The Comprehension of Irony in High and Low Emotional Contexts

Valeria A. Pfeifer and Vicky T. Lai

Department of Psychology, University of Arizona, Cognitive Science Department, University of Arizona

Author Note

Portions of these findings were presented as a poster at the 32rd Annual CUNY Sentence Processing Conference in Boulder, Colorado, United States. We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Valeria A. Pfeifer, Department of Psychology, University of Arizona, 1503 E University Blvd, Tucson, AZ 85719, United States. Email: vpfeifer@email.arizona.edu

Abstract

Verbal irony are words that intend the opposite of their literal meaning. We investigated the emotional function of irony by asking whether irony intensifies or mitigates negative feelings. Experiment 1 used ratings to assess the mental state of a speaker using irony or literal language following a negative event in either a high or a low emotional context. We found that regardless of context emotionality, speakers using irony were perceived as being in a less negative and less aroused mental state than speakers using literal language. In Experiment 2, we examined the time course of this process with ERPs. Initially, literal statements elicited a larger N100 than irony regardless of context emotionality, suggesting that irony mitigates negative feelings overall. Later on, irony elicited a larger LPC than literal statements in high emotion contexts, but not in low emotion contexts. This suggests that irony required more mental state processing or/and more speaker emotion processing than literal language in emotionally loaded situations. These results indicate that whether irony intensifies or mitigates negative feelings depends on context and the point in time at which we assess its function.

Keywords: Emotion, ERP, Figurative Language, N100, LPC

Public Significance statement:

Using brainwave and behavioral measures, we found that in a negative situation, people initially find literal statements more threatening and irony more difficult to process. After they have a second to integrate and re-analyze semantic, pragmatic, and emotional information, they think that the person using irony is less negatively impacted by the emotional situation. This study contributes to a broader understanding on the interaction between emotion and language.

The Comprehension of Irony in High and Low Emotional Contexts

Verbal irony is used to communicate the opposite of what is being said – e.g., 'How careful of you' if someone spills a drink. What motivates a speaker to actively choose irony? Most people (94%) believed that they use irony to convey negative emotions (Roberts & Kreuz, 1994). But it is unclear whether and how irony accomplishes this. Here we investigated a factor that has received little attention so far: the emotional intensity of the situation. Specifically, we used ratings and event related potentials to test whether a speaker using irony is perceived to be in a different mental state compared to a speaker using literal language in more and less emotional situations.

Two opposing views about the function of irony have been postulated. The *tinge hypothesis* (Dews & Winner, 1995) posits that the use of irony decreases the emotional impact of the statement, such that it mutes criticism. They showed that a speaker using irony is perceived as tactful, because irony is rated less critical than literal language. Further support comes from an eye-tracking study by Filik et al. (2017). Participants read literal or ironic statements, followed by a description of either the protagonist's intention 'Ray intended for this to be a very …' or the victim's feeling 'Charlie felt that this was a very …'. In the continuation 'mean thing to say', the critical word 'mean' exemplified the protagonist's hurting intention or the victim's hurt response. Facilitated reading times at 'mean' in the ironic context indicated that readers expected hurt responses/intentions. Interestingly, in the post-critical region ('thing to say'), it was harder to integrate a hurt response in ironic than in literal contexts. When the critical word 'mean' was replaced with 'amused', no difference was found at the critical word, but following irony, readers expected amusing intention in the post-critical-word region. The authors proposed a two-stage process: Initially, a hurt response is more congruent with the readers' hypothesis about the

victim's emotional state. But later on, the amusing intent is easier to integrate. One caveat is that there are multiple referents, which might have increased working memory load and made it hard for perspective taking.

Opposed to this stands the view that irony enhances criticism and stings as it increases condemnation (Colston, 2002). The *contrast and assimilation theory* (CAT) suggests that the amount of contrast between what is said and what is true determines how negative irony appears. A stark contrast makes irony more negative, whereas a weak contrast makes irony less negative through assimilation. In Colston (2002), participants rated how bad a situation was, how condemning the speaker was, and how personally affected they as readers were by a situation, in varying degrees of situational negativity. Results demonstrated that how condemning the speaker was corresponded to the speaker's intent to make the situation appear negative. However, the study did not address the speaker's emotional state. Ivanco and Pexman (2003) also provided support for CAT. They manipulated the degree of incongruity between the situation and the ironic utterance by varying situational negativity and assessed readers' interpretation of speaker intent with ratings and reading times. In strongly negative contexts, irony was read slower and was rated more mocking and sarcastic than literal. In contrast, in weak negative contexts, irony was read faster and was rated more polite. Boylan and Katz (2013) found that irony simultaneously mutes and enhances negativity. Participants read short vignettes that contained either ironic or literal criticisms and evaluated several aspects of the vignettes. It was more polite and positive to use irony than literal statements, yet simultaneously more mocking and sarcastic. Behavioral literature so far suggests no consensus as to how irony impacts negativity.

In Event-related potentials (ERPs) studies of irony, one of the first (Katz, Blasko & Kazmerski, 2004) contrasted utterances following contexts that biased a literal or sarcastic

interpretation. At the utterance-final words, sarcasm elicited a more negative-going potential at ~650 ms and a more positive-going wave at ~900 ms. The authors suggested that the effect at ~650 reflects either integration difficulty or emotion and the effect at ~900 ms reflects humor. In the following years, most irony studies reported an irony-related P600/Late Positive Component (LPC), a positive-going potential peaking at ~600 ms after stimulus onset, centrally distributed, and larger for irony than literal language (Caillies et al., 2019; Filik et al., 2014; Regel et al., 2010). In psycholinguistics generally, P600 has been associated with the need to re-analyze a statement after the initial interpretation fails (Kuperberg, 2007), while in irony, it has been associated with integrating linguistic and world knowledge after rejecting the initial interpretation (Regel et al. 2010). In emotion, LPC is increased for emotional compared to non-emotional content (Citron, 2012), including both negative words (Kanske & Kotz, 2007), and positive words (Herbert et al., 2008, Kissler et al., 2009). Beyond language, frontal LPC/LP has been associated with Theory of Mind (ToM). An LP effect (300-500 ms) was present for mental state judgements in the Mind-In-The-Eyes test, but not for gender judgements (Sabbagh et al., 2004). Likewise, an LP effect (500 ms onwards) was found for reasoning about the mental state of a cartoon character during a false-belief task, compared to reasoning about reality (Liu et al. 2004). An LP (600-840 ms) was found for reasoning about beliefs (Sabbagh & Taylor, 2000).

Less clear is the presence of N400, a negative-going potential peaking at ~400 ms. While traditionally associated with semantic processing (Kutas & Federmeier, 2011), Caillies et al. (2019) found that N400 is modulated by the emotional connotation of an ironic remark. They examined ironic praise (negative words pronounced ironically) and ironic criticism (positive words pronounced ironically), and found an increased N400 effect for ironic praise, but not for ironic criticism. However, frequency of use may have been a confounding factor, since ironic

praise occurs less frequently in naturalistic language. Similarly, Filik et al. (2014) found an N400-like effect for unfamiliar irony, but not familiar irony. Notably, both studies reported enhanced P600 effects for irony regardless of familiarity, associated with competing word meanings (Filik et al., 2014) and negative affect (Caillies et al., 2019). Lastly, some reported that irony elicited a larger P200, a positive going wave peaking at ~200 ms, than literal language (e.g., Regel et al. 2010). Given its early occurrence, it has been implicated in congruency detection (Regel et al. 2010), early attention allocation (Carretie et al., 2004), and alert responses to negative events (Kanske et al., 2011).

The present study examines whether context emotionality influences the perceived mental state of an ironic speaker and the time course of this process. Ironic and literal statements were embedded in high and low emotionally loaded stories. Each story began with a negative event that created either high or low emotional situations. The critical statement was an immediate ironic or literal response to the event uttered by the character suffering the negative consequences. In Experiment 1, we collected valence and arousal ratings about the mental state of the speaker. Based on the *tinge hypothesis*, speakers using irony would be in a less negative mental state, regardless of context negativity. Based on CAT, an ironic speaker would be in a less negative mental state in the low emotional condition, but in a more negative mental state in the high emotional condition. In Experiment 2, based on the *tinge hypothesis*, irony would mute negativity regardless of context, indexed by an enhanced P600/LPC effect in both contexts, reflecting similar semantic integration effort, ToM processing, or emotional processing. Based on CAT, irony would appear more negative in high emotional contexts, resulting an enhanced LPC effect compared to all others. In low emotional context, irony would assimilate negativity and be no different from literal language.

Experiment 1

Method

Participants. This and all succeeding experiments were approved by the local Institutional Review Board (University of Arizona).Eighty-three native-speaking participants gave informed consent and participated for course credit. Eighteen were excluded based on attention check questions. The remaining 65 had a mean age of 19.14 (*SD* = 1.4) years.

Materials and Design. We employed a 2 emotion (high emotion, low emotion) X 2 literality (literal, ironic) design. One hundred and twenty-one quadruplets (a total of 484 items) were selected from the initially constructed 133 quadruplets. Examples can be found in Table 1 and in supplementary Table A. Each item consisted of a context story and a target statement. The context briefly described a negative event, and the target statement was a comment by one of the protagonists. Minimal changes to the context rendered its emotional impact as either high or low (e.g., water being spilled on an <u>open</u> vs. <u>closed</u> computer). Likewise, one word in the target statement rendered its literality as either literal or ironic (e.g., how <u>clumsy/considerate</u> of you). Items were rotated via Latin Square, such that each list had one version of each item. Items within each list were randomized. The list number was counterbalanced with subject number.

We conducted two web-based pretests. Details are provided in supplementary materials. (1) *Emotionality of context* (high vs. low) was verified via ratings of valence and arousal (*N*=61) on the context stories without their target sentences. (2) *Literality of statement* (irony vs. literal) was verified via ratings of literality (*N*=80) on target sentences preceded by their context stories.

Table 1

Example stimuli used in experiment 1 and 2.

Emotionality (high emotion, low emotion)	Literality (literal, irony)

Context	Max is helping Jenny with her computer when he accidentally spills a glass of water over the open [*high emotion]* / closed [*low emotion*] computer. Jenny says:	**Target**	How clumsy [*literal*] / considerate [*irony*] of you!

Procedure. Participants were instructed what valence and arousal mean with examples and practice. The overall rating instruction was: "Please read each story carefully and only judge how the **protagonist** feels". Valence rating instruction was: "How does the protagonist feel?" (0 = *very negative*, 4 = *very positive*). Arousal rating instruction was: "How intense does the protagonist feel?" (0 = *very low arousal*, 4 = *very high arousal*). The ratings took 45-60 minutes.

Results

Table 2

Experiment 1 results.

Condition	**Arousal (4 = high)**	**Valence (4 = positive)**
	M (SE)	*M (SE)*
High emotion: irony	2.18 (0.04)	1.45 (0.04)
High emotion: literal	2.30 (0.04)	1.01 (0.05)
Low emotion: irony	1.90 (0.04)	1.63 (0.04)
Low emotion: literal	2.07 (0.05)	1.18 (0.05)

Table 2 summarizes the results. Data were entered in a Repeated Measures (RM)MANOVA with valence and arousal as dependent variables. A complete statistical report is in supplementary materials, Table B. Results revealed significant main effects of emotionality ($F(2,119) = 48.5$, $p < .001$) and literality ($F(2,119) = 35.9$, $p < .001$), with no interaction ($p = .74$). Follow-up RM-ANOVAs revealed significant main effects for literality in speaker valence ratings ($F(476) = 7.34$, $p = .007$) and speaker arousal ratings ($F(476) = 32.59$, $p < .001$), and a main effect of emotionality for speaker arousal ratings ($F(476) = 9.47$, $p = .002$). When a speaker used irony, regardless of how emotional the context was, they were perceived as being in

a more downregulated mental state (higher in valence and lower in arousal) compared to a literal statement. Similarly, a speaker in a high emotional context was rated as in a more aroused state compared to the low emotional context, regardless of using literal or ironic language.

Experiment 2

Experiment 1 results indicated irony dilutes negativity. However, differences in processing may arise earlier and may have already been resolved by the time rating occurs. Additionally, effects of valence and arousal affect processing at different timepoints based on ERP data (Citron, 2012) and independently based on reaction times (Kuperman et al., 2014). In Experiment 2, we used ERP to examine the initial time-course of irony processing in different emotional situations.

Method

Participants. Forty-four right-handed native English speakers gave informed consent and participated for course credit. All had normal or corrected to normal vision and reported no history of neurological abnormalities. None participated in the pretests or experiment 1. Four were excluded due to excessive noise in the EEG (> 40% trial loss in one condition). Five were excluded due to poor comprehension (< 80% accuracy). The remaining 35 participants (23 females) had a mean age of 18.81 years (*SD* = 0.94). To assess natural variations in participants' abilities to understand mental states, we administered the Faux-Pas test for Adults (Stone et al. 1998) and the Autism Quotient (AQ, Baron-Cohen et al. 2001). All participants received an average Faux-Pas recognition score of .95 (range: .9-1.0), suggesting intact ToM abilities. Effects of these measures will be discussed only when they modify the main findings.

Materials and Design. Same as Experiment 1.

Procedure and data acquisition. Brainwaves were recorded from 64 channel active

electrodes, while participants read context stories presented as a whole on the screen. Via button press, they advanced to the critical utterance, presented word-by-word. Further details are provided in supplementary materials.

Analyses. ERPs were timelocked to the critical word of each utterance (e.g. how *careful* of you!). Selection of time windows and electrodes were based on literature (see Intro) and visual inspection. We analyzed N100 (100-140 ms), P200 (220-280 ms), N400 (300-500 ms), and LPC (600-800 ms) in an anterior and a posterior location that include the majority of the electrodes (see supplementary materials, Figure C). Mean amplitudes were entered in a RM- ANOVA of 2 emotion (high, low) X 2 literality (literal, ironic) X 2 location (anterior, posterior). Greenhouse-Geisser correction was applied to correct for the violated assumption of sphericity. For the correlational analysis, the LPC effect was calculated by subtracting the mean amplitudes of the literal from the irony condition. The AQ was calculated according to Baron-Cohen et al. (2001).

Results

Comprehension accuracy was 95.5% (*SD* = 0.05%). The AQ had a mean of 17.94 (*SD* = 4.8) out of 50, range 10-29. This is comparable to Baron-Cohen et al. (2001) where healthy controls had a mean of 16.4 (*SD* = 6.3) and >32 is considered high. Thus, we did not exclude participants based on their AQ. ERPs had sufficient segments per condition (~27) and are displayed as grand averages in Figure 1.

N100 (100-140 ms). There was a main effect of literality ($F(1,34) = 9.640$, $p = .004$, $\eta^2 = .221$) and a literality x location interaction ($F(1,34) = 5.339$, $p = .026$, $\eta^2 = .137$). Literal elicited a more negative N100 than irony, in the anterior location ($F(1,34) = 12.178$, $p = .001$, $\eta^2 = .264$), but not in the posterior (*n.s.*). There was no literality x emotion interaction.

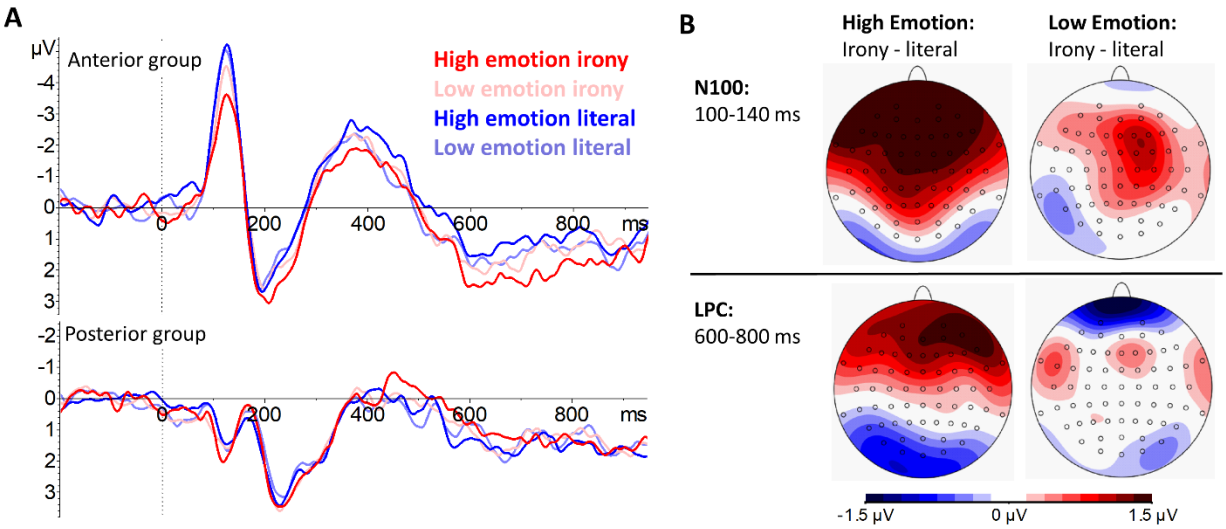
LPC (600-800 ms). There was a literality x location interaction ($F(1,34) = 5.712$, $p =$

.023, η^2 = .144) and an emotion x literality x location interaction (F(1,34) = 4.398, p = .043, η^2 = .115). Breaking down the interactions, irony elicited a more positive LPC than literal statements in the anterior channels in the high emotion condition (F(1,34) = 5.690, p = .023, η^2 = .143). Such LPC effect was not present in the low emotion condition.

Correlation between LPC effects and AQ scores. The subscale *attention switching* is correlated with the LPC effect in the high emotion condition, both anteriorly (r(33) = -.36, p = .031) and posteriorly (r(33) = .34, p = .044). The subscale *imagination* is correlated with the LPC effect, in the high emotion condition anteriorly (r(33) = -.39, p = .022), but not posteriorly (r(33) = .32, p = .064). These two subscales were not correlated (r = .08). A complete statistical report is in supplementary materials, Table B.

Figure 1

Experiment 2 results.



Note. A: Grand Averaged ERPs for the critical word in anterior group (top) and posterior group (bottom). B: Scalp distribution of irony minus literal for N100 effect (top) and LPC effect (bottom) in high emotion (left) and low emotion (right) condition.

General Discussion

We investigated the effects of context emotionality on the perception of irony using ratings and ERPs. In particular, we asked whether irony intensifies or mildens negative feelings depending on the emotionality of the context. Experiment 1 used behavioral ratings to examine readers' perception of an ironic language user's mental state in high and low emotional contexts. It was found that people who use irony are associated with a less negative and less aroused mental state, regardless of context emotionality. Experiment 2 used ERPs to examine the time course of irony perception. Literal statements elicited a more negative frontal N100 (100-140 ms) than irony. In addition, irony elicited a more positive frontal LPC (600-800 ms) than literal in high emotion, but not in low emotion conditions.

Based on ratings, we found that irony makes a speaker appear to be less negative and less aroused, consistent with the *tinge hypothesis*. Our findings expand the scope of the hypothesis by addressing the mental state of the ironic speaker rather than the emotional impact of the ironic remark on the victim. In addition to valence, we also assessed arousal and provided a finer grained resolution for the emotion-irony interface. Our findings differ from Ivanko and Pexman (2003), who found that irony was more mocking in strongly negative situations than in weak ones. This discrepancy may be explained by our focus on the speaker's mental state, wherein a less negative mental state does not exclude a more mocking intention.

In the ERPs, literal language elicited a larger N100 than irony in both emotional contexts. This N100 effect likely reflects an affective difference between irony and literal language. Hofmann et al. (2009) found that high-arousal negative words elicited a larger N100 (80-120 ms) than low-arousal negative or neutral words. Arousal has been linked to an automatic "fight or flight" response and may require a quick re-allocation of resources to ensure survival. León et al.

(2010) showed that emotionally inconsistent sentences elicited a more negative N100 (70-150 ms), associated with emotional consistency. Thus, our N100 effect may reflect how highly arousing and negative (potentially threatening) a literal reply is to the readers, as compared to an ironic reply. This difference in arousal is also evident in the Experiment 1 ratings and is partially inconsistent with Ivanco and Pexman (2003), who found that in strongly negative contexts, literal statements were rated more expected than ironic ones. However, given its rapid timing, the N100 likely reflects a more automatic response to high arousal in literal language and not expectancy within the context more broadly. Thus, our results at N100 are consistent with the *tinge hypothesis* that irony mildens negativity, while literal language "stings" more than irony.

Our second ERP finding is that irony elicited a more positive LPC than literal language in high emotion, but not low emotion contexts. We consider three accounts: (1) Semantic integration, (2) mental state processing, and (3) emotion. Most irony studies that reported a P600 associated it with cognitive effort for re-analysis / integration (cf. Introduction). Our findings can further support this interpretation. High arousing emotion context captures the readers' attention and occupies their neural resources. By the time the ironic remark comes on, the cognitive system does not have sufficient resources readily available and needs a boost for integrating the (more effortful) ironic remark with the context, indexed by the enhanced LPC. In contrast, in low emotion contexts, readers' neural reserve is not maxed out, thus they can integrate irony with the context accordingly. One issue that is not supportive of this interpretation is that irony related P600 is typically of centro-parietal distribution (e.g. Regel et al., 2010), whereas our effect is frontally distributed.

A second interpretation is that the LPC reflects the processing of the mental state of others. The AQ subscale *imagination* correlated with the LPC. The worse a participant's

imagination ability was, the less their LPC effect differentiated irony and literal remarks. The subscale *imagination* consists of questions such as "When I am reading a story, I find it difficult to work out the character's intentions" and "I find it difficult to imagine what it would be like to be someone else". These clearly require perspective taking. Specifically, in high emotional situations, processing the mental state of an ironic speaker requires more cognitive resources, because it is in stronger contrast to the contextual emotion, as shown by our rating results. This is not the case for literal utterances, where the mental state of the speaker is consistent with the context. Additional processing is not needed in the low emotion contexts, because the mental state of the speaker may seem less relevant or show a less stark contrast. Consistent with this interpretation, Liu et al. (2004) and Sabbagh & Taylor (2000) found a similar, frontally distributed LP effect when participants reasoned about beliefs and mental states. Thus, our frontal LPC effect may reflect more ToM in high emotional situations.

The third interpretation that LPC reflects emotion processing is directly relevant in answering whether irony mutes/enhances negativity (*tinge/contrast and simulation theories*). However, this interpretation inherits the effect directionality issue in emotion research. Namely, if the increased LPC for irony relative to literal language reflects the processing of positive emotion conveyed by text (e.g., Kissler et al., 2009), then the *tinge hypothesis* holds in high emotion contexts. Equally, if the LPC reflects humor, as proposed by Katz et al. (2004), our data would support the *tinge hypothesis,* as it would index more humor being conveyed in the high emotion condition. However, if the increased LPC reflects the processing of negative emotion (Kanske & Kotz, 2007), then our data support the *contrast and assimilation theory*, in that in highly negativity situations only, using irony requires continued processing. In the context of a larger N100 for literal language compared to irony, and an enhanced LPC for irony in high

emotion situations only, we believe that our LPC reflects the processing of negative emotion. Notably, the result of this continued processing is a less negative mental state, as evidenced by Experiment 1 ratings, suggesting that this processing reflects emotion regulation.

Across ERP time windows, one could alternatively argue that high emotion irony displays a sustained positive shift. However, differential patterns of results across all time-windows reduce the likelihood of a single neuro-cognitive process underlying all time windows. Taken together, the N100 illustrates that literality matters, while the enhanced LPC effect shows that emotionality matters. Thus, context emotionality poses differential processing constraints on irony, especially in high emotion situations. In other words, the high emotion context acts as a magnifying glass for the irony-emotion-interface.

Our behavioral and ERP findings paint the following picture: In a negative situation (e.g., water spilled on a laptop), a literal response is highly negative and unexpected, as reflected by the N100. When the situation is highly arousing (e.g., the laptop was open instead of closed), the difference between literal and ironic statements becomes more significant, as reflected by the LPC. Such significance is possibly due to inferring the speaker's mental state, or/and continued emotional evaluation of the irony user or the situation at large. While our behavioral data cannot discern these two interpretations, they indicate that downstream of processing, irony indeed mitigates negativity, because the readers considered the victims to be in a less negative and less aroused mental state.

In conclusion, both *tinge* and CAT are supported at different time points. While the *tinge hypothesis* is supported downstream of processing in ratings, CAT seems to hold during irony processing under different contextual constraints. This study contributes by further clarifying what "conveying negative emotion" means (cf. Roberts & Kreuz, 1994) and by presenting a

temporally more fine-grained picture of emotion processing during irony comprehension.

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders, 31*(1), 5-17.

Boylan, J., & Katz, A. N. (2013). Ironic expression can simultaneously enhance and dilute perception of criticism. *Discourse Processes, 50*(3), 187-209.

Carretié, L., Hinojosa, J. A., Martín-Loeches, M., Mercado, F., & Tapia, M. (2004). Automatic attention to emotional stimuli: neural correlates. *Human brain mapping, 22*(4), 290-299.

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.

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological & hemodynamic neuroimaging studies. *Brain and language, 122* (3), 211-226.

Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics, 34*(2), 111-142.

Dews, S., & Winner, E. (1995). Muting the meaning a social function of irony. *Metaphor and Symbol, 10*(1), 3-19.

Filik, R., Brightman, E., Gathercole, C., & Leuthold, H. (2017). The emotional impact of verbal irony: Eye-tracking evidence for a two-stage process. *Journal of Memory and Language, 93*, 193-202.

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.

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology, 45*(3), 487-498.

Hofmann, M. J., Kuchinke, L., Tamm, S., Võ, M. L., & Jacobs, A. M. (2009). Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 389-397.

Ivanko, S. L., & Pexman, P. M. (2003). Context incongruity and irony processing. *Discourse Processes*, *35*(3), 241-279.

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, *49*(11), 3121-3129.

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain research*, *1148*, 138-148.

Katz, A. N., Blasko, D. G., & Kazmerski, V. A. (2004). Saying what you don't mean: Social influences on sarcastic language processing. *Current Directions in Psychological Science*, *13*(5), 186-189.

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.

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.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax.

Brain research, 1146, 23-49.

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065-1081.

León, I., Díaz, J. M., de Vega, M., & Hernández, J. A. (2010). Discourse-based emotional consistency modulates early and middle components of event-related potentials. *Emotion, 10*(6), 863-873.

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: an ERP study of theory of mind. *NeuroReport*, *15*(6), 991-995.

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain research, 1311*, 121-135.

Roberts, R. M., & Kreuz, R. J. (1994). Why do people use figurative language?. *Psychological science*, *5*(3), 159-163.

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: an event-related potential study. *Journal of cognitive neuroscience*, *16*(3), 415-426.

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: an event-related potential study. *Psychological Science*, *11*(1), 46-50.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of cognitive neuroscience, 10*(5), 640-656.

Supplementary Table A. Example stimuli used in Experiment 1 and 2.

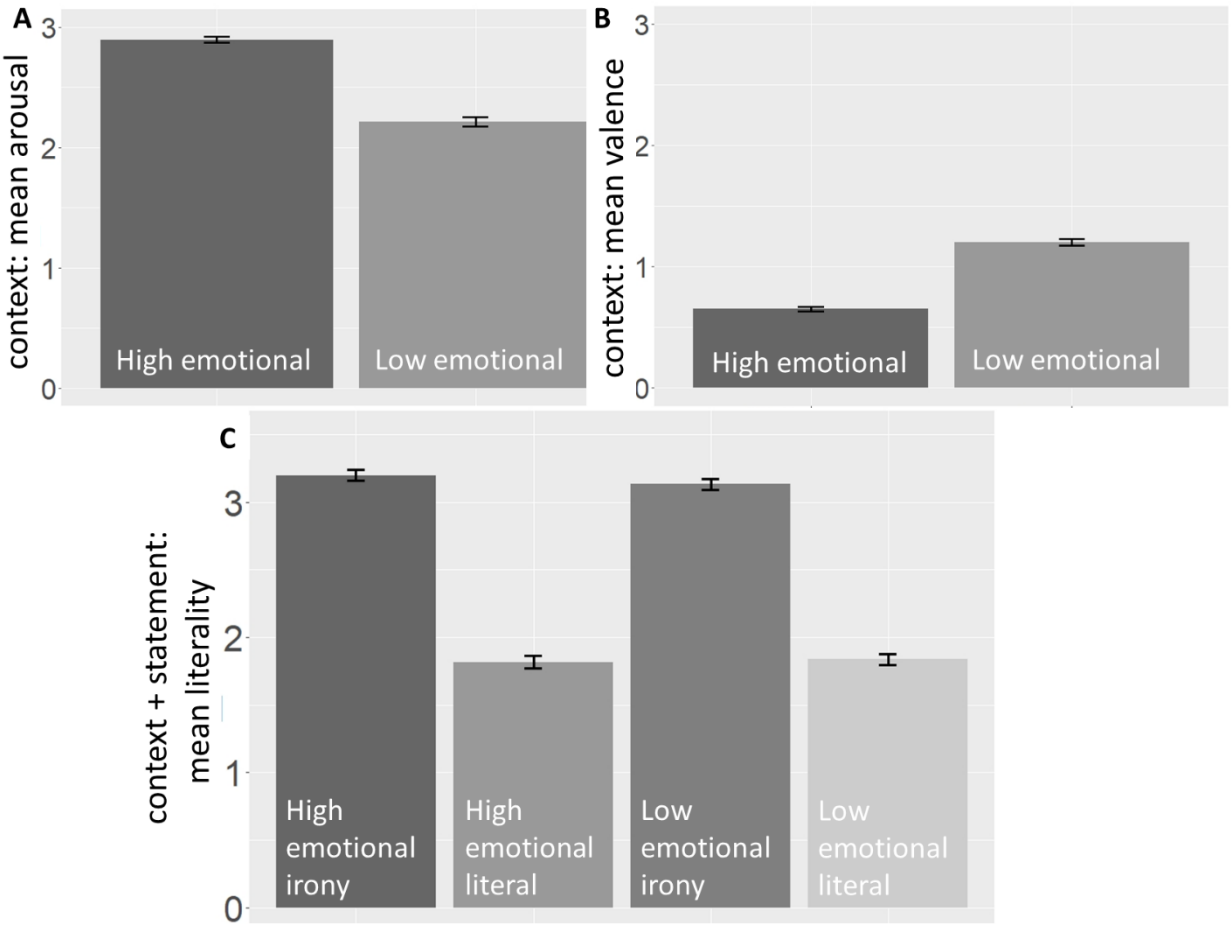
Emotionality (high emotion, low emotion)		Literality (literal, irony)	
Example 1			
High Emotion	Max is helping Jenny with her computer when he accidentally spills a glass of water over the open computer. Jenny says:	**Literal**	How clumsy of you!
Low Emotion	Max is helping Jenny with her computer when he accidentally spills a glass of water over the closed computer. Jenny says:	**Irony**	How considerate of you!
Example 2			
High Emotion	When getting into the car, Ricarda accidentally sits on her boyfriend's new expensive sunglasses. Ricarda's boyfriend says:	**Literal**	I knew you were not as considerate as you said!
Low Emotion	When getting into the car, Ricarda accidentally sits on her boyfriend's new expensive scarf. Ricarda's boyfriend says:	**Irony**	I knew you were not as clumsy as you said!
Example 3			
High Emotion	Audrey completely forgot all of her lines during the premier of her play. Audrey says:	**Literal**	I guess I wasn't really advancing today's performance.
Low Emotion	Audrey completely forgot all of her lines during the first practice of her play. Audrey says:	**Irony**	I guess I wasn't really spoiling today's performance.
Example 4			
High Emotion	When Dominic is on a date, his ex-girlfriend walks in and starts a scene. Dominic says:	**Literal**	Thanks for spoiling my day!
Low Emotion	When Dominic is at home, his ex-girlfriend walks in and starts a scene. Dominic says:	**Irony**	Thanks for advancing my day!
Example 5			
High Emotion	Frankie has a new job in a restaurant. When he tries to impress his boss by carrying multiple plates with food to a table, he drops all of them. Frankie says:	**Literal**	I wish I was more professional than this!
Low Emotion	Frankie has a new job in a restaurant. When he tries to impress his boss by carrying multiple empty plates to the kitchen, he drops one of them. Frankie says:	**Irony**	I wish I was more amateurish than this!
Example 6			
High Emotion	Amy and Michelle are on a hiking trip, when Amy stumbles and falls down onto a rocky slope. Amy says:	**Literal**	I really am an amateurish hiker!

Low Emotion	Amy and Michelle are on a hiking trip, when Amy stumbles and falls down onto the soft grass. Amy says:	**Irony**	I really am a professional hiker!

Additional pretest information.

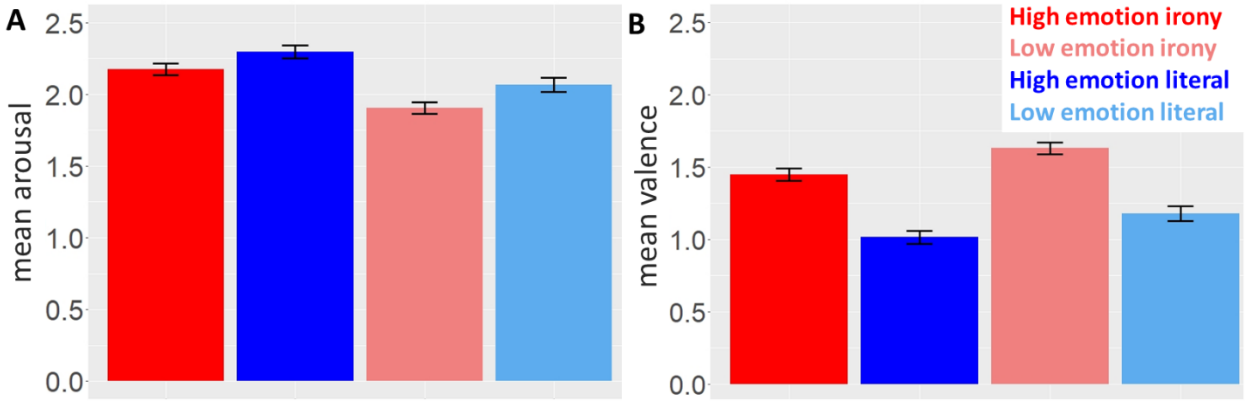
Pre-test 1: *Emotionality of context* (high vs. low) was verified via a web-based pre-test (Subject *N*=61) on the context stories without their target sentences. Sixty-one native speakers of English (*M* age = 19.2, *SD* = 1.3) who did not participate in the main experiment participated for course credit. Items were split in two lists and pseudo-randomized, such that no more than four stories from one condition were presented consecutively. Participants rated how the protagonist felt in each context story, using 5-point-Likert-scales for valence and for arousal (Valence: 0 = negative, 4 = positive; Arousal: 0 = low arousal, 4 = high arousal). To verify participation, three types of attention checks were inserted: (1) "I am still paying attention" checkboxes, (2) enforced ratings ("Please choose option 3"), and (3) definition checks (T/F: "Valence describes how positive or negative something is"). Eighteen participants were excluded due to failure to answer attention check questions. For the high emotion stories, the mean valence was 0.65 (*SD* = 0.3) and the mean arousal rating was 2.90 (*SD* = 0.41). For the low emotion stories, the mean valence was 1.20 (*SD* = 0.45) and the mean arousal rating was 2.21 (*SD* = 0.58). Results are summarized in Figure A. A two-tailed t-test revealed a significant lower valence ($t(2, 264) = -10.93$, $p < .001$) and higher arousal ($t(2,264) = 10.79$, $p < .001$) for high emotion stories than for low emotion stories, verifying our intended manipulation. Results are illustrated in supplementary Figure A, subplots A and B.

Pre-test 2: Literality of statement (irony vs. literal) was verified via a web-based pre-test (Subject *N*=80) on the context stories followed by their target sentences. The results are in Figure A. Eighty native speakers of English (*M* age = 20.2, *SD* = 5.1) who did not participate in the main experiment participated for course credit. Items were split across four lists and pseudo-randomized, such that not more than four ironic stories were presented consecutively. Participants read the context story and target statement and judged how literal the statement was on a 4-point-Likert scale (1 = very literal, 4 = very non-literal). We used the label "very non-literal" instead of "very ironic", to avoid raters being biased towards choosing irony. We, however, used ironic statements as examples to illustrate non-literalness in the instructions. Nineteen participants were excluded due to attention check questions (e.g. "Please rate this statement as a one."). A total of 12 statements were excluded because their literality rating was higher in the literal than the ironic condition. The results of the remaining 121 items are summarized in Figure 1. For the high emotion condition, literal statements received a mean rating of 1.82 (*SD* = 0.52), and ironic statements, 3.2 (*SD* = 0.46). In the low emotion condition, literal statements received a mean rating of 1.84 (*SD* = 0.47) and ironic statements, 3.12 (*SD* = 0.47). A two-tailed t-test revealed a significant difference between ironic and literal statements ($t(2,482) = 28.65$, $p < .001$), verifying our manipulations. Results are illustrated in supplementary Figure A, subplot C.



Supplementary Figure A. Norming results. A. Mean arousal ratings for context stories (left/dark bar: high emotion; right/light bar: low emotion). B. Mean valence ratings for context. C. literality ratings of the target statement. Light bars indicate literal statements, dark bars indicate ironic statements. Left two bars show high emotion context. The error bars indicate Standard Errors (*SE*).

Attention check questions. Three types of attention checks verified participation: (1) "I am still paying attention" (checkboxes), (2) enforced ratings ("Please choose option 3"), and (3) definition checks (T/F: "Valence describes how positive or negative something is").



Supplementary Figure B. Experiment 1 results. Mean speaker arousal (left) and valence (right) ratings. Blue bars indicate literal statements, red bars indicate ironic statements. Darker colors indicate high emotion context (right two bars).

Supplementary Table B. Experiment 1 full statistical results.

MANOVA (Arousal & Valence as dependent variables)		
	Emotion: $F(2,119) = 48.5$, $p < .001$, $\eta^2 = .449$, Wilk's $\lambda = .552$	
	Literality: $F(2,119) = 35.864$, $p < .001$, $\eta^2 = .376$, Wilk's $\lambda = .624$	
ANOVA		
	Arousal	Valence
	Emotion $F(1,476) = 9.470$, $p = .002$	Emotion: $F(1,476) = 2.881$, $p = .09$
	Literality $F(1,476) = 7.342$, $p = .007$	Literality: $F(1,476) = 32.587$, $p < .001$
Pairwise comparisons		
	Arousal	Valence
High Emotion	Literality: $F(1,238) = 4.163$, $p = .04$	Literality: $F(1,238) = 15.64$, $p<.001$
Low Emotion	Literality: $F(1,238) = 3.244$, $p = .07$	Literality: $F(1,238) = 16.95$, $p < .001$
Irony	Emotion: $F(1,238) = 5.025$, $p = .026$	Emotion: $F(1,238) = 2.092$, $p = .15$
Literal	Emotion: $F(1,238) = 4.508$, $p = .03$	Emotion: $F(1,238) = 0.985$, $p = .32$

Additional Experiment 2 EEG-methodology information.

Procedure. Brainwaves were recorded from 64 channel active electrodes, placed using the international 10-20 system. Participants were fitted the electrode cap and seated 80 cm away from an LCD screen in a sound attenuated booth. Stimuli were presented visually in white font (Courier New, size 20) against a black background via E-prime 3.0 software (Psychology Software Tools, Inc.). Each trial started with the context story, presented at once on the screen, self-paced. Then, a fixation cross (500 ms) appeared. The target statement was presented word-by-word for a length dependent duration (mean: 386 ms, range: 320-420 ms), with a blank screen

(500 ms) between words. After each trial, a "blink or continue" screen appeared, during which participants could rest before the next trial. Participants were instructed to read silently. After 20% of the trials, basic yes/no comprehension questions appeared to verify participation. Participants completed five practice trials to familiarize themselves with the procedure. Trials were randomized and spatial order of yes/no keys was counterbalanced. After recording, participants filled out the Autism Quotient questionnaire (Baron-Cohen et al., 2001) and completed the Faux-Pas for Adults test (Stone et al., 1998). The total session lasted between 1.5 - 2 hrs.

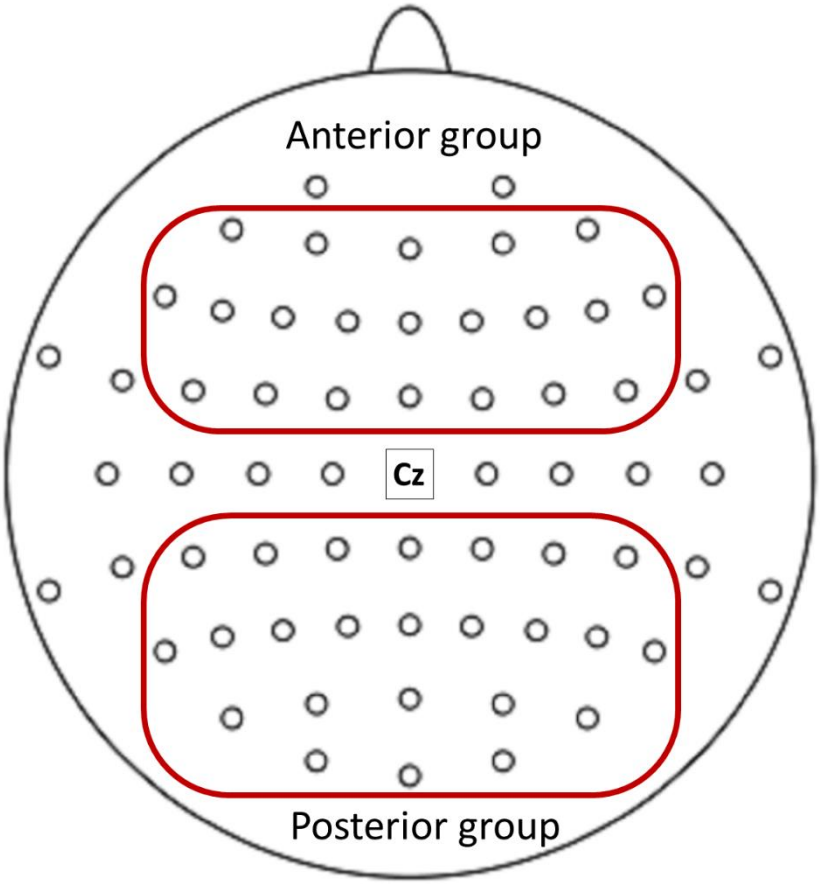
EEG data acquisition and pre-processing. EEG was recorded continuously from 64 active electrodes (Brain Vision, actiCHamp) at a sample rate of 500 Hz. Electrodes were placed using a 10-20 system. Impedances were kept below 10 kΩ. Data was analyzed using Brain Vision Analyzer 2.0. Data were band-pass filtered at 0.01 Hz and 30 Hz, order 4, with a notch filter at 60 Hz. Data were re-referenced to the average of the left and right mastoid. Ocular movements were corrected with an Independent Component Analysis algorithm (Infomax). Data were segmented into epochs 200 ms before the target word onset and 950 ms after. Segments with artifacts exceeding an amplitude difference of 100 µV were removed. All segments were normalized to a 200 ms pre-stimulus baseline, and averaged ERPs time-locked to the onset of the critical word in each condition were computed.

Segments kept. After artifact rejection, average remaining segments per condition were as follows: 27.2 high emotion literal, 27.3 low emotion literal, 26.9 high emotion irony and 26.7 for low emotion irony.

Determining the Regions of Interest (ROI) and Electrodes: Upon visual inspection of the effects, it was noted that the observed effects were visible mostly in anterior and posterior channels. No differences between the hemispheres were visually present, as visible in Figure 1 B. Thus, we decided to choose the regions displayed in supplementary Figure C based on visual inspection of the data. The anterior group spanned the electrodes AF3, AF7, F1, F3, F5, F7, FC1, FC3, FC5, AF4, AF8, F2, F4, F6, F8, FC2, FC4, and FC6, while the posterior group spanned electrodes CP1, CP3, CP5, P1, P3, P5, P7, PO3, PO7, O1, CP2, CP4, CP6, P2, P4, P6, P8, PO4, PO8, and O2. In previous literature, significant effects of laterality in irony processing have not been reported. For example, Regel et al. (2010) did not report any effects of laterality or anteriority. Filik et al. (2014) used similar electrodes than we did to form two regions of interest (ROIs): anterior (AFz, Fz, FCz, Cz) and posterior (CPz, Pz, POz, Oz) for their laterality analysis and found a significant interaction of experimental manipulation with anteriority, but not with laterality at P600/LPC. Equally, their factor hemisphere did not produce any main effects or interactions throughout their analyses. Lastly, Caillies et al. (2019) did not find any significant effects of hemisphere in their data. While some authors (Regel et al. 2010, Filik et al. 2014) have chosen to conduct omnibus ANOVAs using all recorded electrode sites, we have decided to conduct data-driven comparisons to avoid the artificial inflation of type I errors. Others, such as Caillies et al. (2019) have chosen a similar, data-driven approach for analysis.

Additionally, to confirm our visual inspection, we conducted a second analysis using a Quadrant approach to include factors of laterality, which did not produce any effects of laterality at either timewindow, except for P200. Here, a marginally significant effect of emotion * hemisphere * anteriority with p=.08, in addition to a main effect of literality with p=.079 (similar to the reported effect in the manuscript) was observed. Based on the marginality of the effect, the

lack of interaction with the whole-scalp effect of literality and the amount of comparisons conducted, we consider this effect based on chance or minor and decided that the use of two ROIs is appropriate for characterizing the effects within our data.



Supplementary Figure C: Electrodes used for analysis in anterior group (AF3, AF7, F1, F3, F5, F7, FC1, FC3, FC5, AF4, AF8, F2, F4, F6, F8, FC2, FC4, FC6) and posterior group (CP1, CP3, CP5, P1, P3, P5, P7, PO3, PO7, O1, CP2, CP4, CP6, P2, P4, P6, P8, PO4, PO8, O2).

Additional results for P200 and N400 timewindow.

P200 (220-280 ms). There was a marginally significant main effect of literality ($F(1,34) = 3.250$, $p = .08$).

N400 (300-500 ms). There was no significant result.

AQ scores: No gender difference.

Participants' Autism Quotient ranged from 10 to 29 (out of 50), with an average score of 17.94 ($SD = 4.8$). Females' mean score was 18.19 ($SD = 4.29$) with a range from 10-25, while males scored on average 17.69 ($SD = 5.25$), with a range from 10-29. A two-tailed t-test revealed no statistically significant difference between AQ-scores for males and females ($p = .78$). Thus, we disregard participant gender for AQ in all further analysis.

Supplementary Table C: Correlations between the Autism Quotient overall and subscales and the LPC effect for irony relative to literal statements. Values reflect correlation coefficients, with *p* values in parenthesis. Bold reflects significant results.

	High emotion		**Low emotion**	
	Anterior	Posterior	Anterior	Posterior
AQ (overall)	-.15	.06	-.23	-.15
subscales				
Communication	-.08	-.06	.23	.04
Attention to Detail	.13	-.16	-.07	-.29
Attention switching	***-.36 (*p* = .031)**	***.34 (*p* = .044)**	-.07	-.02
Imagination	***-.39 (*p* = .022)**	.32 (*p* = .064)	.17	.04
Social Skills	.12	.23	.30 (*p* = .076)	-.04