) = -2.309, p = .0249, d = -0.31$), and that *reinterpret* also made participants feel significantly less negative ($M = 1.88, SD = 0.48$) than *attend* ($M = 2.74, SD = 0.48$), ($t(53) = -11.631, p < .001, d = -1.58$) (Figure 1B). The interaction means that the *reinterpret-attend* difference (0.86) was significantly greater than the *irony-literal* difference (0.1). No effect of block order was present.

3.2. ERP results.

On average, 28.69 segments were kept in each condition, which corresponds to 89.66% of all trials. Average trial counts per condition (out of 32) were as follows: Verbal Block: Image before: 29.1 (literal), 29 (irony). Image after: 28.14 (literal), 28.56 (irony). Critical word: 29.67 (literal), 29.3 (irony). Non-verbal block: Image before: 28.91 (attend), 28.3 (reinterpret). Image after: 28.42 (attend), 28.47 (reinterpret). Instruction: 29.28 (attend), 28.91 (reinterpret). Grand averaged ERPs are presented in Figure 1D. Data were entered into a 2-factor Repeated Measures ANOVA of 2 manipulation (irony/reinterpret, literal/attend) by 2 location (frontal, parietal) with block order as a between factor. Full statistical results are listed in Table 1.

[Figure 1; Table 1]

For the verbal block, at the critical word, *irony* elicited a larger N400 (300 – 500 ms) than *literal* ($p < .001$). Following the significant Condition x Location interaction ($p = .022$), pairwise comparisons within each Location revealed that such irony-literal N400 difference was significant in both the frontal ($F(1,42) = 15.19, p < .001, \eta^2 = .266$) and parietal channels ($F(1,42) = 24.85, p < .001, \eta^2 = .372$). Irony also elicited a larger prolonged negativity than *literal* in the LPC (600 – 900 ms) time window ($p = .012$). Following the significant Condition x Location interaction ($p = .010$), the pairwise comparisons revealed irony-literal LPC difference was not significant frontally ($F(1,42) = 2.77, p = .104, \eta^2 = .062$), but was significant parietally ($F(1,42) = 11.764, p = .001, \eta^2 = .219$). No significant effects of block order were observed for either time window.

For the second image presentation, counter our predictions, there were no statistically significant differences, neither between *attend* and *reinterpret*, nor between *literal* and *irony*. (Table 1).

We reasoned that participants might have started to attend or reinterpret as soon as they saw the instruction (*attend* or *reinterpret*) screen. Analyzing ERPs at the instruction (*attend* or *reinterpret*) screen for the non-verbal block, we found a significantly larger N400-like effect (300 – 550 ms) for *attend* compared to *reinterpret* ($p = .022$) across the scalp. This effect interacted with block order ($p = .041$), and using Bonferroni correction, in pairwise comparisons within presentation order, we found that the N400-like effect was only present in those who saw the verbal block first, in both frontal ($F(1,19)=5.35, p = .027, M_{attend} = -1.59, SD = 3.77; M_{reinterpret} = 0.39, SD = 3.82$) and parietal channels ($F(1,19)=15.051, p < .001, M_{attend} = -1.39, SD = 3.09; M_{reinterpret} = 1.21, SD = 3.63$). In addition, we found a significantly larger positivity / reduced negativity (800 – 1500 ms) for *reinterpret* compared to *attend* in the LPP time window (800 - 1500 ms) ($p = .032$). Pairwise comparisons following the significant Condition X Location interaction ($p = .032$) at LPP revealed a significant difference of *attend* and *reinterpret* in the frontal channels ($F(1,42) = 6.70, p = .013, \eta^2 = .138$), but not the parietal channels ($F(1,42) = 2.44, p = .126, \eta^2 = .055$). No effect of block order was observed for the LPP time window.

4. Discussion

We set out to compare the impact of irony and cognitive reappraisal on negative emotion on a behavioral and a neural level, to examine whether irony acts similarly to cognitive reappraisal. Behaviorally, while both irony and cognitive reappraisal significantly reduced how negative participants felt, cognitive reappraisal was more effective than irony in reducing

negativity, comparable to prior findings from affect labeling (Torre & Lieberman, 2018). Neurally, *irony* at the target word elicited an N400 (300-500 ms) and a prolonged negativity in the LPC time window (600-900 ms), compared to *literal* statements. No difference between conditions were observed at second image presentation. However, at the instruction prompt in the non-verbal block, cognitive reappraisal (*reinterpret*) showed a reduced N400 (300-550 ms) and a larger frontal LPP (800-1500 ms) compared to *attend*.

4.1. Verbal block: irony vs. literal

Irony elicited a broad N400 along with a prolonged negativity compared to literal statements, and no LPC. This finding was in part unexpected, because while some reported increased N400 amplitudes for irony (c.f. Filik et al., 2014; Thompson, Leuthold, & Filik, 2021; Caillies et al., 2019; Baptista et al., 2018), most found an additional, increased positivity (P600 / LPC). Our results are in line with Pfeifer and Lai (2021), who did not find P600 in the mildly negative condition, which may be equivalent to the mildly negative pictures used here.

The ERP waveforms allow for two possibilities: (1) a single N400 effect but prolonged, (2) an N400 effect followed by a late negativity. While the present data cannot differentiate between the two possibilities, we believe that a single N400 effect, due to the unexpectedness of the ironic word and the continued process reducing negative emotion, is the most likely scenario.

First, our effect could be a prolonged N400 effect, reflecting (semantic) incongruity between visual and verbal modalities. Unlike past studies that mostly used verbal context, our study used picture stimuli as the context for verbal irony. Picture context may have created rich and precise expectations for upcoming words, pre-activating a very specific, descriptive meaning

and resulting in additional processing costs for ironic words. Picture context and specificity have been examined in other studies using literal language. For instance, in Dikker & Pylkkänen (2011), participants saw a specific image (apple/banana) or a general image (grocery bag/zoo) first, followed by a target word (e.g., “apple”). Image-word mismatch elicited the equivalent of N400 in MEG responses, only when the image prime was specific, not when it was general. Applying this to the current study, the image context in our study was more specific than the verbal context in past irony studies. Additionally, prior work that used a verbal context essentially left the specificity to the participants’ imagination, which might explain the lack of consensus on the involvement of N400 in irony processing. Moreover, Baptista et al. (2018) also used pictorial contexts for verbal irony and reported N400 effects. Specifically, they found a difference for literal and ironic language at N400, which was absent when mPFC was stimulated via tDCS, suggesting that semantic retrieval (indexed by N400) is impacted by mentalizing abilities (manipulated by tDCS on mPFC). Our results provide support for the mediating role of semantic retrieval in irony processing, particularly in the context of images. However, neither Dikker & Pylkkänen (2011) nor Baptista et al. (2018) observed a prolonged effect like our present negativity, suggesting that neither specificity, nor mentalizing ability alone can explain our prolonged effect. Ultimately, we suggest that our negativity initially indexes the contrast between the visually presented scene and the affective content of the target word and is subsequently sustained for the continued processing of the word image combination, which results in the experience of irony and the reduced ratings of negativity downstream.

Secondly, it is possible that our N400 is followed by a Late Negativity (LN). LN has been associated with humor processing during joke comprehension (Coulson & Kutas, 2001), which might be an additional layer of meaning within the ironic statements. While not typically found

for ironic statements, we conducted additional humor ratings ($N = 28$ participants) to explore this possibility and found that while ironic statements were slightly funnier than literal statements (a 0.54-point difference on a 5-point scale), they were overall not rated as being “very funny” (irony: $M = 1.33$, literal: $M = 0.78$ with 4 = “very funny”), meaning our items did not cover a wide range on the humor scale and were not very humorous. Thus, we believe that this explanation is less likely than a single, prolonged N400 effect.

There is a possibility that inherent affective differences between words used ironically and words used literally might modulate the N400. Specifically, as reported in Methods, without context, words used in the literal condition were significantly more negative and more arousing than words used in the ironic condition. However, we do not think that affective word differences alone are sufficient to explain our N400 effects. To begin with, studies on emotional words embedded in sentences do not always find N400 effects (Schacht & Sommer, 2009; Fields & Kuperberg, 2016; but see Holt et al., 2009). Furthermore, word valence typically modulates the Early Posterior Negativity (EPN, Kissler et al., 2009; Schacht & Sommer, 2009; Bayer & Schacht, 2014) and word arousal or task demands modulates the LPC (Delaney-Busch, Wilkie, & Kuperberg, 2016; see Lai et al., (to appear), for more extensive discussion). In contrast, our effects start at the N400 time window and continue until 900ms, which makes it unlikely that affective word differences *alone* account for our effect.

4.2. Second image presentation

We did not observe any notable ERP differences at our second image presentation in either the verbal or the non-verbal block. We can interpret the absence of findings at the second image presentation in two ways. Either our pictures were not negative enough to induce negative

emotional responses in participants, meaning there are indeed no differences between the conditions, or the effects of emotion regulation are not captured at this point in time.

Regarding the first interpretation, our images (e.g., a dropped cake) depicted less negative and less arousing situations than those used in prior studies (e.g., amputation). Thus, our participants may have reinterpreted the images, but not felt the need to actively maintain such reinterpretation. Note that we explicitly made the design choice of not including images as negative as those used in prior studies of emotion regulation, because using strongly negative images such as mutilated bodies or burning houses could have impacted effects in the irony block, as irony is considered socially inappropriate in the context of highly negative events. Thus, our methodological choice may have inadvertently led to the lack of effect in the second image presentation.

The second interpretation is that effects occur at different timepoints. It is possible that our data do not display differences at the second image presentation, simply because differences already started earlier, possibly during the instruction screen (see section 4.3), and therefore, conditions do not visually appear different at this later stage. Indeed, there is some uncertainty around when in time cognitive reappraisal starts, because prior work on emotion regulation did not report ERPs for the instruction prompts. However, it is not unreasonable to assume that participants started following instructions and regulating their emotional responses at the instruction prompt, but further research into the timing of cognitive reappraisal is needed.

4.3. Non-verbal block: reinterpret vs. attend

Because prior studies did not report ERPs to the instruction onset (e.g., Hajcak & Nieuwenhuis, 2006) and typically time-locked ERPs to the second image presentation (c.f.

Hajcak, McNamara, & Olvet, 2010), we cautiously offer possible interpretation of our findings at the instruction screen below. Note that some may consider these speculative.

The frontal N400-like larger negativity for *attend* compared to *reinterpret* is likely not a language or meaning related N400, since participants saw the same prompt word repeatedly. This effect also interacted with block order, showing that it was only present in participants who saw the non-verbal block second, which further points to a demand or task difference.

Further, *reinterpret* elicited a larger positivity / reduced negativity (800 – 1500 ms) than *attend*, most pronounced in frontal channels. We cautiously suggest that this frontal effect reflects cognitive engagement and mental simulation/imagery (Schendan & Ghanis, 2012). In essence, our participants used more mental imagery during the *reinterpret* screen than the *attend* screen, which would also explain the frontal distribution of the effects. They likely recalled details from the image to construct their reappraisal strategy.

A second interpretation is that instead of showing an increased negativity for *attend* at N400, we might be seeing an increased positivity for *reinterpret* at both timepoints. Previous research has associated the earlier part (P300) of the late positivity (LPP) in the context of emotion regulation with attention (Hajcak, MacNamara & Olvet, 2010), so perhaps the *reinterpret* condition required more attentional resources than the *attend* condition, likely because participants had to recall details of the image from its first presentation in order to successfully reinterpret the situation when prompted (c.f. Gibbons et al., 2018). This, together with the vivid mental image participants sustain at LPP, could indicate that participants indeed begin to use the cognitive reappraisal strategy as soon as the instructions appear on the screen. The presence of sustained effects at the instruction screen points to cognitive, rather than a perceptual difference between the *attend* and *reinterpret* cues. However, due to the lack of prior

work reporting ERPs relative to the instruction screen, we cannot make any firm conclusions about the functional interpretation of the ERPs here.

4.4. General Discussion

Our behavioral data show that irony reduces feelings of negativity in recipients compared to literal language, which is consistent with behavioral ratings in Pfeifer & Lai (2021), in line with the *tinge hypothesis* (Dews & Winner, 1995) and a recent view of irony as a socio-emotional skill that provides cognitive benefits (Pfeifer & Pexman, 2024). Comparing the verbal and non-verbal blocks, irony and cognitive reappraisal both downregulate negativity with a different magnitude, as reflected by the rating results. This is congruent with prior findings from affect labeling wherein affect labeling is effective in reducing emotional experience, but not as effective as cognitive reappraisal. Torre & Lieberman (2018) proposed that this is because while cognitive reappraisal is an explicit emotion regulation strategy, affect labeling is an implicit strategy. Our study expands this prior work by suggesting that verbal irony can also be effective in reducing felt negativity. Moreover, our novel approach offers new insights into the cognitive mechanisms that govern emotional experiences and poses important questions that future studies combining language and emotion regulation could address.

At the beginning of the study, we considered the similarities between irony and cognitive reappraisal, but we also acknowledge that there are differences between irony and cognitive reappraisal. For example, irony and cognitive reappraisal differ in the amount of agency or involvement of the participant. Perceiving an ironic comment puts the participants in a passive state, whereas engaging in cognitive reappraisal puts participants in an active state. Relatedly, cognitive reappraisal is voluntary, whereas irony processing is involuntary. A recipient cannot

avoid processing a contrast conveyed in irony, once introduced. Therefore, a recipient of an ironic comment cannot easily avoid receiving the benefits of such contrast in the form of reduced negative emotions. Our behavioral findings suggest that irony can be a tool for speakers to regulate listener's emotions, one that crucially does not require active engagement of the recipient. Lastly, cognitive reappraisal involves creating alternative cognitions, which requires more creativity and mental imagery than irony processing, which is more passive and involuntary. Actively regulating one's emotions is likely more effortful than passive, involuntary processing of ironic comments, and what is effortful might result in stronger effects downstream. Based on these differences, we acknowledge that our rating results could also represent some alternative picture interpretation strategy other than cognitive reappraisal.

The present study suffers from a number of limitations. Most notably, in contrast to our expected outcomes, we did not observe any effects at the second image presentation, which means we are unable to draw definite conclusions about our initial aim, whether irony and cognitive reappraisal rely on similar neural mechanisms, at least not at the time point when the image appeared for the second time. This absence of effect may be explained by a number of methodological decisions we made in this study, including the use of written instructions and mildly negative images, for the purpose of comparing the verbal block and the non-verbal block.

Our images were mildly negative and moderately arousing (e.g., dropped food, flat tire), compared to the strongly negative and highly arousing images (e.g., mutilated bodies, burning buildings) used in other studies of emotion regulation. However, we believe it would be inappropriate to use irony in such strongly negative situations, e.g., mutilated bodies or burned buildings. Future studies could try to overcome this limitation by using more negative images and risking inappropriateness of irony, or using verbal descriptions, a novelty in emotion

regulation research. Second, we used written irony, and written strategy cues, presented as words on the screen. Such presentation may have led to additional visual artifacts, processing costs, or in other ways impacted our ability to compare the neural signature of irony and cognitive reappraisal.

The present data indicated that there are some similarities and differences on the behavioral level between irony and emotion regulation. It presents a first step into exploring *how* irony reduces negativity, based on physiological data from participants experiencing negative emotions and encountering irony, which is an important and novel change to prior work.

We observed both similarities and differences between irony and cognitive reappraisal. Perhaps one way to think about our findings in the context of prior studies is that in mildly negative situations, using irony to deter negativity is possible, but in very negative situations, using cognitive reappraisal is more powerful. In addition, our results seem to suggest that irony affects emotions in those encountering it, not (only) those producing it, meaning that in mildly negative situations, one can regulate someone else's emotions with irony.

Acknowledgements: We thank Dr. Matthias Mehl for his comments on an earlier version of this article.

Declaration of Interest: The authors have no conflicts of interest to declare.

5. References

- Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, 39, 445-459. <https://doi.org/10.3758/BF03193014>
- Baptista, N. I., Manfredi, M., & Boggio, P. S. (2018). Medial prefrontal cortex stimulation modulates irony processing as indexed by the N400. *Social neuroscience*, 13(4), 495-510. <https://doi.org/10.1080/17470919.2017.1356744>
- Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces—a cross-domain comparison. *Frontiers in psychology*, 5, Article 1106. <https://doi.org/10.3389/fpsyg.2014.01106>
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, 1446, 127-143. <https://doi.org/10.1016/j.brainres.2012.01.055>
- Caillies, S., Gobin, P., Obert, A., Terrien, S., Coutté, A., Iakimova, G., & Besche-Richard, C. (2019). Asymmetry of affect in verbal irony understanding: What about the N400 and P600 components? *Journal of Neurolinguistics*, 51, 268-277. <https://doi.org/10.1016/j.jneuroling.2019.04.004>
- Citron, F. M. (2012). Neural correlates of written emotion word processing: A review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and language*, 122(3), 211-226. <https://doi.org/10.1016/j.bandl.2011.12.007>
- Coulson, S., & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience letters*, 316(2), 71-74. [https://doi.org/10.1016/S0304-3940\(01\)02387-4](https://doi.org/10.1016/S0304-3940(01)02387-4)

- Dews, S., & Winner, E. (1995). Muting the meaning a social function of irony. *Metaphor and Symbol, 10*(1), 3-19. https://doi.org/10.1207/s15327868ms1001_2
- Delaney-Busch, N., Wilkie, G., & Kuperberg, G. (2016). Vivid: How valence and arousal influence word processing under different task demands. *Cognitive, Affective, & Behavioral Neuroscience, 16*(3), 415–432. <https://doi.org/10.3758/s13415-016-0402-y>
- Dikker, S., & Pyllkanen, L. (2011). Before the N400: Effects of lexical–semantic violations in visual cortex. *Brain and Language, 118*(1-2), 23-28.
<https://doi.org/10.1016/j.bandl.2011.02.006>
- Fields, E. C., & Kuperberg, G. R. (2016). Dynamic effects of self-relevance and task on the neural processing of emotional words in context. *Frontiers in Psychology, 6*, Article 2003.
<https://doi.org/10.3389/fpsyg.2015.02003>
- Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 811-828. <https://psycnet.apa.org/doi/10.1037/a0035658>
- Gibbons, H., Seib-Pfeifer, L.-E., Koppehele-Gossel, J., & Schnuerch, R. (2018). Affective priming and cognitive load: Event-related potentials suggest an interplay of implicit affect misattribution and strategic inhibition. *Psychophysiology, 55*(4), Article e13009.
<https://doi.org/10.1111/psyp.13009>
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of general psychology, 2*(3), 271-299. <https://doi.org/10.1037/1089-2680.2.3.271>
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes, 8*(4), 439-483.
<https://doi.org/10.1080/01690969308407585>

- Hajcak, G., & Nieuwenhuis, S. (2006). Reappraisal modulates the electrocortical response to unpleasant pictures. *Cognitive, Affective, & Behavioral Neuroscience*, 6(4), 291-297.
<https://doi.org/10.3758/CABN.6.4.291>
- Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental neuropsychology*, 35(2), 129-155.
<https://doi.org/10.1080/87565640903526504>
- Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of Cognitive Neuroscience*, 21(11), 2245–2262. <https://doi.org/10.1162/jocn.2008.21151>
- Ivanko, S. L., & Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse processes*, 35(3), 241-279. https://doi.org/10.1207/S15326950DP3503_2
- Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biological Psychology*, 80(1), 75–83.
<https://doi.org/10.1016/j.biopsycho.2008.03.004>
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, 62, 621-647. <https://doi.org/10.1146/annurev.psych.093008.131123>
- Lai, V. T., Pfeifer, V., & Ku, L. C. (to appear). Emotional language processing: An individual differences approach.
- Lai, V. T., Hubbard, R., Ku, L. C., & Pfeifer, V. (2023). Electrophysiology of non-literal language. In: M. Grimaldi, E. Brattico, & Y. Shtyrov (Eds.), *Language electrified: Principles, methods, and future perspectives of investigation*, pp. 613-646. Humana. https://doi.org/10.1007/978-1-0716-3263-5_19

- McRae, K., Ciesielski, B., & Gross, J. J. (2012). Unpacking cognitive reappraisal: Goals, tactics, and outcomes. *Emotion, 12*(2), 250-255. <https://psycnet.apa.org/doi/10.1037/a0026351>
- Morawetz, C., Bode, S., Derntl, B., & Heekeren, H. R. (2017). The effect of strategies, goals and stimulus material on the neural mechanisms of emotion regulation: A meta-analysis of fMRI studies. *Neuroscience & Biobehavioral Reviews, 72*, 111-128.
<https://doi.org/10.1016/j.neubiorev.2016.11.014>
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language, 31*(6), 785-806. [https://doi.org/10.1016/0749-596X\(92\)90039-Z](https://doi.org/10.1016/0749-596X(92)90039-Z)
- Pfeifer, V. A., & Lai, V. T. (2021). The comprehension of irony in high and low emotional contexts. *Canadian Journal of Experimental Psychology, 75*(2), 120–125.
<https://doi.org/10.1037/cep0000250>
- Pfeifer, V. A., & Pexman, P. M. (2023). Mixed and ambiguous emotions can be studied with verbal irony. *Cognitive Neuroscience, 14*(2), 65-67.
<https://doi.org/10.1080/17588928.2023.2181320>
- Pfeifer, V. A. & Pexman, P. M. (2024). When it pays to be insincere: On the benefits of verbal irony. *Current Directions in Psychological Science, 33*(1), 43-50.
<https://doi.org/10.1177/09637214231205312>
- Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language? *Psychological Science, 5*(3), 159-163. <https://doi.org/10.1111/j.1467-9280.1994.tb00653.x>
- Schacht, A., & Sommer, W. (2009). Time course and task dependence of emotion effects in word processing. *Cognitive, Affective, & Behavioral Neuroscience, 9*(1), 28–43.
<https://doi.org/10.1016/j.bandc.2008.11.005>

- Schendan, H. E., & Ganis, G. (2012). Electrophysiological potentials reveal cortical mechanisms for mental imagery, mental simulation, and grounded (embodied) cognition. *Frontiers in Psychology*, 3, 329. <https://doi.org/10.3389/fpsyg.2012.00329>
- Thompson, D., Leuthold, H., & Filik, R. (2021). Examining the influence of perspective and prosody on expected emotional responses to irony: Evidence from event-related brain potentials. *Canadian Journal of Experimental Psychology*. 75(2), 107–113. <https://doi.org/10.1037/cep0000249>
- Torre, J. B., & Lieberman, M. D. (2018). Putting feelings into words: Affect labeling as implicit emotion regulation. *Emotion Review*, 10(2), 116-124. <https://doi.org/10.1177/1754073917742706>