Psychology of Language and Communication 2024, Vol. 28, No. 1

sciendo

DOI: 10.58734/plc-2024-0014

Valeria A. Pfeifer[1], Matthias R. Mehl[1], and Vicky Tzuyin Lai[1]
[1] Psychology, University of Arizona, United States
[2] Cognitive Science Program, University of Arizona, United States

That was clever of you! Perspectives and verbal irony

What does ironic language indicate about an emotional state? When we experience negative events, we may produce an ironic comment about our experience to cope. We tested whether using irony indexes less negative emotional states of speakers and listeners compared to literal statements. In addition, we tested whether comments directed at the self ("How clumsy/careful of me!") lead to a different emotional state compared to statements directed at someone else ("How clumsy/careful of you!"). We found that listeners were rated as feeling more negative and more aroused when a statement was directed at them, whereas for speakers, direction did not matter. Further, self-irony led to lower ratings of negativity in speakers. Our results show that whether or not someone is an addressee of a statement matters for their perceived emotional state, and that self-irony may index downregulated emotional states.

Key words: speaker, listener, verbal irony, emotion, emotion regulation, self-talk.

Address for correspondence: Valeria A Pfeifer
Psychology, University of Arizona, Tucson, 1503 E University Blvd Tucson, AZ 85721 USA.
E-mail: vpfeifer@email.arizona.edu

Verbal irony is a form of nonliteral language that can express the opposite of a speaker's attitude towards a matter, with the goal that the recipient of the expression uncovers this discrepancy, for example, exclaiming "What a giant portion!" after you have received a very small serving of your favorite dish. While a wide range of definitions of verbal irony exists, including speakers mentioning ironic words to echo prior attitudes (Sperber & Wilson, 1981), or the speaker engaging in pretense when using irony (Clark & Gerrig, 1984), scholars agree that a detectable opposition between expression and intention of language is a hallmark of verbal irony.

Verbal irony is often divided into ironic criticisms and ironic compliments, where ironic criticisms are when a literal positive statement expresses a negative meaning, for example, when saying "great job" to your co-worker that just spilled their coffee over your paperwork (also called blame-by-praise). On the other hand, ironic compliments are negative expressions that convey a positive meaning, for example, complimenting your partner's gorgeous new outfit by saying: "looks awful!" Using positive words to comment on a negative situation is the more frequent and easier to grasp form of verbal irony (c.f. Harris & Pexman, 2003) and it is the main focus of the current study.

Verbal Irony and Emotion

Verbal irony has been found to be associated with negative emotion. Roberts and Kreuz (1994) asked participants why they use a particular figure of speech, and 94% of respondents listed reasons classified as "expressing negative emotions" to use verbal irony. Two questions can be asked regarding this phenomenon: What does "negative emotions" mean, and does irony simply serve to express negative emotions, or does it (also) help to manage them? The first question was not our focus, but is important when considering prior work. In verbal irony, especially in studies focusing on ironic criticism, emotion, affect, attitudes, and evaluations are commonly used terms, but they are not often directly classified as "negative." Such an evaluation is implied and often left to the reader. However, a communicative context that is unpleasant or undesirable is charged with some form of negative affect, and so is a person's adversary response to a comment, such as feeling insulted or criticized, meaning a "negative" emotion, evaluation, or attitude is expressed by or elicited in one or more of the interlocutors of the communication. These terminological difference might stem from methodological differences, where emotions are either assessed categorically on specific, discrete scales (e.g., one scale each for embarrassment, criticism, ridicule, and amusement), or continuously on scales capturing more general dimensions of emotional experience (e.g., valence and arousal). Alternatively, authors might not want to generalize to emotional experience overall, but rather retain the precision of their measurement. In either case, one can discuss irony in terms of specific aspects of emotional experience, (e.g., embarrassment) or using the more general

underlying coloration of such experience (e.g., negative feelings).

The current study focused on the second question, which was what irony does to negative emotion. One hypothesis is the tinge hypothesis (Dews & Winner, 1995), which suggests that the literal meaning of an utterance tinges the perception of the ironic meaning. For example, when someone says "very elegant" to someone who stumbles, the literal positive meaning of "elegant" infuses into the negativity of the intended, critical meaning (not elegant), thereby making it appear less negative. In their work, Dews and Winner (1995) found that ironic criticism was rated as less critical (meaning less negative) than literal criticism, while ironic compliments were rated as less praising than literal compliments. Dews et al. (1995) further found that ironic criticism conveys different social information than literal criticism. Specifically, it was rated as funnier and less aggressive, suggesting that irony can act as face-saving for both speaker and addressee and is more humorous than literal language. In addition, irony's impact on the speaker-addressee relationship was rated as less negative than that of literal criticism, although the authors noted that the impact of irony on social relationships was small. Last, evidence in favor of the tinge hypothesis in the context of emotion comes from Pfeifer and Lai (2021), where participants rated the emotional state of speakers of literal and ironic statements on valence and arousal in either highly negative or mildly negative context situations. They found that irony speakers were rated as being in more positive and less aroused emotional states compared to literal speakers, regardless of context emotion (high or low).

However, not all forms of irony in the context of negative events result in reduced negative emotion. Colston (1997) suggested that ironic criticism is more condemning than literal criticism, as well as being meaner. The contrast and assimilation theory (Colston, 2002) claims that the effects of irony depend on the context, such that in strongly negative situations, irony appears more negative due to establishing contrast, while in mildly negative situations, irony appears less negative due to assimilation. The theory is supported by Ivanko and Pexman (2003), who embedded irony in strongly negative, mildly negative, and neutral contexts (e.g., being a few hours late vs. a few minutes vs. not late). They found that while in mildly negative contexts, irony was rated as more polite (meaning more positive), and that it was rated as more mocking (meaning more negative) in strongly negative contexts than literal statements. These results were also reflected in reading times, where irony was read slower in strongly negative contexts, but faster or equally fast in mildly negative contexts compared to literal ones, suggesting more integration difficulties in the strongly negative contexts.

Additionally, the strength of the contrast between the context and the ironic comment might matter for irony's emotional impact. Colston and O'Brien (2000) compared irony with weak contrasts (e.g., commenting on someone's "agreeable mood" when they are in a bad mood) and strong contrasts (e.g., using "magnificent mood") by changing how hyperbolic the modifiers used were, and

found that effects of irony are enhanced when irony uses stronger contrasts to the situation. In particular, while all irony expressed more condemnation and more humor and speaker protection compared to literal statements, irony with strong contrasts did so to a greater degree.

Interlocutor Perspectives During Irony Processing

To understand irony's impact on emotional experiences, we must consider the different interlocutor roles, such as speaker, addressee, or uninvolved bystander, during an ironic exchange. A few studies have considered different interlocutor roles and perspectives in irony (Bowes & Katz, 2011; Boylan & Katz, 2013; Pexman & Olineck, 2002; Toplak & Katz, 2000). For example, Toplak and Katz (2000) contrasted literal and ironic criticism in eight scenarios and asked participants to take the speaker's, the addressee's, the audience's, or no specific perspective when rating how the speaker intended their statement on 18 different scales, including humor, mocking, aggression, and impact on social relationship. Results indicated that from a speaker's perspective, irony indexes verbal aggression, whereas from an addressee's/listener's perspective, it seems to be a more severe form of criticism than literal criticism. Thus, results of Toplak and Katz (2000) support the idea that irony, specifically when used to criticize, increases negativity in being more aggressive for speakers and more critical for listeners/addressees. One caveat is that this study used few scenarios and 18 different rating scales, together with several different perspectives, which makes the results hard to interpret.

Pexman and Olineck (2002) were the first to propose that enhanced criticism (negative emotional experience) is achieved when raters consider the speaker's intent, but increased politeness (and thus, decreased negativity) is achieved when raters consider the social impression of the statement (i.e., the listener's perspective). They tested this using metaphoric comments that contained positive or negative modifiers and were used in either positive or negative contexts. For ironic criticism (expressed through metaphors using positive modifiers in negative contexts), they found that comments were rated as more positive and polite than literal criticism, as well as more mocking and sarcastic. They argued that positivity and politeness amplified the social impression of the statement, while mocking and sarcasm amplified speaker intent. However, they did not explicitly manipulate whether a speaker, a listener, or a bystander's perspective was considered, and they used both metaphors and irony. This study nevertheless provided testable hypotheses for others, such as Bowes and Katz (2011), who examined sarcasm in the context of aggression. Instead of stories describing actions, Bowes and Katz (2011) used written dialogues, and their participants rated humor and politeness, as well as a number of other variables, including appropriateness, aggression, and social standing, from both the speaker's and the victim's perspectives. They found that when considering the speaker's perspective,

sarcasm was perceived as more humorous and less aggressive compared to when considering the addressee's perspective, illustrating that differences in the impact of verbal irony might be due to the perspective of speaker versus addressee. Note that Bowes and Katz (2011) considered only very aggressive, critical forms of irony, which is only one variety of the irony bouquet, and their participants rated both perspectives simultaneously for the same scenario.

Lastly, Boylan and Katz (2013) took on testing Pexman and Olineck (2002)'s hypothesis by having the participant imagine they are the speaker of literal and ironic criticism that either contained a hyperbolic modifier or not (see Colston and O'Brien, 2000). Half of the participants rated speaker intent, and the other half rated speaker impression, which was considered as the listener's perspective by Boylan and Katz (2013), in line with Pexman and Olineck (2002). Results suggested that irony can both, enhance and reduce negativity simultaneously by enhancing features like humor, politeness, and positivity, while also increasing mocking and sarcasm. However, ratings did not differ by perspective, as predicted by Pexman and Olineck (2002). Results from the hyperbolic modifier manipulation showed that the presence of hyperbolic modifiers increased mocking and decreased politeness, but did not affect sarcasm ratings, which only partially replicates Colston and O'Brien (2000).

Humor in Irony

A third facet to the emotional aspects of verbal irony is humor. While some argue that irony is always implying something negative (Clark & Gerrig, 1984) and therefore, is not suited for communication, others argue that irony is primarily a form of humor (Anolli et al., 2002; Ortega, 2012). In our opinion, humor could either act as a mechanism to reduce or replace the negative emotion in interlocutors (such as replacing criticism with amusement) or occur simultaneously with the negative emotional consequences of verbal irony, a case of mixed emotional experience (cf. Pfeifer & Pexman, 2023). As described above, in Boylan and Katz (2013), participants rated ironic criticism as more humorous, polite, and positive, yet also more mocking and sarcastic than literal criticism, suggesting that irony simultaneously increases and dilutes negative emotion. However, the exact relationship between irony and humor remains unclear and requires further investigation (Gibbs et al., 2014). This is in part, because a number of prior studies (e.g., Bowes & Katz, 2011; Dews et al., 1995; Pexman & Olineck, 2002; Toplak & Katz, 2000 amongst others) used humor or amusement as an additional dependent variable to be rated, rather than as a potential explanation for the effects of irony on other aspects of emotion.

Irony Directed at Others Versus at the Self (Self-Irony)

A less considered scenario in psycholinguistics is irony directed at the self, or self-irony. In this case, the speaker still relies on the listener(s) to identify that they

did not mean their statement literally, but the risk of committing a communicative faux pas is smaller, because even if misunderstood, the consequences of the ironic remark are with the speaker, not the listener. In some contexts, the speaker might even engage in self-talk with no interlocutors present, eliminating the problem of possibly being misunderstood. Sometimes, self-irony is confounded with self-deprecating humor, which describes laughing at, or ridiculing the self. While this could be considered a subset of self-irony, such view only spans public, humorous functions of irony, but not possible self-serving functions of it. Another aspect to consider is self-criticism, which is what intentionally or unintentionally happens when ironic criticism is targeted at the self. Bajerski (2016) also used the term "auto-irony" to refer to irony directed at the self. Using an irony production task where participants were asked to produce a possible utterance of the main character following written context scenarios, Bajerski tested if emotional intelligence and irony production were related. He found that emotional intelligence was negatively correlated to self-irony and ironic compliments/praise, but not ironic blame/criticism. In other words, the higher the emotional intelligence measured with EIQ, the less a participant produced self-irony and ironic praise. Bajerski (2016) interpreted his findings as demonstrating that self-irony is not an effective way to reduce negative tension

The Current Study

The current study investigated the influence of perspective on the perceived emotional state of speakers and listeners. Specifically, we contrasted ironic and literal statements made about the self or another person and collected valence and arousal ratings on both the speaker's and the listener's perceived emotional state. Following Pfeifer and Lai (2021) we decided to use both valence and arousal as the dependent measures, as these are considered orthogonal features of the same emotional experience (cf. Barrett, 1998), which allowed us to pick up on subtle differences in emotional experience. Our study had three aims: determining (a) whether ironic and literal statements differentially impact a perceived emotional state, (b) whether a speaker and a listener are perceived to feel differently about a situation following ironic and literal statements, and (c) whether these effects are modulated by the statement targeting the self or another person. Our predictions were as follows.

(A) We predict that irony would impact emotions of those involved in the communicative exchange. Specifically, using irony should reduce negativity in emotional state ratings compared to literal language, based on results from Pfeifer and Lai (2021), who showed that speakers were rated to be in a less negative and less aroused emotional state, and Boylan and Katz (2013), who showed that there was no difference between different interlocutor perspectives. Alternatively, an ironicity × role (speaker or listener) interaction is plausible, too, wherein irony still impacts emotion, but differentially impacts speakers and listeners.

(B) We predicted that speakers' and listeners' perceived emotional state

would differ depending on the kind of comment made (directed at the self or another person). Specifically, statements directed at a listener (other-directed statement) would have stronger effects on perceived listener's emotion than perceived speaker's emotion. By contrast, self-directed statements would impact the (less involved) listener less, compared to the speaker.

Method

Participants

All research activities were approved by the local institutional review board. All participants only completed one part of the study (the experiment or one of the three norming studies). Two hundred and four native speakers of English (M_{age} = 19.20 years, *SD* = 4.07, range = 18-61; 48 males, 155 females, 1 undisclosed) gave consent to participate in main experiment of the study for course credit. None of them participated in any of the norming studies (described below). Twenty-four were excluded from the final analyses due to failure to respond correctly to two or more attention check questions (see the Procedure section for details). An additional 15 were excluded because their answers were outside of two *SD*s above or below the group means in more than 10% of the questions, meaning their responses were very different from the group average on a large number of items.[1] There also were two individuals that were significantly older than the average participant age (42 and 61 years), however, their answers did not exceed the >10% +/-2*SD* cutoff, so we concluded they did not differ from average-aged responses, and the data remained in the analyses. Thus, there were 165 individuals remaining in the final analyses of the main experiment reported here.

Materials and Design

The study used a within-participants design of 2 × 2 × 2 [ironicity (literal, irony) × addressee (self, other) × role (speaker, listener)]. One-hundred-and-sixty stimulus items were initially constructed. Each item consisted of a short vignette followed by a critical remark/statement. The vignette introduced two characters and described a negative event happening to them. Then, one of the characters made a remark to the other, which was either a statement about the self ("How skillful of me!"), or (coming from the other character) a statement about the other person ("How skillful of you!"). The statements were rendered to be either literal or ironic by replacing one word of the statement (skillful/inept). Words that were literal in one vignette were used as ironic with another vignette, to control for

1 We considered these participants attentive to the attention checks, but inattentive to the actual items. For completeness, we added an analysis including these participants' responses on the OSF webpage as an appendix. While numerically different, no differences in which results reached statistical significance emerged.

Table 1. *Examples of Vignette-Statement Combinations.*

Context vignette	Since Robyn lost her job, she has been spending her days at home watching Netflix and browsing social media. Ella, her friend, discusses Robyn's recent lack of interest with her.
Self:	Robyn says: How lazy (literal) / ambitious (irony) I am!
	Speaker: Robyn Listener: Ella
Other:	Ella says: How lazy (literal) / ambitious (irony) you are!
	Speaker: Ella Listener: Robyn

word-based effects, including word valence. As such, there were a total of 640 vignette-statement combinations, and an example is listed in Table 1. The full set is available online in the study's OSF webpage at: https://osf.io/u7whr/.

Norming

We conducted three types of norming: (a) context emotion, (b) ironicity, and (c) humor, all administered online via Qualtrics (Qualtrics, Provo, UT). All participants in the norming were native speakers of English and did not participate in any other part of the study. They participated for course credit. The original wording of the instructions can be found in the OSF.

First, the context vignettes without the critical statements were normed for "context emotion" which means valence and arousal, to understand the emotional charge of the context in which statements were made, as context emotion occasionally impacts irony perception (Ivanko & Pexman, 2003). This norming only included the context (which was constant across all 4 conditions of each item), not the critical utterance at the end of each vignette, so only one list was needed. Participants ($N = 16$, $M_{age} = 19.25$ years) were given definitions of valence and arousal with examples, and read and rated all 160 vignettes presented without the critical statements. Their task was "to rate the emotional aspects of the situation using valence and arousal" using two 7-point Likert type scales, one for valence (0 = *very negative*, 6 = *very positive*) and one for arousal (0 = *very low arousal*, 6 = *very high arousal*). We included seven attention checks that read: "If you read this, please answer these questions with a 0". Two participants were excluded due to failure to respond correctly to two or more of these attention checks. Based on the remaining 14 participants, the average valence of all context vignettes was 2.0 ($SD = 0.68$) and the average arousal was 2.98 ($SD = 0.7$), meaning that the context vignettes were mildly negative and medium in their level of arousal.

Second, we normed all vignette-statement combinations for ironicity to verify manipulations. Each item existed in four possible conditions, literal-self, literal-other, irony-self, and irony-other. The 640 total items were spread across four lists (160 items each) using Latin square rotation, so that each participant read each vignette only once, but rated items from all four conditions. Their task

was "to rate the final statement on how ironic it is". We included eight attention checks that read: "If you read this, answer this question with a 1" at the item text position. Fifty-four participants (M_{age} = 19.34 years) were introduced to ironicity with a definition and examples and rated how ironic a statement was on a 5-point Likert type scale, with 0 = *not ironic* and 4 = *very ironic*. Four participants were excluded due to failure to respond correctly to two or more attention checks. Based on the remaining 50 participants, the ironic statements (M = 2.95, SD = 0.61) were statistically significantly more ironic than literal statements (M = 1.17, SD = 0.61), $t(638) = 36.87$, $p < .001$. There was no difference in how ironic a statement was based on whether it was directed at the self (M = 2.91, SD = 0.62) or the other person (M = 2.98, SD = 0.61). As expected, nonironic items did not show any difference in ironicity ratings based on the addressee.

Third, we collected humor ratings for all vignette-statement combinations to explore the role of humor in our ratings (cf. Bowes & Katz, 2011; Toplak & Katz, 2000). Humor might be a potential mechanism of irony to reduce negative emotion, and therefore, might moderate any effects of irony on speaker and listener emotional state. Using the same four experimental lists and procedures used in the ironicity norming, participants (N = 64, M_{age} = 21.26) rated how funny a statement was on a 5-point Likert-type scale (0 = *not funny*, 4 = *very funny*). Their task was "to rate the final statement on how funny it is". Three participants were excluded due to failure to respond correctly to two or more attention check questions. Results from the remaining 61 participants using a repeated-measures analysis of variance (ANOVA) of 2 × 2 [addressee (self, other) × ironicity (literal, irony)] showed that irony (M = 1.07, SD = 0.35) was rated as statistically significantly funnier than literal (M = 0.8, SD = 0.30), confirmed by a follow up paired t test, $t(638) = 10.53$, $p < .001$). In addition, statements directed at the other person (M = 0.96, SD = 0.35) were rated as statistically significantly funnier than those directed at the self (M = 0.9, SD = 0.35), as confirmed by a follow up paired t test, $t(638) = 2.04$, $p = .042$. No interaction between ironicity and addressee was observed. Numeric results are presented in Table 2.

The final set contained 160 vignettes in four conditions each, for a total of 640 items. All items were rotated using Latin square rotation and separated into four lists of 160 items, such that each list only had one version of each item. Then, all four lists were duplicated, so that on one version, half of the items were rated from the speaker's perspective, and half of the items were rated from the listener's perspective, and on the second version of each list, the opposite perspective was rated, resulting in a total of eight experimental lists. Items were presented in randomized order, and upon clicking the study link, participants were randomly assigned to one of the eight lists.

Proceure

All parts of the study were administered online using Qualtrics (Qualtrics,

Table 2. *Norming Results for Ironicity and Humor. Mean Values with SDs in Parentheses.*

M (*SD*)	Ironicity (0-4)	Humor (0-4)
irony-other	2.98 (0.61)	1.1 (0.34)
irony-self	2.91 (0.62)	1.03 (0.35)
literal-other	1.18 (0.62)	0.82 (0.3)
literal-self	1.16 (0.59)	0.78 (0.29)

Provo, UT). Participants consented to participate and completed a language screening questionnaire to ensure meeting the inclusion criterion of being a native speaker of English. Then, they were introduced to the concepts of valence and arousal using definitions and examples, and were told that sometimes they might be rating the speaker, and sometimes they might be rating the listener. They were also instructed to consider the impact of each statement during their ratings. During the rating part, they rated how the protagonist (name marked in bold) felt using two 7-point Likert-type scales, one for valence ("How does the protagonist feel?" 0 = *very negative*, 6 = *very positive*) and one for arousal ("How intense does the protagonist feel?" 0 = *very low arousal*, 6 = *very high arousal*). The procedure was piloted using undergraduate research assistants, who confirmed that the factor of role (rating from the speaker's vs. the listener's perspective) highlighted via boldfacing the protagonist's name was clear during the ratings. Two types of attention checks were implemented to ensure active participation during the rating part of the study: (1) mock items that read: "This is an attention check. Please answer the two questions on this page as follows: Please answer these questions with 0" And (2) definition questions, where participants were asked to describe what valence and arousal mean, respectively. After all items were rated, participants completed the Interpersonal Reactivity Index (IRI, Davis, 1980), a 28-item measure of empathy, where participants read statements (e.g., I really get involved with the feelings of the characters in a novel.) and indicate if a statement describes them well or not on a 5-point letter scale. The experiment lasted about 90 minutes.

Analyses

All statistical analyses were computed using R (4.2.2) and R Studio (version 2022.12.0). Additional analyses are reported in the supplementary materials in the study's OSF page. Post-hoc power was determined via MorePower 6.0 (Campbell & Thompson, 2012). With the sample size of 164 participants, we could detect effects as small as $\eta^2 = .05$ at 82.9% power, suggesting that even small effects could be detected with our present sample size. All analyses scripts are available in the study OSF: https://osf.io/u7whr/.

Results

Valence Ratings

We used two ways of analyzing the data. First, we used the preassigned labels for items. A model using this data revealed a main effect of role, as well as main effects of context valence and humor, and a statistically significant interaction between addressee and role. It did not reveal statistically significant main effects or interactions for ironicity. The random effect of participant was 0.21 (*SD* = 0.46), and the random effect of item was 0.06 (*SD* = 0.24). The total variance explained by this model was R^2 = 22.9%, with R^2_{adj} = 8.4% of the variance explained by fixed effects, and ICC = 15.9% explained by random effects. The full model is reported in Table 3.

Second, an anonymous Reviewer suggested to better utilize the available ironicity ratings in the model. We did this in two ways. First, we divided the items based on their observed ironicity ratings, such that items with an ironicity rating below 2 were labeled "literal," and those with ratings above 2 were labelled "irony." The results are shown in Table 4. We found similar main effects for role, context valence, and humor, as well as the addressee × role interaction that we observed with the preassigned label. In addition, we observed a main effect of addressee, and statistically significant interactions of addressee × ironicity, as well as a statistically significant three-way-interaction of addressee × role × ironicity. The random effect of participant was 0.21 (*SD* = 0.46), and the random effect of item was 0.06 (*SD* = 0.24). The total variance explained by this model was R^2 = 22.9%, with R^2_{adj} = 8.4% of the variance explained by fixed effects, and ICC = 15.9% explained by random effects. An ANOVA comparing model fit between the observed and assigned binary ironicity factor favored neither model, but the observed binary ironicity model had a slightly lower Akaike information criterion (AIC), a measure of model fit) score.

In addition, we used a model using the actual observed ironicity rating for each item on a 0–4 scale rather than the binary factor of ironicity in the models reported above. The results for this model are reported in Table 5. For the valence model, we found main effects for addressee, role, context valence, and humor, but not for ironicity. We further found a statistically significant addressee × role interaction, and an addressee × ironicity interaction, and a statistically significant three-way interaction of role × addressee × ironicity. Specifically, as seen in Figure 1, speakers in the self condition were rated as feeling more positive (higher in valence) with increased ironicity scores, compared to listeners in the self condition, who were rated as feeling the same level of valence, regardless of item ironicity. Notably, speakers in the other condition were rated as feeling more positive than listeners, regardless of item ironicity. The random effect of participant was 0.21 (*SD* = 0.46), and the random effect of item was 0.06 (*SD* = 0.24). The total variance explained by this model was R^2 = 22.9%, with R^2_{adj} =

Table 3. *Results From the Linear Mixed-Effect Model with Valence as a Dependent Variable. This Model Used the Preassigned Condition Labels for Ironic and Literal Statements.*

Assigned	Estimate	SE	*df*	*t*	*p*	2.50%	97.50%
(Intercept)	1.322	0.104	249.712	12.749	**< .001**	1.118347	1.524755
Addressee - self	−0.021	0.029	26082.77	−0.725	.468	−0.07914	0.036393
Role - listener	−0.254	0.029	26073.55	−8.615	**< .001**	−0.31143	−0.196
Ironicity	−0.015	0.031	26224.59	−0.484	.628	−0.07527	0.045428
Context valence	0.56	0.03	157.993	18.79	**< .001**	0.501375	0.618384
Humor	0.131	0.031	10062.95	4.144	**< .001**	0.068813	0.192205
Addressee × Role	0.24	0.042	26072.07	5.751	**< .001**	0.157896	0.321119
Addressee × Ironicity	0.018	0.042	26077.71	0.436	.663	−0.06349	0.099803
Role × Ironicity	−0.016	0.042	26075.47	−0.394	.694	−0.09803	0.065233
Addressee × Role × Ironicity	0.006	0.059	26075.6	0.096	.924	−0.10981	0.121076

Table 4. *Results From the Linear Mixed-Effect Model With Valence as a Dependent Variable. This Model Used the Observed Binary Condition Labels for Ironic (< 2) and Literal (> 2) Statements.*

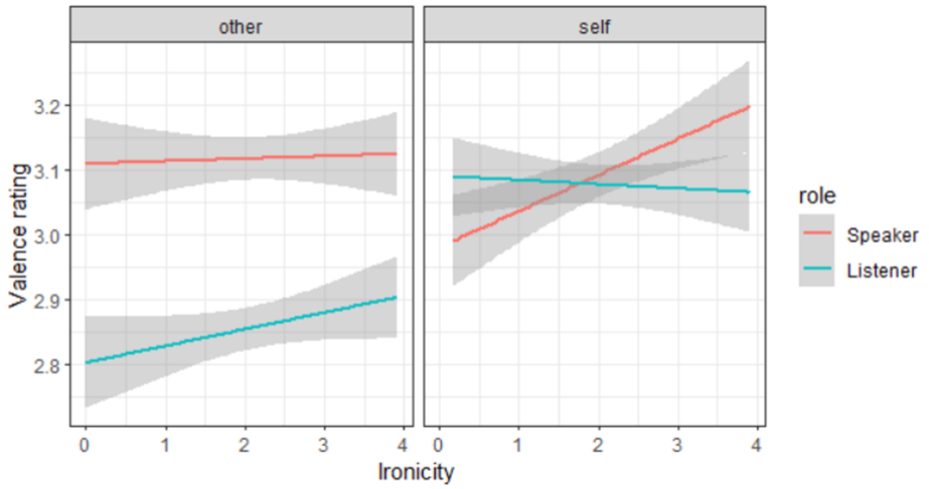
Observed	Estimate	SE	*df*	*t*	*p*	2.50%	97.50%
(Intercept)	1.352	0.104	251.971	13.006	**< .001**	1.148127	1.555594
Addressee - self	−0.075	0.03	26128.56	−2.545	**.011**	−0.13356	−0.01735
Role - Listener	−0.293	0.03	26074.02	−9.906	**< .001**	−0.35045	−0.23469
Ironicity	−0.058	0.032	25637.03	−1.826	.068	−0.12031	0.004223
Context Valence	0.56	0.03	158.08	18.78	**< .001**	0.501092	0.6181
Humor	0.122	0.033	9444.025	3.717	**< .001**	0.057922	0.187036
Addressee × Role	0.318	0.042	26072.01	7.616	**< .001**	0.235949	0.399454
Addressee × Ironicity	0.125	0.042	26176.88	2.974	**.003**	0.042583	0.207129
Role × Ironicity	0.061	0.042	26076.24	1.463	.144	−0.0207	0.142575
Addressee × Role × Ironicity	−0.15	0.059	26075.42	−2.55	**.011**	−0.26561	−0.03476

8.4% of the variance explained by fixed effects, and ICC = 15.9% explained by random effects. An ANOVA comparing model fit between the observed binary ironicity factor and the continuous observed ironicity factor favored neither model, but the observed continuous ironicity model had a slightly lower AIC score, meaning it explained more, but not significantly more of the variance in valence ratings.

Arousal Ratings

Mirroring the analyses for valence, we ran three models for arousal. Using the preassigned binary irony factor, we found a statistically significant main effect of addressee, role, and context valence, in addition to a statistically significant interaction of addressee by role. The full model is reported in Table 6. No main effects or interactions with ironicity were observed. The random effect of participant was 0.68 (*SD* = 0.82), and the random effect of item was 0.23 (*SD*

Figure 1. *Interaction of role (listener/speaker), addressee (self/other), and ironicity (literal/irony) on valence ratings.*



Note. Left panel is the other-condition, right panel is the self-condition. Speaker is in red and listener is in blue. Ironicity ratings are on the X-axis, valence ratings are on the Y-axis.

Table 5. *Results From the Linear Mixed-Effect Model with Valence as a Dependent Variable. This Model Used the Observed Continuous Ironicity Ratings for Each Item, Rather than a Binary Condition Factor.*

	Estimate	SE	*df*	*t*	*p*	lower CI	Upper CI
(Intercept)	1.374	0.109	307.003	12.611	**< .001**	1.160527	1.587501
Addressee - Self	−0.178	0.063	26151.43	−2.812	**.005**	−0.30174	−0.0539
Role - Listener	−0.332	0.063	26077.51	−5.291	**< .001**	−0.4543	−0.20872
Ironicity	−0.012	0.015	25505.16	−0.813	.416	−0.0412	0.017011
Context Valence	0.56	0.03	157.904	18.784	**< .001**	0.501262	0.618284
Humor	0.107	0.034	8912.849	3.128	**.002**	0.040049	0.174289
Addressee × Role	0.497	0.089	26076.84	5.592	**< .001**	0.322765	0.671081
Addressee × Ironicity	0.054	0.02	26168.13	2.758	**.006**	0.0156	0.092182
Role × Ironicity	0.023	0.019	26078.46	1.178	.239	−0.015	0.060188
Addressee x Role × Ironicity	−0.083	0.027	26077.87	−3.045	**.002**	−0.13722	−0.02977

= 0.48). The total variance in arousal ratings explained by this model was R^2 = 35.2%, with R^2_{adj} = 2.3% of the variance explained by fixed effects, and ICC = 33.6% explained by random effects.

For the model using the observed ironicity ratings as a binary factor, we found the same main effects and interactions as in the originally assigned labels. Full results are reported in Table 7. The random effect of participant was 0.68 (SD = 0.82), and the random effect of item was 0.23 (*SD* = 0.48). The total variance in arousal ratings explained by this model was R^2 = 35.2%, with R^2adj

Table 6. *Results From the Linear Mixed-Effect Model with Arousal as a Dependent Variable. This Model Used the Preassigned Condition Labels for Ironic and Literal Statements.*

assigned	Estimate	SE	*df*	*t*	*p*	2.50%	97.50%
(Intercept)	4.682	0.192	214.722	24.403	**< .001**	1.118347	1.524755
Addressee - self	0.085	0.033	26073.74	2.569	**.01**	−0.07914	0.036393
Role - Listener	0.269	0.033	26069.87	8.165	**< .001**	−0.31143	−0.196
Ironicity	−0.017	0.035	26220.43	−0.494	.621	−0.07527	0.045428
Context Valence	−0.345	0.058	158.005	−5.991	**< .001**	0.501375	0.618384
Humor	0.021	0.037	22544.45	0.582	.561	0.068813	0.192205
Addressee × Role	−0.358	0.047	26069.37	−7.674	**< .001**	0.157896	0.321119
Addressee × Ironicity	0.035	0.047	26071.52	0.749	.454	−0.06349	0.099803
Role × Ironicity	−0.039	0.047	26070.52	−0.831	.406	−0.09803	0.065233
Addressee × Role × Ironicity	0.017	0.066	26070.57	0.259	.796	−0.10981	0.121076

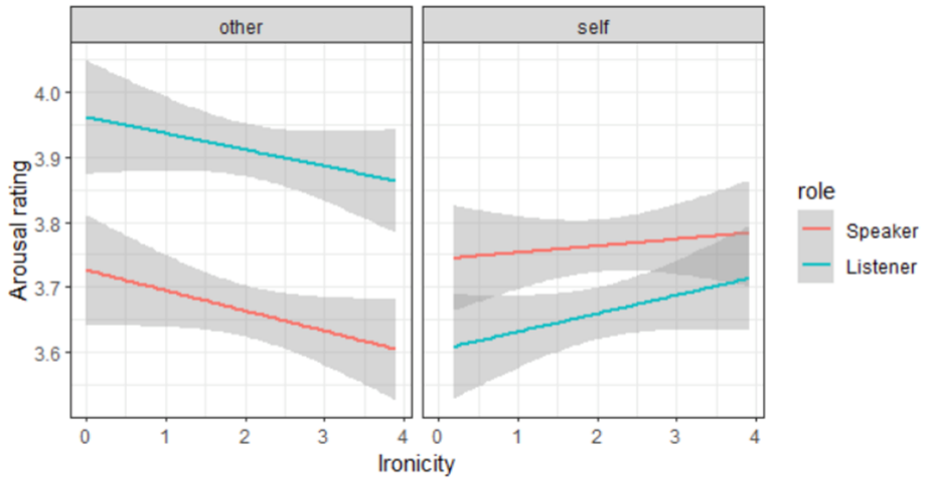
Table 7. *Results From the Linear Mixed-Effect Model with Arousal as a Dependent Variable. This Model Used the Observed Binary Condition Labels for Ironic (< 2) and Literal (> 2) Statements.*

observed	Estimate	SE	*df*	*t*	*p*	2.50%	97.50%
(Intercept)	4.668	0.192	215.57	24.28	**< .001**	1.148127	1.555594
Addressee − self	0.081	0.033	26090.27	2.429	**.015**	−0.13356	−0.01735
Role - Listener	0.262	0.033	26070.03	7.918	**< .001**	−0.35045	−0.23469
Ironicity	−0.066	0.036	26249.3	−1.854	.064	−0.12031	0.004223
Context Valence	−0.344	0.058	158.034	−5.968	**< .001**	0.501092	0.6181
Humor	0.059	0.038	22146.63	1.532	.126	0.057922	0.187036
Addressee × Role	−0.362	0.047	26069.32	−7.747	**< .001**	0.235949	0.399454
Addressee × Ironicity	0.047	0.047	26112.01	1	.317	0.042583	0.207129
Role × Ironicity	−0.024	0.047	26070.82	−0.513	.608	−0.0207	0.142575
Addressee × Role × Ironicity	0.025	0.066	26070.48	0.376	.707	−0.26561	−0.03476

= 2.4% of the variance explained by fixed effects, and ICC = 33.7% explained by random effects. An ANOVA comparing the observed versus assigned binary ironicity factor favored neither model, but the observed model had a lower AIC.

Finally, for the arousal model using the continuous ironicity ratings (see Table 8), we found statistically significant main effects for addressee, role, and context valence, as well as an addressee × role interaction. Specifically, in the other condition, listeners were rated as feeling more aroused compared to speakers. In the self condition, this was reversed, such that listeners were rated as feeling less aroused compared to speakers. This effect is illustrated in Figure 2. The random effect of participant was 0.68 (*SD* = 0.82), and the random effect of item was 0.23 (*SD* = 0.48). The total variance in arousal ratings explained by this model was R^2 = 35.2%, with R^2_{adj} = 2.4% of the variance explained by fixed effects, and ICC = 33.6% explained by random effects. An ANOVA comparing the observed binary ironicity factor versus the observed continuous ironicity factor favored neither model, but the observed binary model had a lower AIC.

Figure 2. *Interaction of role (listener/speaker) and addressee (self/other) on arousal ratings.*



Note. Left panel is the other condition, right panel is the self condition. Speaker is in red and listener is in Blue. Ironicity ratings are on the X axis, arousal ratings are on the Y axis.

Table 8. *Results From the Linear Mixed-Effect Model with Arousal as a Dependent Variable. This Model Used the Observed Continuous Ironicity Ratings for Each Item, Rather Than a Binary Condition Factor.*

	Estimate	SE	*df*	*t*	*p*	lower CI	Upper CI
(Intercept)	5.001	0.195	231.431	25.583	**< .001**	4.617336	5.383905
Addressee / Self	−0.367	0.071	26099.76	−5.188	**< .001**	−0.50589	−0.22848
Role / Listener	−0.253	0.07	26071.38	−3.597	**< .001**	−0.39008	−0.11492
Ironicity	−0.033	0.017	26265.12	−1.997	.046	−0.06616	−0.00062
Context Valence	−0.345	0.058	158.004	−5.985	**< .001**	−0.45791	−0.23167
Humor	0.048	0.04	21736.26	1.188	.235	−0.03097	0.126115
Addressee × Role	0.38	0.1	26071.15	3.814	**< .001**	0.184628	0.574895
Addressee × Ironicity	0.039	0.022	26107.72	1.807	.071	−0.00334	0.082309
Role × Ironicity	0.001	0.021	26071.73	0.036	.971	−0.04135	0.04289
Addressee × Role × Ironicity	−0.01	0.031	26071.52	−0.322	.748	−0.07009	0.05031

Discussion

We set out to investigate the influence of irony on perceived emotional states. Specifically, we contrasted ironic and literal statements made about the self or another person, and asked participants to either rate the speaker's or the listener's perspective. We hypothesized that the presence of irony would result in less negative and less aroused perceived mental states, and that speakers and listeners would be impacted differentially by statements directed at them or not.

Our results showed that the perceived emotional state of a protagonist depends

on a number of factors. First, we found a main effect of role and an addressee × role interaction in context valence ratings. This is in line with our prediction that whether a character is a speaker or a listener, and whether statements are directed at the self or another person, matters for the perceived emotional state. This result and shows that participants engaged with the story content in a meaningful way because it seems intuitive that the listener in the other condition is feeling more negative than the speaker, as the speaker is making a statement directed at the listener, possibly about something they have done, which might contain blame, criticism, or negative sentiment.

Regarding our hypothesis concerning the effects of ironicity on perceived emotional states, we found mixed support. When using the labels we assigned to our stimuli before the experiment, we did not find any statistically significant main effects or interactions with ironicity. However, when we considered the ironicity ratings for each item collected during the norming of our materials, we found that ironicity interacted in a three-way interaction with the factors of role and addressee. Model fit assessed with AIC favored this model, due to more explained variance. Breaking down the interaction of ironicity, role, and addressee in valence ratings, the more ironic the items were, the more speakers in the self condition were rated as feeling more positive (higher in valence). Listeners in the self condition were rated as feeling the same level of valence, regardless of how ironic items were. This finding can be interpreted such that a speaker engaging in self-irony is perceived as feeling more positive, or less negative, than a speaker engaging in self-literal talk. This finding is in line with the idea that verbal irony can index a more downregulated emotional state (cf. Pfeifer & Lai, 2021) and highlights that self-irony might be a special case of irony. Such a finding is not in line with Bajerski (2016)'s idea that self-irony is unsuitable to release emotional tension.

Notably, in the other condition, regardless of how ironic items were, listeners were rated as feeling more negative (lower in valence) than speakers. This is in line with the idea that when a statement is directed at a person, the receiver of the statement (the listener) is feeling more negative compared to the sender of the statement (the speaker).

In addition, we found a main effect of humor and context valence on valence ratings. This means that, the more positive a communicative context situation is, the more positive the characters in it are perceived. Similarly, if a statement, regardless of whether it is ironic or not, is more humorous, characters speaking and listening to it are also rated as feeling more positively. Both of these findings make intuitive sense and confirm that our participants attentively read stories and rated characters based on the information they extracted from the text.

Regarding arousal, we found main effects of role and addressee as well as their interaction. These findings mirror findings in the valence analysis, such that a listener in the other condition (meaning when a statement is directed at them) is perceived as significantly more aroused than a speaker, and significantly more

aroused than speakers and listeners in the self condition. Once more, this is in line with the idea that when a statement, regardless of whether it is ironic or not, is directed at the listener, they are rated as being higher in arousal compared to when they are not targeted with a statement.

We did not observe any main effects or interactions with ironicity in arousal ratings, neither when using the assigned label or the observed ironicity ratings. This is in contrast to prior work that found that ironic speakers were rated to be less aroused than literal speakers (Pfeifer & Lai, 2021). However, our stories were less negative and less arousing than the stories in Pfeifer and Lai (2021). Pfeifer and Lai had two types of contexts, high and low emotion, rated on a 0-4 scale, with high emotion rated 0.65 for valence and 2.90 for arousal, and low emotion rated as 1.20 for valence and 2.21 for arousal. Converting their numbers to our 0-6 scale, the high emotion condition would be 0.975 for valence and 4.35 for arousal, and the low emotion condition would be 1.8 for valence and 3.315 for arousal. In comparison to ours (2.0 for valence and 2.98 for arousal), our stories were indeed less negative and especially less arousing than their stories, which may have resulted in overall more similarity perceived emotional states for all characters. While more research is needed, it is possible that context arousal impacts the role of irony in emotional states.

We found a significant main effect for context valence on arousal ratings, such that the more positive a context was, the less aroused a character was rated, in line with the valence findings. Notably, no effect of humor on arousal ratings was found, suggesting that humor did not affect the perceived intensity of emotion in characters. Another noteworthy finding is that in arousal ratings, random effects of item and participant explained almost 34% of the variance, whereas in valence ratings, they only explained 16% of the variance, suggesting that arousal might either be a more abstract, or more personal and situationally dependent aspect of emotional experience than valence.

Our norming results showed that irony is equally ironic whether it is directed at the self or another person, when used in the same context. This result demonstrates that despite serving different communicative functions when directed at the self or another person, in the same context, an utterance is perceived as equally ironic. In the case of self-directed irony, the speaker is conveying something about themselves, either their own negative emotions about the event via self-talk, or a signal to the listener that they understand that they have made a mistake. In both cases, the listener is a bystander and has little involvement. On the contrary, in the case of other-directed irony, the speaker might be conveying criticism towards the other person, or accusing them by directing their statement at them, which makes the statement socially more relevant from the listener's perspective. Thus, our finding raises interesting questions about what makes a statement ironic, and how ironicity and social signals work in unison.

We did not find main effects of ironicity in valence or arousal ratings, which means we failed to replicate, in part, prior literature demonstrating differences

between irony and literal language on various measures of emotional experience. Such effects include reduced ratings of speaker negativity (Pfeifer & Lai, 2021), irony being more mocking (Ivanko & Pexman, 2003), more condemning (Colston, 1997; 2002), more humorous (Dews, Kaplan & Winner, 1995), or in general different from literal language in any form (Colston & O'Brien, 2000; Bowes & Katz, 2011; Boylan & Katz, 2013; Pexman & Olineck, 2002; Toplak & Katz, 2000). A few possible explanations are outlined below.

Limitations

Limitation of the method used in the current study is that we used participants, who were uninvolved bystanders, to read short stories about characters and then rate the characters' emotional state. This requires considerable perspective-taking ability and imagination, and might not reflect the actual emotional state of speakers and listeners when encountering irony "in the wild." While such bystander ratings are common practice in (irony) research and beyond, this methodological choice may have further contributed to the absence of effects. This limitation could be addressed by involving participants as irony producers or recipients, albeit this would require different experimental approaches.

Relatedly, irony might create several ambiguous or mixed emotional states (such as being simultaneously amused and embarrassed) that cannot be fully captured with unidimensional scales such as valence and arousal, as used here. Thus, future work might consider using additional or different scales to account for the potential of mixed or ambiguous emotional states (c.f. Pfeifer & Pexman, 2023).

Another limitation lies in the materials used. While the context emotion in our items was generally negative, the magnitude of contrast between the ironic statements and the context could have been stronger (cf. Colston & O'Brien, 2000; Boylan & Katz, 2013). In Colston and O'Brien (2000), ironic statements using stronger contrasts to the context (e.g., using "magnificent mood" instead of "agreeable mood") resulted in larger differences compared to literal language for condemnation, humor, and speaker protection ratings. We did not measure the amount of contrast expressed between our literal and ironic words and the situation, and neither did Pfeifer and Lai (2021), Ivanko and Pexman (2003), Dews et al. (1995), or Toplak and Katz (2000); but it is possible that the lack of contrast between the situation and the statement was too weak to reflect upon the emotional state of the speaker. Future studies could improve materials by creating more emotional contexts (cf. Ivanko & Pexman, 2003), more aggressive exchanges (cf. Bowes & Katz, 2011), or by employing more hyperbolic language (cf. Colston & O'Brien, 2000). Related to this, since we used written context, implicit prosody may have impacted sentence interpretation (cf. Breen, 2014). Thus, future studies could use spoken language.

Lastly, a limitation, or opportunity for further investigation, is humor. Our norming data revealed that ironic statements were rated as more funny than

literal ones, both were rated as 0.5-1.5 on the 0-4 scale, suggesting they were not ”very funny” to participants. We did observe that humor had a significant effect on character’s perceived emotional states in valence, but not arousal. Thus, exploring the role of humor more directly in irony could lead to further insight into ironic processes.

Conclusion

Taken together, speakers and listeners are perceived to be in different emotional states, depending on whether they are addressed by a statement (ironic or literal) or not. This is reflected in two key results. First, self-irony seems to be a special case, where the use of such irony indicates a more positively perceived emotional state for the speaker. In other words, self-irony might serve emotional self-regulation functions, and our findings underline the need to study such irony more. Importantly, our study found such effect only when considering the ironicity ratings of each individual item, suggesting that the more ironic an item is, the more positive the speaker feels. This might be an issue of ambiguity, where the more easily an item is recognized as ironic, the more clearly a bystander perceives the speaker as feeling positive. However, more research into self-irony is needed to solidify this finding.

Second, when listeners are addressed by statements (“How clumsy/careful of you”), regardless of whether they are ironic or not, they are rated to feel more negative and more aroused than when statements are not directed at them (“How clumsy/careful of me”) or, when compared to speakers of these statements. Thus, our results empirically demonstrate, as intuition suggests, that being the addressee of a statement matters for a perceived emotional state.

In conclusion, our study demonstrates that both, whether we say or hear a comment in a communicative exchange, and whether we are addressed by a comment or not, matter for the emotional impact of the comment. Self-irony presents a special case deserving of further empirical inquiry.

Conflict of Interest Disclosure

The Authors report no conflicts of interest.

Funding

The Authors report no sources of funding.

Research Ethics Statement

Criteria for Approval has been met (45 CFR 46.111): The criteria for approval listed in 45 CFR 46.111 have been met (or if previously met, have not changed) in that (1) Risks to subjects are minimized: (i) By using procedures which are consistent with sound research design and which do not unnecessarily expose subjects to risk, and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes. (2) Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research). The IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility. (3) Selection of subjects is equitable. In making this assessment the IRB should take into account the purposes of the research and the setting in which the research will be conducted and should be particularly cognizant of the special problems of research involving vulnerable populations, such as children, prisoners, pregnant women, mentally disabled persons, or economically or educationally disadvantaged persons. (4) Informed consent will be sought from each prospective subject or the subject's legally authorized representative, in accordance with, and to the extent required by 46.116. (5) Informed consent will be appropriately documented, in accordance with, and to the extent required by 46.117. (6) When appropriate, the research plan makes dequate provision for monitoring the data collected to ensure the safety of subjects. (7) When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data. (b) When some or all of the subjects are likely to be vulnerable to coercion or undue influence, such as children, prisoners, pregnant women, mentally disabled persons, or economically or educationally disadvantaged persons, additional safeguards have been included in the study to protect the rights and welfare of these subjects.

Authorship Details

Valeria A. Pfeifer: research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article.

Matthias R. Mehl: research concept and design, critical revision of the article, final approval of the article.

Vicky Tzuyin Lai: research concept and design, writing the article, critical revision of the article, final approval of the article.

References

Anolli, L., Infantino, M. G., & Ciceri, R. (2002). " You're a real genius!": Irony as a miscommunication design. *Emerging communication, 3*, 135–158.

Bajerski, M. (2016). " I understand you, so I'll not hurt you with my irony": Correlations between irony and emotional intelligence. *Psychology of Language and Communication, 20*(3), 235–254. https://doi.org/10.1515/plc-2016-0015

Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion, 12*(4), 579–599. https://doi.org/10.1080/026999398379574

Bowes, A., & Katz, A. (2011). When sarcasm stings. *Discourse Processes, 48*(4), 215–236. https://doi.org/10.1080/0163853X.2010.532757

Boylan, J., & Katz, A. N. (2013). Ironic expression can simultaneously enhance and dilute perception of criticism. *Discourse Processes, 50*(3), 187–209. https://doi.org/10.1080/0163853X.2012.747041

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass, 8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

Campbell, J. I., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods, 44*(4), 1255–1265. https://doi.org/10.3758/s13428-012-0186-0

Clark, H. and Gerrig, R. 1984. On the pretence theory of irony. *Journal of Experimental Psychology: General, 113*(1),121–126. https://doi.org/10.1037/0096-3445.113.1.121

Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions of ironic criticism. *Discourse Processes, 23*(1), 25–45. https://doi.org/10.1080/01638539709544980

Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics, 34*(2), 111–142. https://doi.org/10.1016/S0378-2166(02)80008-X

Colston, H. L., & O'Brien, J. (2000). Contrast of kind versus contrast of magnitude:

The pragmatic accomplishments of irony and hyperbole. *Discourse processes, 30*(2), 179–199. https://doi.org/10.1207/S15326950DP3002_05

Davis, M. H. (1980). *Interpersonal Reactivity Index (IRI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t01093-000

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

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

Gibbs, R. W., Bryant, G. A., & Colston, H. L. (2014). Where is the humor in verbal irony? *Humor, 27*(4), 575–595. https://doi.org/10.1515/humor-2014-0106

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes, 36*(3), 147–165. https://doi.org/10.1207/S15326950DP3603_1

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

Ortega, M. B. A. (2013). An approach to verbal humor in interaction. *Procedia-Social and Behavioral Sciences, 95*, 594–603. https://doi.org/10.1016/j.sbspro.2013.10.687

Pexman, P. M., & Olineck, K. M. (2002). Does sarcasm always sting? Investigating the impact of ironic insults and ironic compliments. *Discourse Processes, 33*(3), 199–217. https://doi.org/10.1207/S15326950DP3303_1

Pfeifer, V. A., & Lai, V. T. (2021). The comprehension of irony in high and low emotional contexts. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 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

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

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. *Philosophy, 3*, 143–184.

Toplak, M., & Katz, A. N. (2000). On the uses of sarcastic irony. *Journal of Pragmatics, 32*(10), 1467–1488. https://doi.org/10.1016/S0378-2166(99)00101-0

Appendix

Individual Differences in Irony

Individual differences in verbal irony processing, most notably empathy and perspective taking, have occasionally been considered in past research, but no clear picture has emerged (see Kałowski et al., 2023, for a recent review). In healthy adults, Pfeifer and Lai (2021) used the Autism Quotient (AQ, Baron-Cohen et al., 2001) to measure perspective taking ability and found a relationship between the subscales of attention switching and imagination and their irony-related event-related potential (ERP) effect amplitude. Specifically, those who showed worse attention switching and imagination skills showed smaller ERP amplitude differences at P600. However, their sample size for the correlation was small ($N = 35$) and the correlation strength was moderate. Puhachneuskaya and Järvikivi (2022) used the Interpersonal Reactivity Index (IRI, Davis, 1980) to approximate empathy in a study where participants rated how ironic a speaker was following literal and ironic compliments and criticism but did not find statistically significant correlations between IRI and irony ratings. Finally, Kieckhäfer et al. (2019) administered an irony detection test, and collected measures on several additional scales, including the IRI, the AQ, and two measures of personality (schizotypal and borderline). They found a relationship between irony detection skills and personality traits, but not for autistic personalities (individuals who scored high on the AQ) as measured by the AQ. In particular, in individuals with schizotypal personalities, the IRI subscale of personal distress was negatively correlated with irony detection such that those who scored higher on the personal distress subscale were worse at irony recognition. Based on these mixed results, it has been hypothesized that tests like the AQ and IRI are not sensitive enough to pick up differences in healthy young adults. In work with patients, the IRI and AQ have been used as well. For example, in those with cerebral damage, a negative relationship between theory of mind as measured with an irony task and empathy was found by Shamay-Tsoory et al., (2005), such that individuals with worse perspective-taking skills also showed lower empathy scores on the IRI. Similarly, Deliens et al., (2018) found that AQ scores could predict performance on an irony detection task via videos, but only in individuals with autism spectrum disorder (but see Nuber et al., 2018 for conflicting results).

Predictions

We predicted our results to be modulated by empathy as measured with the IRI (Davis, 1980). Specifically, we predicted that individuals who scored higher on the perspective-taking subscale of the IRI would show a greater sensitivity/difference in emotional state ratings between ironic and literal statements (perceived emotional state ratings following irony minus perceived emotional

state ratings following literal statements). Further, we predicted the same relationship of higher perspective-taking skills and larger difference between speaker and listener ratings in the self and other conditions.

Analyses

For the analysis, ratings were averaged per participant across items. Ratings that were outside of 2 *SD*s of the item mean were excluded (less than 1% of total responses). All results from the individual differences analyses were corrected for multiple comparisons using FDR. The IRI responses were scored using Davis' (1980) manual and resulted in one value per participant for each of the subscales: fantasy (FS), personal distress (PD), perspective taking (PT), and empathic concern (EC). Each subscale ranged from 0 to 28, with higher scores indicating stronger tendencies towards that dimension.

Results for Individual Differences

The means of IRI subscales and self-rated irony comprehension are reported in Table 1A. We correlated self-rated irony comprehension ("How well do you think you are able to understand irony? 0 = *never*, 6 = *always*) with each of the subscales of the IRI and found that there was a weak (statistically nonsignificant) positive relationship between PT and self-rated irony comprehension, $r = .14$, $p = .06$. The higher a participant scored on PT, the higher they rated their irony comprehension. No other subscale of the IRI was correlated to self-rated irony comprehension (all *p*s >.1)

Next, we calculated an irony valence sensitivity score, which was a participant's average valence rating in the literal conditions minus the ratings in the respective irony conditions. The idea was that participants with higher perspective-taking scores might be more sensitive to a difference in emotional state based on whether irony or literal language was used (cf. the Predictions section). In other words, participants rating characters in the irony and literal conditions as feeling more similar would have a score closer to zero, whereas participants who differentiated how characters feel in the ironic and literal conditions more would have either positive or negative scores. Self-rated irony comprehension was negatively correlated with irony valence sensitivity when statements were directed at the other person and the listener's perspective was rated ($r = -.15$, $p = .044$). In other words, people who self-rated their irony comprehension as higher tended to rate listeners as feeling more negative in the irony condition compared to the literal condition, or more similar to each other, compared to those who reported less irony comprehension. No such relationship was present

Table A1. *Average IRI Subscale Scores and Irony Comprehension Score from the Individual Differences Analyses.*

	M (*SD*)
Empathic concern (IRI)	20.09 (4.49)
Fantasy (IRI)	17.36 (5.88)
Personal distress (IRI)	12.64 (5.08)
Perspective taking (IRI)	18.44 (4.48)
Irony comprehension	5.32 (1.08)

for speaker ratings, and no relationships between irony comprehension and arousal ratings were found. In addition, the irony valence sensitivity was not related to any of the IRI subscale scores.

Discussion

We did not find a relationship between experimental ratings and individual IRI scores. This is in line with some recent findings on irony ratings (Puhachneuskaya & Järvikivi, 2022), where no relationship between the IRI and participants' ratings was observed. We used the IRI to approximate empathy, yet it seems that either empathy does not significantly affect how a participant rates interlocutors' emotional states, or the IRI is not an ideal measure of empathy. One potential explanation brought forward by Puhachneuskaya and Järvikivi (2022) is their homogenous participant pool (93 young, female participants). Our study had a larger sample size, yet whether or not our sample is more heterogeneous is questionable. However, we did find a weak correlation between self-rated frequency of irony use and the IRI subscale of perspective taking. Thus, it is possible that perspective taking ability, not empathy in general, contributes to how often people use irony.

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.

Davis, M. H. (1980). *Interpersonal Reactivity Index (IRI)* [Database record]. APA PsycTests. https://doi.org/10.1037/t01093-000

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in autism spectrum disorder: Indirect requests versus irony. *Journal of Autism and Developmental Disorders, 48*(9), 2938–2952. https://doi.org/10.1007/s10803-018-3561-6

Kałowski, P., Zajączkowska, M., Branowska, K., Olechowska, A., Siemieniuk,

A., Dryll, E., & Banasik-Jemielniak, N. (2023). Individual differences in verbal irony use: A systematic review of quantitative psycholinguistic studies. *Metaphor and Symbol, 38*(1), 81–111. https://doi.org/10.1080/10926488.2022.2160253

Kieckhäfer, C., Felsenheimer, A. K., & Rapp, A. M. (2019). A new test for irony detection: The influence of schizotypal, borderline, and autistic personality traits. *Frontiers in Psychiatry, 10*, 28. https://doi.org/10.3389/fpsyt.2019.00028

Nuber, S., Jacob, H., Kreifelts, B., Martinelli, A., & Wildgruber, D. (2018). Attenuated impression of irony created by the mismatch of verbal and nonverbal cues in patients with autism spectrum disorder. *PloS One, 13*(10), e0205750. https://doi.org/10.1371/journal.pone.0205750

Pfeifer, V. A., & Lai, V. T. (2021). The comprehension of irony in high and low emotional contexts. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale. 75*(2), 120–125. https://doi.org/10.1037/cep0000250

Puhacheuskaya, V., & Järvikivi, J. (2022). I was being sarcastic!: The effect of foreign accent and political ideology on irony (mis) understanding. *Acta Psychologica, 222*, 103479. https://doi.org/10.1016/j.actpsy.2021.103479

Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired "affective theory of mind" is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology, 18*(1), 55–67. https://doi.org/10.1097/01.wnn.0000152228.90129.99