AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 9, 2438–2462
https://doi.org/10.1037/xge0001377

Ironic Twists of Sentence Meaning Can Be Signaled by Forward Move of Prosodic Stress

Pauline Larrouy-Maestri[1, 2], Vanessa Kegel[3], Wolff Schlotz[4], Pol van Rijn[1], Winfried Menninghaus[3], and David Poeppel[1, 2, 5, 6]

[1] Department of Neuroscience, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[2] Max Planck NYU, Center for Language, Music, and Emotion (CLaME), New York, United States
[3] Department of Language and Literature, Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[4] Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany
[5] Department of Psychology, New York University
[6] Ernst Strüngmann Institute for Neuroscience, Frankfurt am Main, Germany

Prosodic stresses are known to affect the meaning of utterances, but exactly how they do this is not known in many cases. We focus on the mechanisms underlying the meaning effects of ironic prosody (e.g., teasing or blaming through an ironic twist), which is frequently used in both personal and mass-media communication. To investigate ironic twists, we created 30 sentences that can be interpreted both ironically and nonironically, depending on the context. In Experiment 1, 14 of these sentences were identified as being most reliably understood in the two conditions. In Experiment 2, we recorded the 14 sentences spoken in both a literal and an ironic condition by 14 speakers, and the resulting 392 recorded sentences were acoustically analyzed. In Experiment 3, 20 listeners marked the acoustically prominent words, thus identifying perceived prosodic stresses. In Experiment 4, 53 participants rated how ironic they perceived the 392 recorded sentences to be. The combined analysis of irony ratings, acoustic features, and various prosodic stress characteristics revealed that ironic meaning is primarily signaled by a stress shift from the end of a sentence to an earlier position. This change in position might function as a "warning" cue for listeners to consider potential alternative meanings of the sentence. Thus, beyond giving individual words a stronger contrastive or emphatic role, the distribution of prosodic stresses can also prime opposite meanings for identical sentences, supporting the view that the dynamic aspect of prosody conveys important cues in human communication.

Keywords: prominence, pitch accent, intonation, prosody, voice acoustics

Speech conveys meaning through individual lexical elements as well as their combination. Genuinely acoustic information, that is, *how* something is said, is also of crucial importance. The prosodic aspects of speech include suprasegmental features that are mainly described in terms of pitch, loudness, duration, and timbre (Cruttenden, 1997; Crystal, 1969; Ladd, 1980; Lehiste, 1970). Suprasegmental features can also modulate the meaning of sentences through *stress*. We focus specifically on prosodic stress, which is historically associated with pitch accent and commonly defined as the "strength" of a spoken word relative to the other words in an utterance.

Prosodic stresses affect paralinguistic meaning in different ways. Unlike lexical stress (e.g., OBject vs. obJECT), prosodic stress provides information to the listener about what is new or relevant or the focus within a sentence. Typical textbook examples regarding prosodic stress highlight its contrastive or emphatic aspects. For

This article was published Online First April 20, 2023.
Pauline Larrouy-Maestri https://orcid.org/0000-0001-9245-0743
We thank Denis Stannek, Franziska Hannig, and R. Muralikrishnan for technical support, as well as Gerrit Kentner and Isabelle Franz for their helpful comments on an earlier version of the manuscript.

Wolff Schlotz died before the submission of this article and contributed substantially to this work.

Data and material can be found at the following link: https://doi.org/10.17617/3.IZWNZ1

Pauline Larrouy-Maestri served as lead for conceptualization, data curation, formal analysis, investigation, methodology, supervision, visualization, writing–original draft, and writing–review and editing and contributed equally to project administration, Vanessa Kegel served as lead for project administration and served in a supporting role for conceptualization and writing–original draft. Wolff Scholtz contributed equally to formal analysis, visualization, and writing–original draft. Pol van Rijn served in a supporting role for writing–original draft. Winfried Menninghaus contributed equally to funding acquisition and served in a supporting role for writing–original draft. David Poeppel served as lead for funding acquisition and served in a supporting role for formal analysis, supervision, and writing–original draft. Vanessa Kegel, Wolff Scholtz, and Pol van Rijn contributed to data curation equally. Vanessa Kegel and Wolff Scholtz contributed equally to methodology. Winfried Menninghaus and David Poeppel contributed equally to conceptualization and writing–review and editing.

Correspondence concerning this article should be addressed to Pauline Larrouy-Maestri, Max Planck Institute for Empirical Aesthetics, Grüneburgweg, 14, 60488 Frankfurt am Main, Germany. Email: plm@ae.mpg.de

instance, in "I love to write and to be read," a stress on *I* highlights that the person who loves to write is the speaker and not somebody else; a stress on *love* shows the strength of the enjoyment to write; a stress on *write* pinpoints what the preferred activity is; while a stress on *and* would indicate that being read is perhaps even more important than writing. Depending on this choice of stress placement, the meaning of a sentence can vary. Attracting attention to specific information increases processing efficiency (e.g., Bock & Mazzella, 1983). It has been shown repeatedly that prominent words are better recalled (e.g., Fraundorf et al., 2010; Kember et al., 2021) and elicit faster responses in detection tasks (Akker & Cutler, 2003; Cutler & Foss, 1977). Prosodic stress is thus a useful communicative tool that modulates the meaning of a sentence and influences speech processing. It is hence not surprising that features supporting focus or pitch accents are used in the context of public speech or political debates (e.g., Braga & Marques, 2004; Holliday et al., 2020).

Whereas the described ways of affecting paralinguistic meaning are closely dependent on the semantic content of the words being stressed, one can hypothesize that the number of stresses itself as well as their position also convey information. Using a gating paradigm, Pell and Kotz (2011) demonstrated for emotional prosody that emotion recognition improves as an utterance unfolds, supporting that acoustic (and thus perceptual) information accumulates over time. Stresses at an early position can provide (potentially relevant) meaningful cues earlier than stresses in the middle or end of a sentence. Also, prosodic stresses' location within a sentence shapes the general contour of an utterance. Contours or dynamic trajectories in speech prosody have been shown to be perceptually relevant for the identification of intended messages (e.g., Grichkovtsova et al., 2012) or affective states (e.g., Rodero, 2011). Even within single words, pitch movements are associated with distinct intentions/attitudes of speakers (e.g., Goupil et al., 2021; Ponsot et al., 2018). Such results support the notion that contrasts emerging from the distribution of acoustic information (and by extension prosodic stresses) over time might be particularly informative. Notably, the studies mentioned here make use of the same material across conditions (i.e., across different speakers' attitudes or emotional states). While not implying that the semantic content of the material is irrelevant, this does speak to the assumption that dynamic aspects of prosody make important contributions to conveying paralinguistic information that is relevant for human communication.

The precise acoustic nature of prosodic stress remains debated. Since Lehiste (1976), several studies successfully identified associations between stress or prominence and acoustic characteristics such as hyperarticulation, increased duration, increased intensity, increased spectral energy in high-frequency regions, and salient fundamental frequency (F0) movements (e.g., Breen et al., 2010; Cole, Mo, & Hasegawa-Johnson, 2010; Katsika & Tsai, 2021). However, no straightforward correspondence between stress and any single acoustic parameter (or their combination) has yet been determined that might allow the automatic identification of stressed syllables/words by means of acoustic analyses. As a consequence, the identification of stress/prominence mainly relies on transcribers' perception. Cole and Shattuck-Hufnagel (2016) proposed a prominence score (*p*-score) that can be used for this purpose. To arrive at the *p*-score, a certain number of transcribers are asked to mark the prominent words of a speech sample, and then each word is assigned a score that represents the proportion of transcribers who marked it as prominent. Ranging from zero to one, the *p*-score is thus a quasicontinuous score with zero representing no perceptual salience and one reflecting strong perceptual salience.

In this study, we use the *p*-score measure and its derivatives (capturing positions and patterns) to examine how prosodic stress modulates meaning. Specifically, we investigate the role of prosodic stress in the perception of verbal irony. Irony is a fuzzy concept that has been discussed in competing theories (see Appendix A for an extensive discussion of the definition of irony). Despite the lack of agreement on the definition, there is a relative consensus regarding the existence of *discrepancy* between what is (literally) said and the meaning that is likely intended by the speaker. This discrepancy arises as the recipient tries to map the wording of a sentence onto an interpretation that is a good fit with the context. The literal and ironic meanings associated with a sentence are often (yet by no means always) diametrically opposed; in such cases, the perception of irony primes a full-blown inversion of the literal semantic meaning of the words that are used. Notably, the discrepancy between literal and nonliteral meanings is described as complex by some authors (e.g., Giora, 1995; Sperber & Wilson, 1981) and as requiring a demanding process that involves the contribution of several cues (including speakers' tone of voice or general knowledge about them), which are coordinated with the linguistic content itself to derive the best fit (Pexman, 2008). Another relatively consensual aspect of irony in the literature is that irony usually lacks a factual content dimension, that is, appears to be wholly an evaluative component of communication. Irony is associated with the expression of a feeling or attitude (Grice, 1978); it appears to be a suitable tool for ostensive communication (see Scott-Phillips, 2014 for a discussion of this mode of communication) and more generally to serve interpersonal functions (Pell & Kotz, 2021). This raises the question of the diversity of information (i.e., not only literal vs. ironic interpretation) and its integration in the comprehension of ironic meaning. Typically associated with negative valence (e.g., Dynel, 2018; Garmendia, 2010; Roberts & Kreuz, 1994; Sperber & Wilson, 1986), irony can also be associated with positive meaning (i.e., humor and jocularity, see Gibbs, 2000; Partington, 2007; Roberts & Kreuz, 1994; Schwarz-Friesel, 2009) and even occupy any position on the evaluative continuum (Alba-Juez & Attardo, 2014).

Irony can be signaled by several types of cues: visual (Rockwell, 2001), contextual (Bryant, 2010; Bryant & Fox Tree, 2002; Wang et al., 2006; Woodland & Voyer, 2011), semantic (Kunneman et al., 2015; Liebrecht et al., 2013), and acoustic. However, the relevance of acoustic cues has occasionally been challenged. In fact, irony (or sarcasm, which is often considered as a special case of irony) can be understood based on written words, or with unmarked intonation, suggesting that the vocal attributes might only be optional, nonessential markers of irony (Attardo et al., 2003). In this connection, Rockwell (2000) found that sarcastic utterances that were spontaneously produced (i.e., prompted/induced by a context that tends to prime a sarcastic meaning intention) were not acoustically distinguishable from nonsarcastic utterances; in contrast, when speakers were explicitly instructed to speak sarcastically, their utterances were successfully recognized as sarcastic. This might suggest that prosody is used as a marker of irony in cases where there is not sufficient contextual information. Understood in this way, a characteristic acoustic signature would only be an optional, but not a necessary ingredient of ironic or sarcastic communication.

Moreover, the prosodic marker might not be specific to irony but could serve as a contrastive marker (Attardo et al., 2003) or as a tool for exaggeration (Kreuz & Roberts, 1995) in many contexts. The role of acoustic markers seems also to depend on the valence of the linguistic content of the material (Mauchand et al., 2020). With another approach (presenting filtered material from which local prosodic and verbal cues are deleted), Bryant and Fox Tree (2005) come to the conclusion that there is no particular ironic tone of voice.

On the other hand, several studies support the existence of acoustic cues that drive recognition of irony without the help of visual, contextual, or semantic cues (e.g., Cheang & Pell, 2008; Woodland & Voyer, 2011), particularly when the material is presented in the listeners' mother tongue (Cheang & Pell, 2011). Various acoustic cues have been found to be associated with irony (or sarcasm) in different languages. Table 1 summarizes features relative to F0 and its derivatives (i.e., range, minimum, maximum, standard deviation), rate of speech, and dynamics (principally melodic contour). These observations are based on a variety of material, ranging from very controlled (e.g., based on a recognition task involving 32 native speakers of English and Cantonese, only 15% of previously recorded material was retained for use in Cheang & Pell's 2011 study) to more natural speech (e.g., material from radio broadcasts in Bryant & Fox Tree, 2002, 2005). As illustrated in Table 1, ironic speech is consistently marked by a slower speech rate (or longer duration) across studies. However, the findings are not consistent with respect to other measures (e.g., the F0 range). Also, it is now documented that acoustic cues vary depending on the function of irony (Mauchand et al., 2018).

In addition to the role that the individual acoustic features listed in Table 1 have in ironic or sarcastic expressions, several studies mention the relevance of cue combinations that are usually associated with prosodic stress (see Anolli et al., 2002; Attardo et al., 2003; Cutler, 1974; Haiman, 1998; Kreuz & Roberts, 1995; Landgraf, 2014). According to Scharrer and Christmann (2011), ironic speech is characterized by higher energy, increased vowel duration, and vowel hyperarticulation (i.e., enhancement of the acoustic difference between vowels) compared to nonironic speech, but not by prominent F0-contour modulations as examined with quantitative eigenshape analysis. Even though these studies do not directly address the role and structural nature of prosodic stress, they strongly suggest that prosodic stress might well be a key component for the perception of irony.

The present study is the first to empirically investigate, with ecologically plausible material, the role of prosodic stress in irony perception. To this end, we performed four experiments (Figure 1). In the first experiment, we created a large set of target sentences that can be interpreted both ironically and nonironically, depending on narrative scenarios that precede the target sentences. Based on an online survey, we selected 14 sentences that showed the highest discriminability in terms of irony ratings (low vs. high irony ratings for nonironic and ironic sentences with identical wording) along with relatively convergent levels of likelihood of being used with an ironic and a nonironic meaning in the given context (Experiment 1). In a second experiment, we recorded the 14 sentences as spoken by 14 speakers in two different conditions (ironic vs. nonironic, as induced by the context and correctly identified by the speakers), and we documented various features (i.e., speech rate and pitch range) of the spoken sentences. In a third experiment, 20 listeners performed an annotation task that was designed to quantify the *p*-scores (Cole & Shattuck-Hufnagel, 2016)

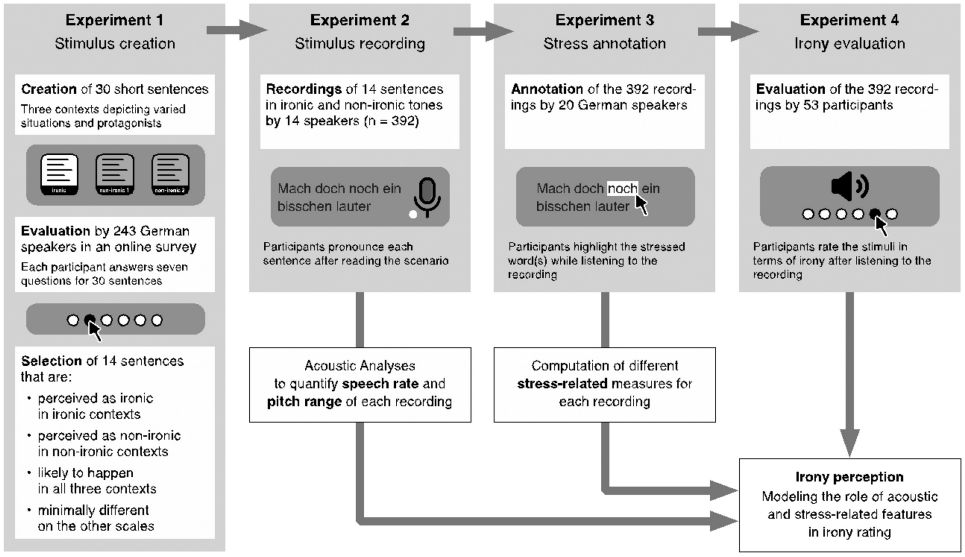
Table 1
Overview of Cues Examined for Comparing Ironic (or Sarcastic) and Neutral Speech Material in a Selection of Studies

Author(s)	Year	Language	Type	Speech material	Rec.	F0 mean	F0 range	Duration	Speech rate	Intensity mean	Intensity range	Voice quality
Rockwell	2000	English	Sarcasm	Sentence	Prof.	–	/	/	–	+	/	/
Anolli et al.	2000 2002	Italian	Sarcasm and kind irony	Sentence	NonP.	+	+	+/–	–	+	+	/
Bryant and Fox Tree	2005	English	Dripping and dry sarcasm	Sentence	Radio	+	n.s	n.s.	/	/	/	/
Cheang and Pell	2008	English	Sarcasm (humor)	Sentence, key phrase	NonP.	–	–	/	–	n.s.	n.s.	–
Cheang and Pell	2009	Cantonese	Sarcasm (humor)	Sentence, key phrase	NonP.	+	+/–	/	–/n.s.	+/n.s.	–	+
Scharrer and Christmann	2011	German	Ironic criticism	Target words	NonP.	–	/	+	/	+	/	/
Loevenbruck et al.	2013	French	Sarcasm	Sentence	NonP.	+	+	+	/	/	/	/
Niebuhr	2014	German	Sarcasm	Sentence	NonP.	–	–	+	–	–	n.s.	n.s.

Note. Different languages and materials were involved. +: higher in irony; –: lower in irony; /: not measured; n.s.: not significant; Rec.: recordings; Prof.: professional encoders; NonP.: nonprofessional encoders; Radio: natural material. Voice quality was measured as the harmonics-to-noise ratio (Cheang & Pell, 2008, 2009) or as the amplitude difference between H1 and H2 averaged across all vowels (Niebuhr, 2014). Only cues that were found significant in at least two of the selected studies are listed in the table. As a consequence, not every measured cue appears in the table. Cues referred only in few publications include nasality, speech reduction, and variability of intensity or F0 (e.g., Anolli et al., 2002; Bryant & Fox Tree, 2005; Cheang & Pell, 2008; Niebuhr, 2014).

Figure 1

The Study's Four Experiments: Stimulus Creation, Stimulus Recording, Stress Annotation, and Irony Evaluation



of the words in the spoken sentences and thus to estimate the perception of stress for each word of each speech sample. Finally, a fourth experiment examined the role of several predictors (characteristics of the speakers, of the listeners, and of the material; common features; the prosodic stress measure and its derivatives) for 53 listeners' perceptions of irony, using statistical modeling.

Experiment 1: Creation and Selection of the Material

Experiment 1 was designed to construct the material needed to examine the role of prosodic stress in irony perception (Figure 1). We first created a set of 30 target sentences that did not correspond to idiomatic ironic sentences (also called familiar irony, e.g., Giora & Fein, 1999; see Appendix B for the examination of the material idiomaticity). We also created 90 associated scenarios or contexts (three per sentence): One scenario was intended to lead to an ironic perception of the sentence, one was intended to lead to a nonironic perception of the sentence, and one scenario was an alternative one. An online survey was performed to identify a subset of 14 context–sentence pairs that is large enough to secure a reasonable variability in the linguistic material while still allowing for a within-subject design for the listening experiments.

The selection itself was grounded on several criteria. First, sentences of the subset should show high discriminability in terms of irony when presented in a given context (despite the identical wording); that is, they should be correctly identified as ironic versus nonironic in the two scenarios by a majority of the participants. Second, sentences should show a high level of likelihood of being used with an ironic and a nonironic meaning. Concretely, the ironic and nonironic meaning of each target sentence should be (roughly) equally likely to occur in two contrasting scenarios. Our decision to examine the likelihood or "fit" of our target expressions in their two contexts (ironic vs. nonironic) was meant to address concerns raised by Toplak and Katz (2000) regarding the ecological validity of constructed examples of irony as opposed to spontaneous occurrences of irony. By selecting only context–sentence pairs that received higher likelihood ratings, we aimed at broader generalizability of our results. In addition, we also suspected that the sentences that were considered to be likely in a given context would be easier to produce for the participants who would read these sentences aloud in our subsequent recording study (Experiment 2).

The main objective of the survey was thus to select items that were highly discriminable with regard to irony and likely to be used. In addition, we aimed at identifying pairs that were not too contrasting in other respects, such as the degree of humor, teasing, critique, or valence of speaker feelings. Indeed, the standard case of irony presented in the literature is a positive sentence spoken to express a negative meaning (Clark & Gerrig, 1984; Kreuz & Glucksberg, 1989; Kreuz & Link, 2002; Kumon-Nakamura et al., 1995; Sperber & Wilson, 1981), and it is generally assumed that negative feelings play an important role in irony (see Appendix A; Dynel, 2018; Garmendia, 2010; Sperber & Wilson, 1986). A negativity bias for the ironic sentence condition could lead to a systematic difference in terms of valence for the two experimental conditions (ironic vs. nonironic). Also, *humor*, *teasing*, and *criticism* are considered to be important pragmatic functions of irony. In Roberts and Kreuz's (1994) study, *to be humorous* was listed as a discursive goal for the rhetorical figure of irony by 65% of the participants. Moreover, lighthearted forms of irony support the pragmatic function of friendly teasing (e.g., Gibbs, 2000; Partington, 2007). To prevent large differences in terms of these potential characteristics (i.e., humor, teasing, critique, and positive and negative speaker feelings) between ironic and nonironic context–sentence pairs of the final subset, the online survey included specific scales capturing valence as well as critical/humorous/teasing aspects of the material.

Notably, the presence of the alternative context–sentence pairs in the complete material, in addition to the pairs designed to be specifically ironic or not, had the advantage of increasing the variability of the stimuli set to be evaluated in the online survey and to provide an

alternative option in case some of the context–sentence pairs would outperform the ironic and nonironic ones with regard to the mentioned selection criteria.

Method

The experimental procedure was approved by the Ethics Council of the Max Planck Society.

Participants

A total sample of 243 native German speakers completed the online survey. The data obtained from seven participants were excluded due to incorrect answers for quality check items (see the Procedure section), leaving a final sample of 236 participants (169 self-reported as females, 67 self-reported as males). The participants' ages ranged from 18 to 81 years old ($M = 38.3$, $SD = 17.1$). Participants were recruited via the research participant database at the Max Planck Institute for Empirical Aesthetics, via flyers and posters distributed at art centers and Goethe University in Frankfurt and via online platforms. The inclusion criteria were a minimum age of 18 years and having German as a native language. Responses to items from the Sarcasm Self-Report Scale developed by Ivanko et al. (2004; see the "Procedure" section for this experiment) showed that the participants used or believed they used irony rather frequently ($M = 3.6$ on the 6-point scale, $SD = 1.1$) and with different intentions ($M_{\text{critical}} = 3.0$, $SD = 1.1$; $M_{\text{humorous}} = 3.7$, $SD = 1.0$; $M_{\text{teasing}} = 3.3$, $SD = 1.2$). In addition, they self-reported a relatively high ability to understand irony ($M = 3.5$, $SD = 1.0$) and to generally appreciate the irony of others ($M = 3.6$, $SD = 1.2$). Participants received no compensation for taking part in the online survey; however, after completing the experiment, they had the opportunity to participate in a lottery for one of 40 Amazon vouchers worth €15.

Material

Target Sentences. We created 30 target sentences that can be interpreted either as ironic or nonironic, depending on the narrative scenarios or contexts that precede the target sentences (see below and Appendix B for details about contexts). In addition, all target sentences were constructed to meet the following criteria:

- They are relatively short (6–11 syllables) and thus can be easily spoken without pausing to take a breath;
- They do not contain any low-frequency words (i.e., the words had a minimum frequency of 2 on the scale used in the *Digitales Wörterbuch der Deutschen Sprache* [DWDS], a German-language online dictionary that provides word profiles based on large corpora, with frequency scores ranging from 0 = *rare* to 6 = *frequent*);
- They feature different syntactic structures, including expressions that do not include a verb (to potentially allow for a higher generalizability of the findings);
- They refer to a variety of aspects of the scenarios, such as other people's behavior (the addressee's or a third person's), the situation, the content, or the speakers themselves.

To ensure the low idiomaticity of the material, we presented all written target sentences standing alone (i.e., without a scenario that primes either a literal or an ironic meaning) to 35 German native speakers and examined their rating of irony on a 6-point scale ranging from not at all ironic to very ironic. The results, reported extensively in Appendix B, are reassuring because the mean irony rating (across participants) of the 30 target sentences was highly variable (ranging from 0.51 to 3.77, $M = 2$, $SD = 0.92$) and skewed toward the nonironic end of the scale. This control experiment confirmed that the created target sentences were not idiomatic. Therefore, all of the 30 target sentences were used in Experiment 1.

Contexts. For each of the 30 sentences, three contexts were created, resulting in a total of 90 different scenario/sentence items grouped in three conditions. Two conditions, called *ironic* and *nonironic*, differed in their specific intention, that is, they primed an ironic or nonironic meaning of the target sentence. An additional context was created, referred to as *alternative*, to increase the variability of the material to be evaluated and to provide another option if the results of the online survey show that they fit the selection criteria better than the nonironic contexts.

Regardless of the condition, each scenario begins by framing a communicative exchange between two individuals, and the target sentence is then uttered by one of the interlocutors.

The construction of the scenarios was guided by the following criteria:

- They should evoke a compact everyday situation that can immediately be intuitively understood and hence needs no additional explanation;
- The interlocutors featured in the scenarios are familiar with one another, for example, they are friends, family, or roommates (on the influence of the type of relationship between interlocutors on irony perception, see Pexman & Zvaigzne, 2004);
- Fifteen of the 30 sentences presented to each participant should be uttered by female speakers and the other 15 by male speakers, thus balancing speaker gender. For each sentence, the speaker gender was kept constant in the three conditions.

Procedure

The rating task was implemented as an internet survey using Unipark Enterprise Feedback Suite (QuestBack GmbH, Cologne, Germany). The 90 context–sentence pairs were divided into three sets of 30 items in which each of the 30 sentences appears (i.e., no sentence would be repeated for a participant) but only in one of the three context conditions (*ironic*, *nonironic*, and *alternative*). The three conditions were equally represented in each of the three sets (i.e., each set contained 10 items for each condition) that each participant could rate items from the three different conditions while avoiding sentence repetition and keeping the task short. After confirming that the participants met the inclusion criteria, they were randomly assigned to one of the three sets of stimuli. The participants' age, gender, education, and time to complete the survey did not differ significantly between the three groups (all $ps > .25$).

Each of the 30 target sentences was presented following its given scenario, one at a time, and the participants were asked to evaluate the sentences on seven rating items (presented in randomized order):

- How ironic is the utterance meant to be?
- How likely is such a statement in this situation?

- To what extent does the utterance convey speaker's positive feelings?
- To what extent does the utterance convey speaker's negative feelings?
- How humorous is the utterance meant to be?
- How teasing is the utterance meant to be?
- How critical is the utterance meant to be?

Answers were provided by clicking on a specific location on a 6-point Likert-type scale ranging from 0 (*not at all*) to 5 (*very strongly*). To ensure that participants read the questions, we included a "quality check" consisting of three additional questions (i.e., participants were instructed to select a specific tick of the scale) over the course of the experiment. At the end of the survey, participants responded to seven questions (in German) based on Ivanko et al.'s (2004) Sarcasm Self-Report Scale to document how frequently they used (or believed they used) irony, how frequently they used irony with a critical/humorous/teasing intention, their estimate of their ability to understand irony and to appreciate the use of irony by others, on 6-point scales, ranging from 0 (*not at all/never*) to 5 (*very much/very often/always*). The entire survey took on average about 34 min ($SD = 19$) to complete.

Statistical Analyses

Linear mixed-effects regression models were used to test for differences in ratings between the conditions (*ironic*, *nonironic*, and *alternative*). We applied intercept-only linear mixed-effects regression models to estimate between- and within-cluster variance components, and also the intraclass correlation coefficient (ICC) based on the ratio of between-cluster to total (between- and within-cluster) variance. The models included crossed random intercepts for raters and sentences to account for the nonindependence of observations within clusters.

With the goal of selecting a subset of context–sentence pairs that maximally discriminate between perceptions of irony in ironic versus nonironic contexts while maintaining high ecological validity and keeping other potentially relevant factors as constant as possible, we evaluated sentences as potential target sentences based on the following criteria: (a) The proportion of irony ratings of 4 or 5 was at least 75% for contexts that were intended to be ironic; (b) The proportion of irony ratings of 0 or 1 was at least 75% for contexts intended to be nonironic; (c) The proportion of likelihood ratings of 4 or 5 was at least 50% for all contexts; (d) There was minimal dissimilarity between a pair of contexts (i.e., ironic vs. nonironic) for the sentence on the items assessing perceived positive and negative speaker feelings, as well as humor, teasing, and criticism. Regarding (d), we computed between-context mean differences for each of these items across participants and then squared the mean difference to penalize larger deviations. We then summed the aggregated squared mean differences over items to achieve an integrated indicator of dissimilarity between contexts.

Results and Discussion

The irony ratings were rather independent of individual raters (ICC = 0.02, 95% CI [0.012, 0.031]) and sentences (ICC = 0.02, [0.011, 0.036]), suggesting that most of the variance was observed within clusters. The irony ratings differed significantly between conditions. The mean rating for the *ironic* condition was 4.53 ($SE = 0.07$, $p < .001$), while the mean rating for the *nonironic* condition was 0.8 and thus 3.73 scale units lower ($SE = 0.04$, $p < .001$) than for the *ironic* condition. The mean rating for the *alternative* condition was 1.72 and thus 2.81 ($SE = 0.04$, $p < .001$) units lower than for the *ironic* condition.

The irony ratings for the target sentences were compared to the ratings provided in the control experiment (Appendix B). Ratings were higher in the *ironic* condition and lower in the *nonironic* condition than in the context-free condition (both $p < .001$). These opposite effects confirm the antithetical nature of the two context conditions and the relatively neutral nature of the target sentences themselves.

Figure 2 shows the relative frequency of the six irony ratings for each sentence in the three context conditions (*ironic*, *nonironic*, and *alternative*). The results indicate that the contexts are very effective in altering the perception of each sentence along the scale ranging from *not at all* to *very ironic*.

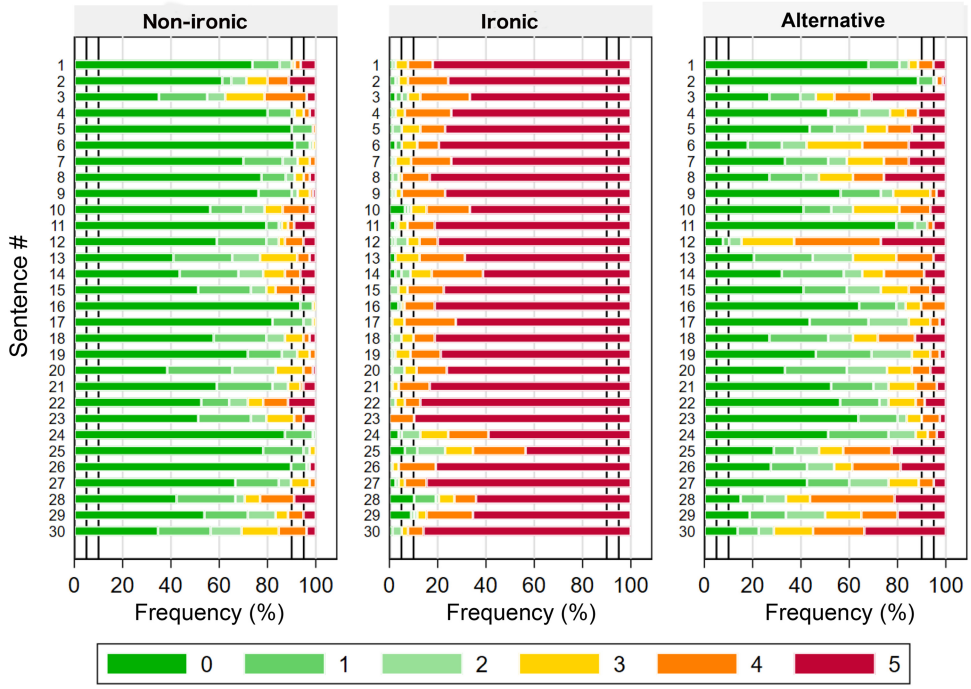
Applying the same models with the likelihood ratings as the outcome variable resulted in larger ICCs for raters (ICC = 0.24, 95% CI [0.207, 0.281]) and sentences (ICC = 0.09, [0.057, 0.147]), indicating larger between-rater differences in the likelihood ratings compared to the irony ratings. The estimated average ratings of likelihood for the *irony* condition were 3.48 ($SE = 0.09$, $p < .001$). The ratings for the *alternative* condition were not significantly different (difference parameter = −0.02, $SE = 0.03$, $p = .61$), whereas those for the *nonironic* condition were slightly higher (difference parameter = 0.31, $SE = 0.03$, $p < .001$).

In contrast to the contextualized ratings, the irony as well as likelihood ratings for the sentences presented without a context (see Appendix B for the details of the method) showed a greater dependence on individual raters (ICC = 0.11, 95% CI [0.064, 0.176]) as well as the sentence (ICC = 0.20, [0.123, 0.301]). Adding the additional ratings of sentences without a context to a model with a random intercept for raters resulted in an average irony rating of 1.95 ($SE = 0.08$, $p < .001$), while the average irony ratings in the *ironic* context were significantly higher (difference parameter = 2.59, $SE = 0.09$, $p < .001$) and the irony ratings in the *nonironic* and *alternative* contexts were significantly smaller (difference for *nonironic* = −1.14, $SE = 0.09$, $p < .001$; difference for *alternative* = −0.22, $SE = 0.09$, $p = .016$). Unlike the results for the irony ratings, the dependence of the perception of a sentence's likelihood on individual raters (ICC = 0.25, [0.170, 0.350]) and sentences (ICC = 0.14, [0.083, 0.227]) was similar to that for the contextualized ratings.

In sum, 17 of the 30 sentences met the combined criteria of having (a) a large majority (75%) of high irony ratings in the *ironic* condition as well as low irony ratings in the *nonironic* condition, and (b) a majority (50%) of high likelihood ratings in both conditions. To select the final subset of these sentences to be used in Experiments 2–4, we examined the sum of the squared mean differences (SSMDs) for the context–sentence combinations. Figures 3A, 3B shows the SSMDs for the 17 sentences. Five sentences that were presented with an *alternative* context had lower SSMD scores; we therefore selected the *alternative* context instead of the initial *nonironic* context for these sentences. Sentences 6, 8, and 9 were dropped due to their high SSMD scores. The SSMDs for the final selected set of 14 sentences ranged from 3.6 to 39.4, with an average $M = 20.1$ ($SD = 11.6$). Figure 3C shows the mean differences for the final selection of context–sentence combinations for each of the assessed items. While these results still reflect some heterogeneity in the ratings of perceived positive

Figure 2

Frequency (%) of Irony Ratings by Context Condition for the 30 Target Sentences



Note. 0 = *not at all*; 5 = *very ironic*. Each target sentence was rated by (different) raters in the context of a scenario that was designed to trigger an ironic or nonironic understanding of the sentence. See the online article for the color version of this figure.

and negative speaker feelings, humor, and criticism, the most heterogenous context–sentence pairs were excluded. The final selected set of sentences and contexts can be found in Appendix B, together with English translations.

Within-rater and within-sentence associations for this final selection of 14 sentences dropped to near zero (both ICCs < 0.001), suggesting that the ratings were almost exclusively influenced by the combination of a sentence with a specific context. The same type of mixed-effects regression model as above showed that the average differences in irony ratings between conditions were even stronger for the final subset of sentences, with average irony ratings in *ironic* contexts estimated as 4.63 ($SE = 0.04$, $p < .001$) and considerably lower irony ratings for the *nonironic* and *alternative* conditions (*nonironic*: difference $= -4.14$, $SE = 0.05$, $p < .001$; *alternative*: -3.81, $SE = 0.07$, $p < .001$).

These results confirm that the different contexts can reliably prime ironic versus nonironic understandings of identical sentences and, moreover, that the perceived likelihood of using these sentences with an ironic or nonironic meaning does not significantly differ. The rather high levels of perceived likelihood, along with the similarity of likelihood ratings across context conditions, support the ecological validity of our materials and strongly reduce the risk of a systematic bias between the ironic and nonironic conditions. This potential bias was further controlled by selecting a subset of 14 context–sentence combinations using the criteria of maximizing differences of irony perception and minimizing differences of perceived critical, humorous, and teasing intentions as well as perceived positive or negative feelings of the speaker.

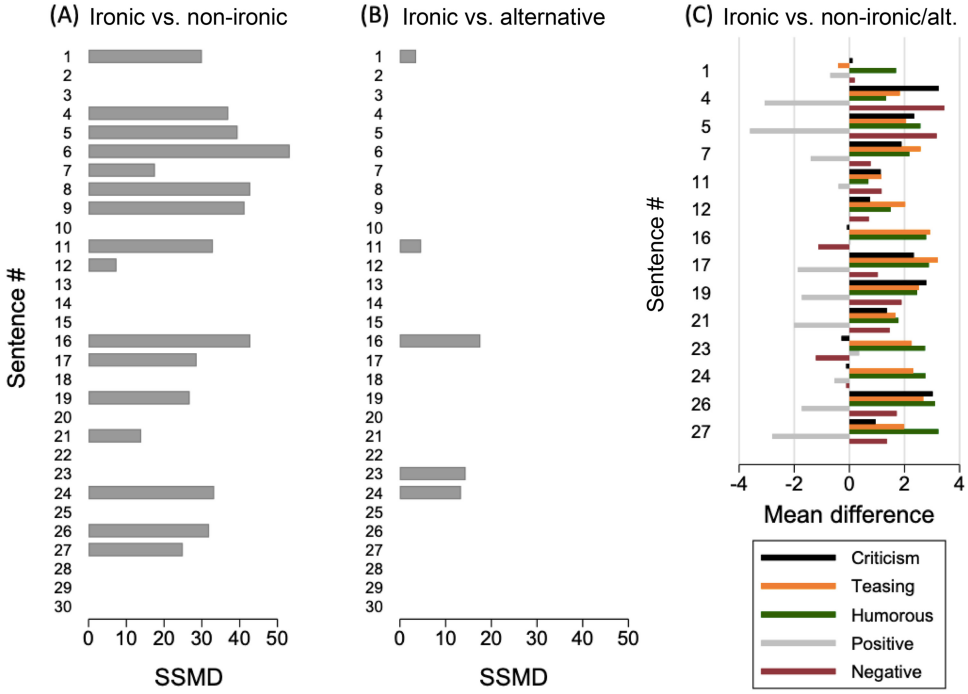
In sum, we identified 14 sentences that show strongly contrasting meanings dependent on contexts that prime either ironic or nonironic understandings while not differing in their likelihood of occurrence in these contexts. We conjectured that the equally high likelihood of our target sentences in the two conditions would facilitate the ease with which nonprofessionals could render equally "natural" performances of these sentences and thus enhance the authenticity of the speech material that we would use in Experiments 2–4 (see Figure 1). We expected that the vocal renditions of these 14 sentences would show acoustic differences dependent on the context-driven understanding of whether they were intended to be understood literally or ironically, thus making them suitable for further experiments. The selected sentences, such as "Was für ein schickes Auto" [What a fancy car] or "Das steht dir richtig gut" [This looks really good on you] are reported in Table B1, together with their contexts and English translations.

Experiment 2: Recording and Acoustic Analysis of the Material

The purpose of this step was to create auditory speech material to be acoustically analyzed, which would then be annotated (Experiment 3), and finally evaluated with regard to irony (Experiment 4). As described in Experiment 1, the ironic and nonironic target sentences were preceded by brief scenarios. We expected that reading these scenarios before the target sentences would not only prime the different meanings for the two conditions, but would also facilitate natural production of the spoken material. In

Figure 3

Output of the Third Selection Criterion Between Ironic and NonIronic Conditions, Between Ironic and Alternative Conditions, and Description of the Final Set of Sentences



Note. Panel A: sum of squared mean differences (SSMD) across items assessing perceived positive and negative speaker feelings, as well as humor and criticism, between *ironic* and *nonironic* contexts for the sentences meeting our selection criteria. Panel B: SSMD across items assessing perceived positive and negative speaker feelings, as well as humor and criticism, between *ironic* and *alternative* contexts for the sentences meeting our selection criteria. Panel C: mean differences across raters on items assessing criticism, teasing, humor, and perceived positive and negative speaker feelings for the final selection of 14 sentences. See the online article for the color version of this figure.

line with previous research (Anolli et al., 2000, 2002; Lœvenbruck et al., 2013; Scharrer & Christmann, 2011), we chose an induction procedure instead of an explicit request. That is, we did not ask participants to speak the sentences in a nonironic or an ironic tone of voice, as this might have resulted in an exaggerated and unnaturally dichotomous production strategy that could mask more subtle nuances. Instead, participants were asked to read a scenario and to speak the target sentence in a way that seemed appropriate given this scenario. Therefore, we left it to the participants themselves to decide how the target sentences should be appropriately spoken in the two conditions. By selecting statements that obtained a high likelihood to occur in the proposed situations (see selection criteria in Experiment 1), we assumed that the speakers' task would not be too challenging despite the fact that they were not professional actors. To control for individual interpretations of the items by the speakers, we presented the sentences along with their respective scenarios a second time after the recording session and asked for a binary classification of the sentences in their respective scenario contexts as being either ironic or nonironic.

As there is evidence that the use and perception of irony differ depending on the speaker's gender (Colston & Lee, 2004; Jorgensen, 1996; Leykum, 2019; Pexman & Olineck, 2002), we recruited a gender-balanced sample for the recordings. In addition, it has been shown that sarcasm comprehension is modulated by age (Phillips et al., 2015), suggesting that the potential generalizability of findings would require a large age range of speakers (and listeners). Therefore, we invited both female and male participants with a substantial age range. Finally, we aimed to control for the vocal quality of the speakers, to avoid potential biases in perceptions regarding the speaker due to voice disorders that might influence the perception of irony in the subsequent experiments (e.g., an attractiveness bias for smoothed voices, Bruckert et al., 2010; for the perception of pathological voices, see Murry et al., 1987).

Method

The experimental procedure was ethically approved by the Ethics Council of the Max Planck Society, and the recording and classification tasks were undertaken with the written informed consent of each participant.

Participants

Twenty-three participants were recruited via the research participant database of the Max Planck Institute for Empirical Aesthetics. In the recruitment, we informed participants that their performances would be used to investigate the acoustics of speaker intentions. Inclusion criteria were a minimum age of 18 years and

having German as a native language. To control for the vocal quality of the speech material, three experienced speech-language pathologists were asked to independently rate the general Grade as well as Roughness, Asthenicity, Breathiness, and Strain of the voice (GRABS, Hirano, 1981) of two sentences for each speaker ($n = 44$ sentences in total; there was a technical issue during the recording of one participant), presented in random order. Although this procedure is not equivalent to a full-blown medical examination of the speakers' vocal folds, we chose a very conservative criterion that strongly limited the risk of vocal perturbation or nondiagnosed vocal pathology being present in the final sample. The three voice pathologists were asked to perform the task with adequate equipment and were not informed about the purpose of the study. Speakers were discarded from the final sample if at least one of the professionals gave a rating higher than zero (on a scale ranging from 0 = *normal* to 3 = *pathological*) to either of the speakers' two sentences, which resulted in eight speakers being excluded.

The final sample thus comprised 14 participants (seven self-reported as females, seven self-reported as males; aged between 21 and 77 years, $M = 40.36$, $SD = 19.10$). A questionnaire based on Ivanko et al.'s (2004) Sarcasm Self-Report Scale regarding participants' typical usage and appreciation of irony (same as in Experiment 1, with a 6-point scale ranging from 0 to 5) revealed that the participants' frequency of using (or believing that they used) irony was in the high medium range ($M = 3.6$, $SD = 0.9$). Moreover, the participants self-reported as using irony with different intentions ($M_{critical} = 2.64$, $SD = 1.39$; $M_{humorous} = 3.79$, $SD = 0.70$; $M_{teasing} = 3.7$, $SD = 0.8$), rated their ability to understand irony as fairly high ($M = 4.1$, $SD = 0.6$), and generally appreciated the irony used by others ($M = 3.7$, $SD = 1.0$).

Material

The material to be recorded consisted of the 14 target sentences selected in Experiment 1. The literal (nonironic) versus ironic conditions for the 14 sentences were primed using the 28 scenarios described in Experiment 1 and reported in Appendix B. Note that five of the scenarios used for the nonironic condition originated from the *alternative* contexts as they outperformed the nonironic ones with regard to our selection criteria (see Table B1 for details).

Procedure

Recordings. The participants' readings of the target sentences were recorded in a soundproof room, using a Neumann U87 microphone and an RME Fireface UCX sound card. Before the recording began, the microphone gain was adapted to the speaker, using the program TotalMix FX. Material was displayed and recorded via MATLAB (2017), using the Psychophysics Toolbox (Kleiner et al., 2007). Participants used a response box with two buttons: one for starting and one for ending a recording. They were allowed to repeat each sentence until they were satisfied with their reading. In the cases where participants made use of this opportunity, only the last recording of a sentence was kept for the acoustic analyses and Experiments 3 and 4. Note that, whereas participants were allowed to repeat a sentence if necessary, they were not explicitly asked to do so to avoid potential training effects that might favor the generation of stereotypical productions.

Before the presentation of the context–sentence combinations began, participants were asked to read aloud the 14 target sentences (without the scenario descriptions) in a neutral tone of voice. This step allowed us to calibrate the recording settings and provided the participants with an opportunity to warm up their voices and familiarize themselves with the microphone (psychological stress being a potential source of voice signal perturbation, see Dropuljić et al., 2017; Larrouy-Maestri & Morsomme, 2014) and also served as a training round for pronouncing the sentences properly.

For the actual recordings, participants were informed that they would be reading short descriptions of social interactions that all ended with a sentence uttered by one of the interlocutors. They were instructed to read each scenario description silently and then to speak the target sentence in a way that seemed appropriate given the preceding scenario. The 28 items were presented in pseudo-randomized order. The only constraints for the randomization were: the identical target sentences in their two conditions had to be separated by at least two items, at most three sentences in a row could belong to the same condition (i.e., ironic or nonironic), and the presentations were balanced with regard to which condition was presented first (i.e., seven times nonironic condition first, seven times ironic condition first).

At the end of the recording sessions, the collected speech material was normalized using R128Gain (Version 1.0.11, Belkner, 2014) to adjust the overall volume level across stimuli while preserving the variability and dynamics within sentences.

Classification Task. Subsequent to the recording, participants were presented with the context–sentence combinations once again. This time they were asked whether the sentence was ironic, based on their interpretation. Participants answered by pressing a "yes" or "no" button in the response box. This task allowed us to compute each participant's *d* prime (d'). Each answer was assigned to one of four categories: hit (if the context–sentence combination was ironic and was recognized as ironic), miss (if the context–sentence combination was ironic, but was not recognized as ironic), correct rejection (if the context–sentence combination was not ironic and was perceived as not ironic), and false alarm (if the context–sentence combination was not ironic, but was perceived as ironic).

We computed two measures: the *hit rate* (H, the proportion of *ironic* trials that the participant perceived as ironic) and the *false alarm rate* (F, the proportion of *nonironic* trials that the participant perceived as ironic). We then computed the inverse of the cumulative normal distribution function (with the norminv function, MATLAB, 2017) for these measures, so that the normalized values ranged from −1.8 to 1.8. The optimal performance would be a maximal hit rate and a minimal false alarm rate, meaning the participant recognized all cases of irony and rejected all cases of nonirony. The larger the difference between the normalized hit rate and the false alarm rate, the better the participant's ability to classify the two conditions. The d' measure captures this sensitivity, with 3.6 reflecting the highest and −3.6 the lowest ability to discriminate the two experimental conditions of our target sentences.

The session ended with a short questionnaire on demographic data and the participant's personal use of irony (the latter from Ivanko et al., 2004; see the Experiment 1 "Procedure" section for a description and this experiment's "Participants" section for the results). The entire session lasted about 1 hr, and the participants received €10 for their participation.

Acoustic Analyses

The acoustic analyses of the speech material focused on two features that several previous studies have shown to be relevant for the acoustics of irony: speech rate and pitch range (e.g., Anolli et al., 2000, 2002; Cheang & Pell, 2008, 2009; Lœvenbruck et al., 2013; Niebuhr, 2014; Rockwell, 2000). To calculate the speech rate, the number of syllables in the sentence (between 6 and 11) was divided by the duration of the spoken sentence as measured in seconds. The pitch range or pitch interval size between the minimum (f_1) and maximum (f_2) frequency for each sentence was computed using the formula $1200 \times 3.322 \log_{10} (f_2/f_1)$ and reported in cents (100 cents = one semitone).

Because estimation of complex material (i.e., connected speech) has been shown to be particularly challenging and prone to detection errors when performed automatically (e.g., Hirst, 2011), the minimum and maximum frequencies for each stimulus were determined in two steps using Praat (Version 6.0.39, Boersma & Weenink, 2018). First, automatic pitch detection was performed, with the frequency range adapted for the female and male speakers (60–500 Hz and 50–400 Hz, respectively) and other parameters set as the default. Second, the detection was manually corrected following predefined criteria: pitch points considered to be pitch tracking errors (unvoiced consonants, octave jumps, points at a distance of more than 2 *SD* from the average of the minimum and maximum for each sentence/speaker) and those believed to represent noise, glottal, or breathing modes of phonation were deleted. This time-consuming procedure allowed us to identify realistic minimum and maximum pitch points (and thus the pitch range) for each utterance.

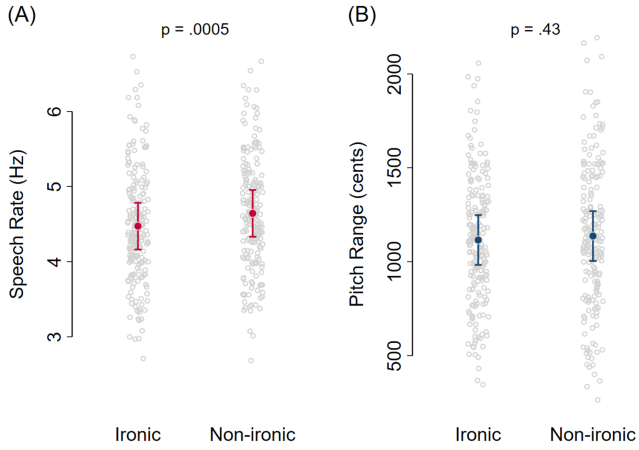
Results and Discussion

The classification task confirmed that the 14 participants in the final sample were able to discriminate the ironic and nonironic conditions as expected: their d' ranged from 2.9 to 3.6 ($M = 3.4$, $SD = 0.2$). Notably, we also observed a high discrimination ability in the participants in the original sample who were discarded from the analyses due to poor (or suspected to be poor) voice quality following the evaluation by the three independent speech-language pathologists. This high overall discrimination of our target sentences in their two conditions confirms the existence of an ability to detect irony in written material (Kunneman et al., 2015; Liebrecht et al., 2013) and also that the context scenarios and sentences selected from the online survey (Experiment 1) were perceived as intended. It is important to note that because responses to the classification task were collected after the recording session, the speakers were already familiar with the context–sentences; in the absence of a previous encounter with the material, the results might not have been as strong. Also, this result of the classification task can not confirm that speakers purposefully produced the sentences as intended (i.e., with an ironic and nonironic tone, without explicitly asking them to do so). Still, it does indicate that the speakers successfully discriminated between the ironic and nonironic contexts of the target sentences.

As illustrated in Figure 4, the speech rate was slowest when the sentence was pronounced ironically, which is in line with Rockwell (2000), Anolli et al. (2000), Cheang and Pell (2008), and Niebuhr (2014). Linear mixed-effects regression models with crossed random intercepts for speaker and sentence showed a significant effect of the condition on the speech rate (nonironic vs. ironic = 0.17, $SE = 0.05$, $p < .001$), but not on the pitch range (nonironic vs. ironic = 21.1, $SE = 26.5$, $p = .43$). Regarding the pitch range, previous studies have reported contradictory findings: while some studies found a larger pitch range in the ironic condition compared to the nonironic condition (e.g., Anolli et al., 2000; Lœvenbruck et al., 2013), other studies reported the opposite (e.g., Cheang & Pell, 2008; Niebuhr, 2014), or that the effect did not reach a significant level (Bryant & Fox Tree, 2005). Our results are in line with the last finding. However, our material might not be comparable to the material that others have used to date, since we aimed at recording "natural" productions by nonactors. The use of an induction procedure (instead of an explicit instruction) as well as the lack of repetitions of the target sentences across the session hopefully prevented stereotypical productions but might have limited the differences in acoustic features such as pitch variations.

Figure 4

Estimated Marginal Means and Scatterplots of all Observations for Speech Rate and Pitch Range for the (Intended) Ironic and Nonironic Conditions



Note. Panel A: speech rate in Hz. Panel B: pitch range in cents. Error bars indicate 95% confidence intervals. The *p*-values are for condition difference coefficients from mixed-effects regression models. See the online article for the color version of this figure.

Overall, Experiment 2 confirmed that our target sentences in the two experimental conditions were correctly discriminated as ironic versus nonironic and that the ironic sentences were spoken more slowly than the nonironic ones. This supports the relevance of the temporal aspect of speech (syllable rate, in the present case) for expressing irony and provides highly controlled material for testing specific hypotheses regarding the role of acoustic or perceptual characteristics in the expression of irony. In sum, Experiment 2 provided us with a set of 392 speech samples that are balanced with regard to gender, are naturally performed, and have no recording quality issues or voice disorders, to be used in the next two steps: quantifying the perceived stress in the recorded speech material (Experiment 3) and identifying predictors of irony perception (Experiment 4).

Experiment 3: Quantification of Stress in Recorded Speech Material

To quantify prosodic stresses in the material recorded in Experiment 2, we followed the procedure proposed by Cole and

Shattuck-Hufnagel (2016) and examined native German speakers' perceptions of prominence in the recorded speech material.

Method

The experimental procedure was ethically approved by the Ethics Council of the Max Planck Society, and the task was undertaken with the written informed consent of each participant.

Participants

Twenty native German speakers (10 self-reported as females, 10 self-reported as males, aged between 21 and 70 years, $M = 44.2$, $SD = 17.8$) were recruited via the research participant database of the Max Planck Institute for Empirical Aesthetics. One participant was a native Swiss-German speaker, and another was bilingual in German and English. All participants declared that they had normal hearing ability and had not participated in the online survey (Experiment 1) or the recording session (Experiment 2).

Due to the lack of pilot or published studies with a similar design, we were not able to rely on effect sizes observed in previous work to estimate the number of participants needed through power analyses. However, Roy et al. (2017, Figure 6) computed Fleiss's kappa for their 32 participants to quantify the degree of intertranscriber agreement. They observed a low-to-moderate agreement level for prominence marking (Fleiss's kappa = 0.28; slightly higher kappas were reported for a similar task in Cole, Mo, & Hasegawa-Johnson, 2010). The authors also examined the number of transcribers that would be required to reach reasonable agreement. Specifically, they computed Fleiss's kappa for resampled groups of increasing size (from two to 31 transcribers). For each sample size, they bootstrapped 10,000 samples to assess how much sampling error occurred across the different values of n. The curve depicting the sampling error was steep between two and eight transcribers, reached a sampling error of 0.05, and stabilized after about 15 transcribers. That is, Cole and colleagues reported a plateau for the estimation of the sampling error after about 15 annotators for prominence annotation of short speech samples, which suggests that a group of fewer than 20 participants might provide reliable judgments in similar tasks.

Material

As reported for Experiment 2, the material consisted of 14 sentences that were each recorded in two context conditions (*ironic* vs. *nonironic*) by 14 native German speakers. In Experiment 3, the 392 trials were played through headphones (Beyerdynamic DT-770 Pro) with an in-house Presentation program (Version 20.0) and simultaneously displayed on a screen.

Procedure

The participants were tested individually in a sound-attenuated booth. They were asked to listen to and read single sentences and to select the words they perceived as being prominent in each sentence with a mouse click. The instruction stressed that their responses should be guided by their personal impressions, and it did not make any reference to irony. Participants were given further information about prominence based on the wording in Roy et al. (2017), who instructed their participants that "in normal speech, speakers pronounce some word or words in a sentence with more prominence than others. The prominent words are in a sense highlighted for the listener, and stand out from other nonprominent words. Your task is to mark words that you hear as prominent in this way" (p. 6).

Before beginning the study, participants completed two practice trials to familiarize themselves with the task. The task itself contained 392 trials presented in eight blocks of 49 trials each, separated by short breaks. After each break, participants could begin the next block as soon as they were ready. There was no time limit for selecting the prominent word(s) in a sentence. When participants indicated that they were done identifying the prominent words in a given sentence, the next sentence was played automatically. Each recording was played to the participants once. At the end of the session, participants were asked to complete the same questionnaire about their use of irony as in the preceding two experiments (for a description, see the "Procedure" section for Experiment 1). The sessions took roughly between 1 and 1.5 hr on average, and the participants received €15 for their participation.

Results and Discussion

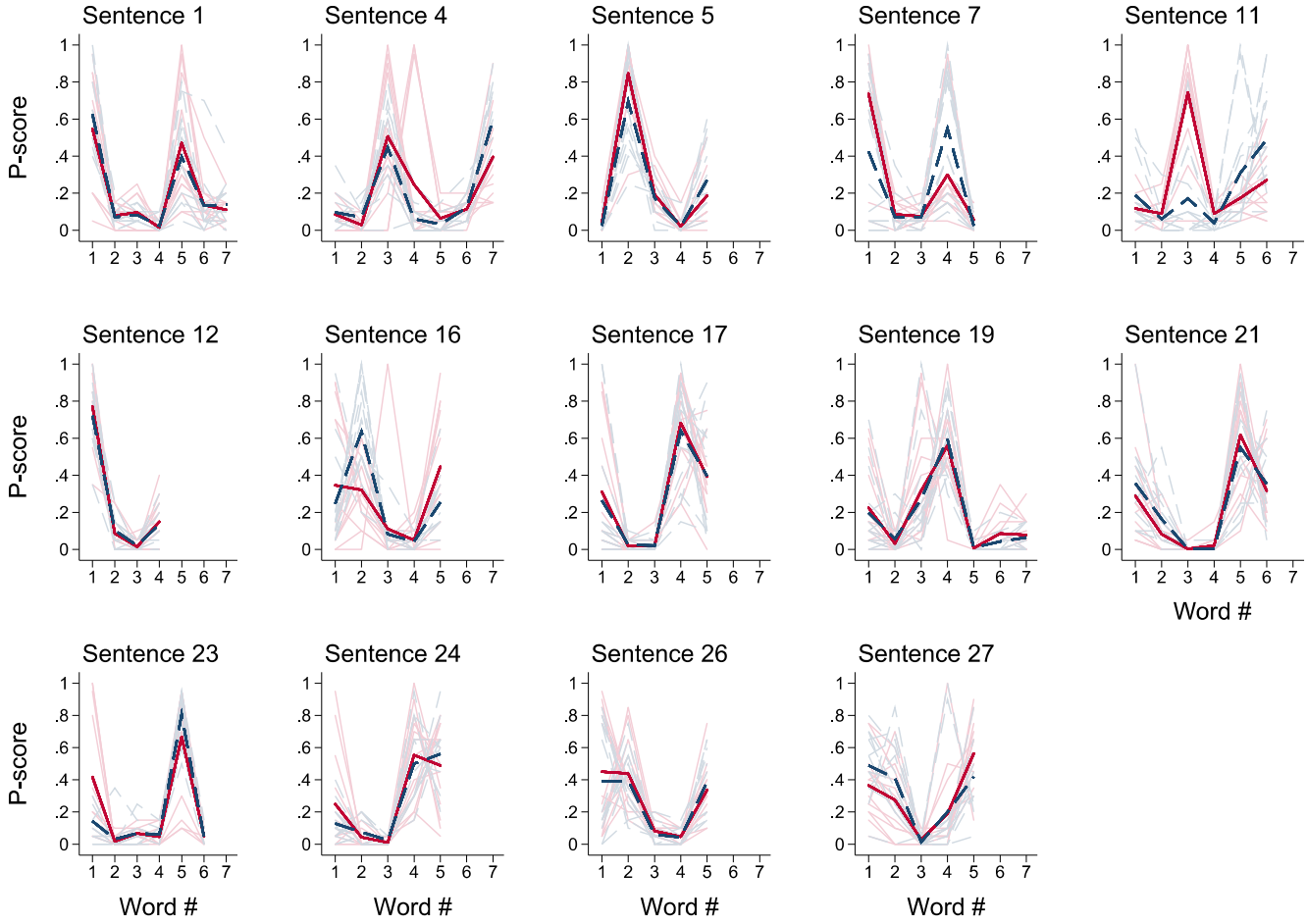
Across the sentences that were produced in the *ironic* condition in Experiment 2, participants on average rated 1.33 words per sentence as being stressed ($SD = 0.78$, Min = 0, Max = 7). Given the sentence lengths of 4–7 words, this corresponds to 24.4% of the words in a sentence being perceived as stressed on average ($SD = 14.2$, Min = 0, Max = 100). The result was similar for the sentences that were produced in the *nonironic* condition ($M = 1.29$, $SD = 0.76$, Min = 0, Max = 6), where on average, 23.5% of the words were perceived as stressed ($SD = 14.0$, Min = 0, Max = 100). The average agreement of raters on whether a word was stressed across sentences and speakers was estimated as $\kappa = 0.41$ ($z = 266.8$, $p < .0001$), which is a moderate agreement, according to Landis and Koch (1977). The interrater agreement was higher for ratings of the target sentences in the *ironic* condition ($\kappa = 0.43$, $z = 197.6$, $p < .0001$) compared to the *nonironic* condition ($\kappa = 0.39$, $z = 179.4$, $p < .0001$).

For each word within a sentence, we computed a p-score indicating the relative frequency of the word being perceived as stressed (i.e., the probability of the word being marked by raters as having been stressed by the speaker). Figure 5 illustrates the p-scores, which ranged from 0.004 to 0.846, for the 14 sentences produced by the 14 speakers in the *ironic* (red) versus the *nonironic* (blue) condition. The average p-score for the words in the *ironic* condition ($M = 0.24$, $SD = 0.22$, Min = 0.004, Max = 0.846) and the words in the *nonironic* condition ($M = 0.23$, $SD = 0.22$, Min = 0.036, Max = 0.807) did not differ significantly ($p > .05$). To test whether more words were perceived as being stressed in the *ironic* condition, a mixed-effects logistic regression model with fully crossed random intercepts of participant, speaker, and sentence was computed, modeling the probability of words being stressed by condition (*ironic* vs. *nonironic*). In the *ironic* condition, the probability of a word being stressed was significantly higher than in the *nonironic* condition; however, the odds ratio indicates that this difference was a rather small increase of 5% ($OR = 1.05$, 95% CI [1.004, 1.098], $p = .032$). Overall, the ratings were rather independent of the rater (ICC-rater = 0.043), sentence (ICC-sentence = 0.008), and speaker (ICC-speaker = 0.003).

For follow-up models testing differences between sentences and speakers as fixed-effect factor variables, as well as their interaction

Figure 5

p-*Scores by Word Position in All Sentences Presented to Raters*



Note. Solid lines show *p*-scores for the *ironic* condition; dashed lines show *p*-scores for the *nonironic* condition. Thick lines show the average *p*-score per word across raters. Thin lines show individual raters' *p*-scores. See the online article for the color version of this figure.

with context conditions, fully crossed random-effect models were not estimable. We therefore included only a random intercept for raters to account for the nonindependence of observations in these models. The model including speaker, condition, and their interaction as factor variables showed main effects for the condition ($p = .036$) and speaker ($p < .001$); however, their interaction was not statistically significant ($p = .67$). Sidak-adjusted multiple comparisons of each speaker's relative deviance from the grand mean of all speakers in a follow-up model without interaction showed that the recordings by one speaker were assigned significantly fewer stress markings compared to the other speakers, whereas more than the average number of stresses were perceived for the recordings of three speakers. For the remaining 10 speakers, stress markings did not differ from the grand mean (Figure 5).

We aggregated *p*-scores across speakers separately for the two conditions to test whether words at a specific position in a sentence were more or less likely to be stressed in the *ironic* than in the *nonironic* condition. Wilcoxon matched pairs signed-rank tests showed that the third word of a sentence (i.e., a fairly central position in the sentence) received higher *p*-scores in the *ironic* condition ($z = 2.34$, $p = .019$) and that this was not simply a result of the outlying observation for sentence 11 ($z = 2.12$, $p = .034$ with sentence 11 excluded from the analysis). In contrast, the *p*-scores of the first, second, and last words of a sentence were similar for the *ironic* and *nonironic* conditions ($ps > .29$).

Additionally, we examined *p*-scores as a function of the word classes. Note that this variable was not manipulated when creating the material, as our focus was on other characteristics (e.g., word frequency, length, type of interaction, gender of the characters). We aggregated *p*-scores across speakers separately for the two conditions to test whether words from different word classes differed between the conditions in terms of their probability of being perceived as stressed by a speaker. The sentences comprised a total of 12 adjectives (15%), five adverbs (6%), 10 articles (13%), nine nouns (12%), 14 particles (18%), three prepositions (4%), 13 pronouns (17%), and 12 verbs (15%). A mixed-effects logistic regression model including condition, word class, and their interaction as factor variables resulted in a significant main effect of the word class ($p < .001$) and its interaction with the condition ($p < .001$). Sidak-adjusted multiple between-condition comparisons of the predicted probabilities of words being perceived as stressed revealed

significant differences for three word classes (Figure 6). Adverbs ($OR = 1.30$, 95% CI [1.05; 1.61], $p = .006$) and particles ($OR = 1.79$, [1.48; 2.16], $p < .001$) were more likely to be perceived as stressed in ironic contexts, whereas nouns were less likely to be perceived as stressed in ironic contexts ($OR = 0.80$, [0.68; 0.94], $p = .002$). No significant differences were observed for any of the other word classes.

Finally, to describe the level and variability of *p*-scores within sentences, we computed the following indicators for each sentence (separately by speakers):

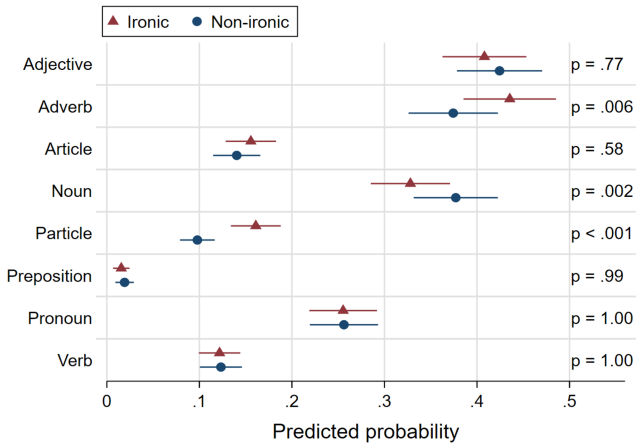
(1) the mean *p*-score across the words in the sentence;
(2) the percentage of *p*-score peaks in the sentence, defined as *p*-score > 75th percentile, that is, *p*-score > .35;
(3) the *p*-score mean-squared successive differences (MSSD), that is, the squared differences of the *p*-scores of successive words in the sentence;
(4) the probability of acute change (PAC) in *p*-scores, which counts the large *p*-score changes within a sentence, defined as changes between two successive words that are larger than the mean *p*-score plus 1 *SD* (across sentences), where the deviation of the first word is calculated from zero.

The mean indicates the central tendency of the *p*-scores within a sentence (i.e., the average stress); the percentage of peaks indicates the relative frequency of highly stressed words; the MSSD indicates the volatility in the change of stress within a sentence; and the PAC indicates the frequency of large changes of stress within a sentence. The latter indicators, developed to describe fluctuations over time (Jahng et al., 2008) and used in clinical contexts (e.g., Powell et al., 2017), were applied here to capture different patterns of *p*-scores that are not visible through summary measures.

Mixed-effects regression models with crossed random intercepts of sentence and speaker yielded significant differences between the *ironic* and *nonironic* conditions for the mean, MSSD, and PAC, but not for the percentage of peak *p*-scores. Figure 7 illustrates the distributions and central tendencies of these integrated indicators of word prominence by condition. Figure 8 shows the *p*-score trajectories for selected sentence–speaker combinations to illustrate that the integrated indicators capture the individual trajectories of stress within sentences in a complementary way. Sentences 19 and 23 yielded a similar mean level close to the median for all sentences ($Mdn = 0.23$), speakers, and conditions, but differed markedly in the volatility of change within sentences, as indicated by the MSSD. Sentences 5 and 24 yielded a rather high mean level, and sentence 5 showed an average volatility while sentence 24 showed a very high volatility as indicated by the MSSD and PAC. In summary, these examples show that specific combinations of complementary integrated indicators of stress trajectories can be used to differentially identify levels and changes of word prominence within sentences.

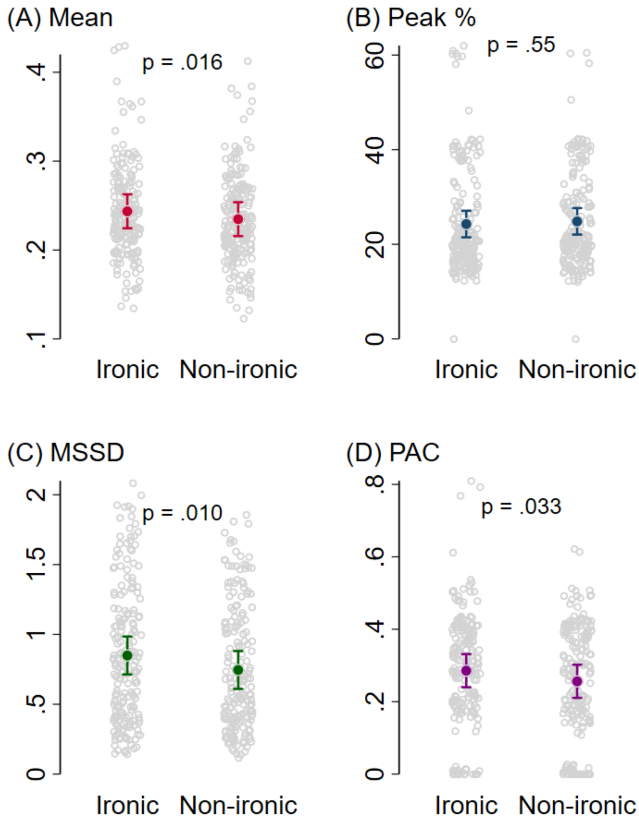
The results show considerable heterogeneity between raters in their perceptions of words as being stressed when listening to sentences that were produced in ironic and nonironic contexts. Although systematic differences between speakers were found, they can only partially

Figure 6
Average Predicted Probabilities of Word Classes Being Perceived as Stressed in the Ironic and NonIronic Conditions



Note. Error bars indicate 95% confidence intervals. The *p*-values are based on Sidak-adjusted multiple between-condition comparisons. See the online article for the color version of this figure.

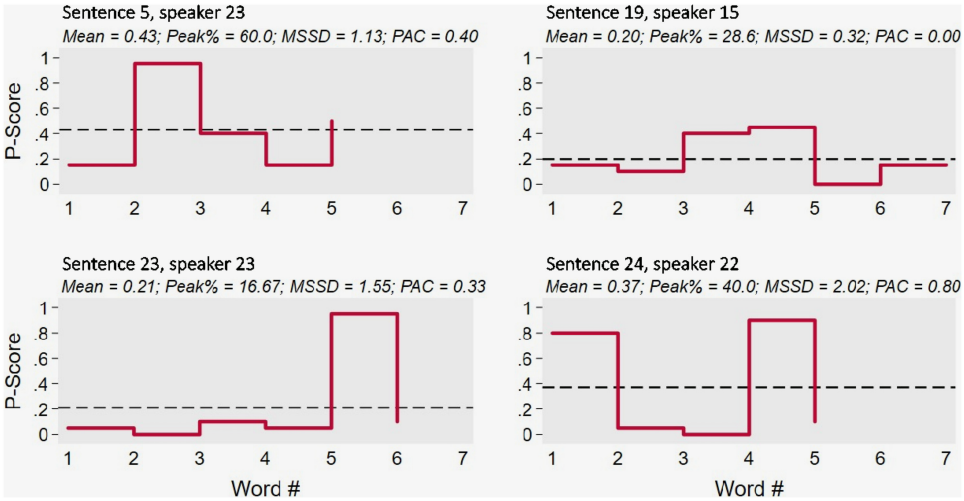
Figure 7
Estimated Marginal Means and Scatterplots of All Observations



Note. Analysis of *p*-scores within sentences (by speaker) by condition. Panel A: mean. Panel B: percentage of peaks. Panel C: mean-squared successive difference (MSSD). Panel D: Probability of acute change (PAC). Error bars indicate 95% confidence intervals. The *p*-values are for condition difference coefficients from mixed-effects regression models. See the online article for the color version of this figure.

Figure 8

Integrated Indicators of Stress (p*-Score) for Four Selected Sentence–Speaker Combinations in the Ironic Condition*



Note. See text for definitions of the indicators. MSSD: mean-squared successive difference; PAC: probability of acute change. See the online article for the color version of this figure.

account for the full heterogeneity in stress perceptions. Nevertheless, significant differences were found between the conditions regarding stress, with the mean, MSSD, and PAC *p*-score measures being higher when irony was intended. Although these differences might be viewed as subtle, Experiment 3 supports the use of *p*-scores and their derivatives as potential predictors of the perception of irony in spoken sentences. Additionally, this experiment provides highly controlled material that can be used to further examine the contribution of acoustic features and the linguistic content of sentences in the perception of prosodic stress, for instance, to disentangle different types of stress, or to clarify the individual differences observed here.

Experiment 4: Predictors of Irony Perception

The fourth experiment was designed to examine the features that are predictive of irony perception in ecologically valid material, with a focus on prosodic stress. Sentences that met our selection criteria (equal likelihood of occurrence with their literal and ironic meanings, marked contrast between these two understandings of the sentence, depending on the preceding context, and homogeneity regarding other dimensions (Experiment 1)) were recorded by speakers in two different conditions (ironic and nonironic) and acoustically analyzed in terms of speech rate and pitch range (Experiment 2) as well as annotated with regard to stress/prominence (Experiment 3). Besides clarifying the relevance and role of acoustic and perceptual features in irony comprehension, this final experiment explored the potential role of derivative measures capturing stress patterns in the communication of irony.

Method

The experimental procedure was ethically approved by the Ethics Council of the Max Planck Society, and the task was undertaken with the written informed consent of each participant.

Participants

Fifty-three participants were recruited via the research participant database of the Max Planck Institute for Empirical Aesthetics (27 self-reported as females, 24 self-reported as males, and two undisclosed; aged between 18 and 69 years, $M = 38.9$ years, $SD = 17.5$). Due to the lack of pilot or published studies with a similar design, we could not rely on effect sizes observed in previous work and therefore did not perform power analyses to estimate the adequate number of participants. Rather, our choice of the sample size was grounded in the usual sample size reported in listening experiments in prosody research.

The inclusion criteria were a minimum age of 18 years, having German as a native language, and intact hearing abilities, along with no participation in any of the previous experiments of this research. On a questionnaire based on Ivanko et al.'s (2004) Sarcasm Self-Report Scale (see the "Procedure" section for Experiment 1), participants self-reported their frequency of using (or believing they use) irony as 3.2 on average on the 6-point scale ($SD = 1.1$); they regularly used irony with different intentions ($M_{\text{critical}} = 3.2$, $SD = 1.2$; $M_{\text{humorous}} = 3.9$, $SD = 1.1$; $M_{\text{teasing}} = 3.4$, $SD = 1.3$); and they self-reported a good ability to understand irony ($M = 3.1$, $SD = 0.9$) and to generally appreciate the irony of others ($M = 3.0$, $SD = 1.3$).

Material

The material consisted of audio recordings of 14 target sentences (selected from the larger stimulus set described in Experiment 1) which had been read aloud and recorded in two conditions (*ironic* and *nonironic*) by 14 speakers (Experiment 2), resulting in a set of 392 recorded stimuli. The speech rate, pitch range, and *p*-score and derivatives for each recording are described in the "Results" sections for Experiments 2 and 3.

Procedure

Participants were seated in front of a computer screen in a single sound-attenuated testing booth. Material was presented via closed headphones (Beyerdynamic DT-770 Pro) at a comfortable sound level, which was kept constant across participants. Participants were informed that they would be listening to single recorded sentences and that their task was to indicate on a scale from 0 (*not at all*) to 5 (*very*) how ironic each sentence was meant to be [in German: "Wie ironisch ist die Aussage gemeint?" 0 = *gar nicht*; 5 = *sehr*]. Participants were offered the option of listening to any recorded sentence a second time. Before the experiment began, participants completed two trials (one precategorized as ironic and one precategorized as nonironic) to familiarize themselves with the procedure.

The 392 stimuli were presented randomly in eight blocks. The session ended with a short questionnaire on demographic data and the participants' personal use of irony (based on Ivanko et al.'s 2004 Sarcasm Self-Report Scale) that addressed their own typical usage and liking of irony in daily life (the results are reported in the "Participants" section for this experiment). The sessions lasted roughly 75 min on average, and the participants received €15 compensation for their participation.

Statistical Analysis

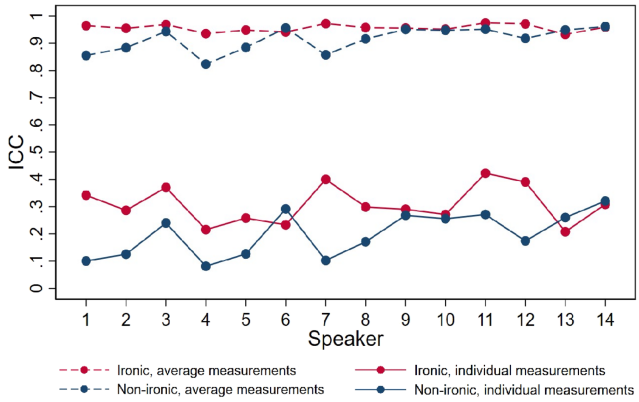
Agreement between listeners was estimated with ICCs from two-way random-effects models for both individual measurements and average measurements. We computed the overall ICCs across all sentences, speakers, and conditions, as well as separate ICCs for condition and speaker to test for differences between conditions using Wilcoxon signed-rank tests. Linear mixed-effects regression models with crossed random intercepts for participant, sentence, and speaker to account for the independence of observations within clusters were used to test for associations between a number of predictors (regarding the rater, speaker, and material characteristics) and the irony ratings. Finally, a multiple predictor mixed-effects regression model was built with the main variable (i.e., the speakers' ironic intention) and variables reflecting acoustic features (from Experiment 2) and prosodic stress characteristics (from Experiment 3) entered as predictors, and the irony ratings as the outcome variable.

Results and Discussion

Across speakers and conditions, there was high agreement on the average measures of ICC (ICC = 0.97; 95% CI [0.96, 0.97]). For individual measures, the listeners' irony ratings showed moderate agreement (ICC = 0.36; [0.33, 0.40]). Figure 9 shows the ICCs for correlations between individual ratings of irony by speaker and condition. The average ICC for speaker-specific irony ratings was significantly higher ($z = 2.67$; $p = .008$) for sentences that were spoken in the *ironic* condition (ICC individual: $M = 0.31$, [0.27, 0.35]; ICC average: $M = 0.96$, [0.95, 0.96]) than for sentences spoken in the *nonironic* condition (ICC individual: $M = 0.20$, [0.15, 0.25]; ICC average: $M = 0.91$; [0.89, 0.94]). The variance in ratings was mainly at the within-cluster level, with very small ICCs (ratio of between-subject to total variance) for the cluster of raters (ICC = 0.06; 95% CI [0.04, 0.09]), sentences (ICC = 0.10; [0.05, 0.18]), and speakers (ICC = 0.07; [0.03, 0.13]). This suggests that none of these factors explained a substantial amount of variance in irony ratings.

Figure 9

Intraclass Correlations for Individual and Averaged Ratings Answering "How Ironic Is the Statement Meant to Be?" for 28 Sentences Spoken by 14 Speakers, Averaged Across 53 Listeners



Note. Spoken sentences were recorded after a speaker read text depicting an ironic or a nonironic context. See the "Method" section for Experiment 2 for further details about the recording sessions. See the online article for the color version of this figure.

The univariate models, reported in Table 2 (where each row represents one model), confirm that the rater and speaker characteristics regarding the use of irony, age, and sex had no effect on the perception of irony (all $ps > .05$). In fact, the only variable that accounted for some variance was the likelihood of a sentence given a context (data from Experiment 1). The less likely a sentence given a context, the more ironic it was perceived to be. This was to be expected, since irony perception is commonly understood to imply a perceived *discrepancy* between what is said (the literal meaning of the relevant expression) and what is actually meant (the context-activated ironic meaning). However, when the sentences were presented in written form without a context (reported in Appendix B), the likelihood of the sentences did not affect the perception of irony, supporting the notion that the effect noted above did not concern the semantic content of the target sentences themselves.

Conversely, and as expected, how a sentence was pronounced strongly impacted the irony rating of the sentence. Figure 10 illustrates the effect size of the main variable, that is, the condition in which the message was intended to be understood (i.e., an *ironic* intention); the results support the quality of the material created in Experiment 1 and recorded in Experiment 2. Also as expected, the effect of the acoustic features (pitch range and speech rate) was in line with the results reported in Experiment 2 (and consistent with the literature). Indeed, a slower speech rate (or longer duration) was strongly associated with irony ratings, whereas the pitch range, which was not different between the two groups of sentences in Experiment 2 and is not consistently reported as relevant in previous studies, did not significantly predict listeners' irony ratings of the recordings.

As illustrated in Figure 10, the prosodic stress characteristics of the recordings (the *p*-scores) did play a role in the irony ratings of the spoken sentences. Specifically, recordings that were more ironic had a greater mean *p*-score, which can be interpreted as "more prosodic stress." Importantly, the derivative measures computed here

Table 2
Results of Simple Mixed-Effects Regression Models (With a Separate Model for Each Predictor)

Predictors	Coefficient (*SE*)	*p*
Rater characteristics		
Irony use and liking	−0.14 (0.09)	.13
Age	−0.002 (0.003)	.63
Sex (1 = female)	−0.04 (0.12)	.74
Speaker characteristics		
Irony use and liking	−0.18 (0.18)	.32
Age	−0.006 (0.006)	.32
Sex (1 = female)	0.05 (0.23)	.83
Material characteristic (likelihood)		
Ratings of written sentences with a context (Experiment 1)	−0.97 (0.05)	<.001
Ratings of written sentences without a context (Appendix B)	−0.04 (0.35)	.91

Note. The likelihood ratings were centered around their grand means. Regression intercept coefficients are not shown.

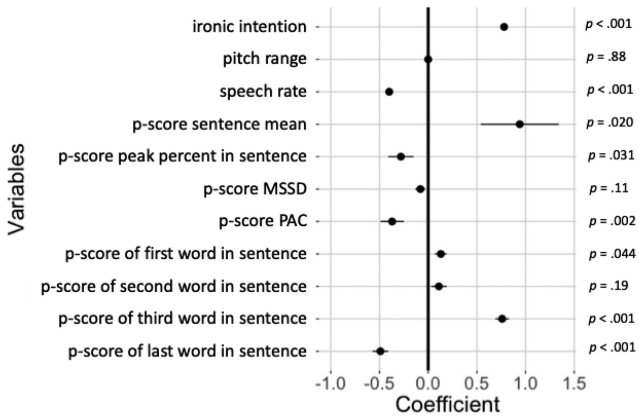
allow us to go beyond the mere magnitude of *p*-scores and to pinpoint which specific type of stress pattern plays a significant role in irony perception. We observed that a lower percentage of *p*-score peaks in a sentence (defined as *p*-score > .35) leads to higher irony ratings, meaning that irony is not associated with constant increased prosodic stresses but might result from the contrast between stressed and unstressed words. If irony were to emerge essentially via contrasts, we should observe large positive and significant effects of the *p*-score MSSD and *p*-score PAC measures, since the first of these represents the squared differences of the *p*-scores of successive words in a sentence and the second represents the *p*-score changes within a sentence. However, the *p*-score MSSD did not reach a level of significance, whereas we observed—unexpectedly—that irony ratings were associated with lower prosodic stress contrasts between successive words (Figure 10).

This apparent contradiction between, on the one hand, the relevance of mean *p*-scores and the role of potential contrasts between consecutive words and, on the other hand, the lack of significance or unexpected effects of contrast measures can be resolved by examining the role of the *p*-scores at different positions within a sentence. This analysis reveals strong effects of the *p*-scores that were calculated at the middle or end of the sentences, whereas the *p*-scores at the first and second positions in the sentences did not contribute to the irony ratings. Therefore, the irony effect might not be about the contrast or shape of the contour per se, but about the occurrence of stress at an earlier position than expected. More specifically, (a) a higher *p*-score of the third word as well as (b) a lower *p*-score of the last word significantly contributed to listeners' irony ratings of the recordings. In other words, a strong stress on the third word and no stress on the last word are important for conveying irony.

Of course, natural speech differs in terms of number of words as well as grammatical and semantic content or syntactic structure, and it would be premature to claim that a change in prominence in the last section of an utterance systematically induces an ironic meaning. Still, the finding of Experiment 3 regarding word classes supports the notion that stresses are not randomly added when speaking ironically. For all the diversity between our target sentences—which range from four to seven words and feature different syntactic structures—and taking into account this variability in sentences by including it as a variable in the mixed-effects model, the present results support some regularity regarding the effect of an earlier stress in detecting irony. Nevertheless, future research that parametrically manipulates stresses (manipulating the stress positions and the number of words within sentences) in highly controlled sentences (with regard to their structure) is strongly encouraged to characterize more narrowly the stress movements between the end and earlier positions of sentences. Additionally, completing such an approach with the investigation of interactions between prosodic cues and other cues known to be relevant (e.g., syntactic content, see Beach, 1991, or Jun & Bishop, 2015, as well as linguistic or visual information) would enrich our understanding of irony comprehension.

Our finding goes beyond the emphatic or contrastive role of prosodic stress that has been repeatedly discussed. Stressing a specific word attracts attention and enhances memory and reaction about a

Figure 10
Results of the Multiple Predictor Mixed-Effects Regression Model Including Acoustic Features (Pitch Range and Speech Rate) and Prosodic Stress Characteristics of the Recordings



Note. Variables were centered around their grand means. Regression intercept coefficients are not shown. The *p*-score peak % (peak = *p*-score > .35 or 75th percentile) indicates the relative frequency of peaks. The *p*-score MSSD (mean-squared successive differences) is a measure of volatility in the stress changes within a sentence. The *p*-score PAC (probability of actual change) is a measure of the frequency of large changes of stress within sentences.

selected information (e.g., Akker & Cutler, 2003; Cutler & Foss, 1977; Fraundorf et al., 2010; Kember et al., 2021). In the present study, the sentences (which were identical in the ironic and nonironic conditions) included differences in terms of syntactic structures (see the criteria in Experiment 1), leading to different word classes at different positions. Note also that we took great care not to use low-frequency words or idiomatic expressions. On the contrary, the material was intended to represent a variety of types of behavior and situations and was designed to span a broad valence range while being limited in terms of criticism/humorous/teasing characteristics (see the discussion of the selection procedure for Experiment 1). Taken together, these criteria were expected to limit the role of the specific emotional and social contents of the target sentences in the observed irony effects. The position of stress might not be random (i.e., on any possible words of the sentences) but we observed that the variable "sentence" did not explain a large amount of variance in irony ratings, suggesting that the linguistic content or grammatical class could not be the primary reasons for the effect observed. Therefore, we hypothesize that the effect of stress might be rather related to the general dynamics or movement of speech prosody. A shift forward in the prosodic stress position (potentially from a nuclear stress; Wheeldon, 2000, p. 258) would thus not have an emphatic or contrastive role but could help the listener infer an ironic meaning. The replication of such findings with filtered speech (Bryant & Fox Tree, 2005), or jabberwocky material (e.g., Banse & Scherer, 1996) or in languages that have other syntactic structures, would confirm that the acoustic signature of irony is not about the distinct degrees of stress placed on specific words but rather about a differential distribution of *stress patterns* over the trajectory of a sentence, and specifically about prosodic stress at an earlier position in the sentence compared to a nonironic use of the same sentence.

A potential explanation of the effect observed here is that shifting the peak sentence stress to an earlier position could prevent the listener, at least temporarily, from interpreting the meaning literally, thus serving as a "warning" or cue (Scott-Phillips, 2014) that, relatively early in the sentence trajectory, motivates a search for—or at least an increased awareness of—potential alternative meanings of the upcoming words. Examining the emergence of alternative or ironic or ambiguous meaning over the course of sentences, for instance with a gating paradigm (see Pell & Kotz, 2011 for a description of this method in the emotional prosody domain) or electrophysiological measures over the course of the utterance (e.g., Mauchand et al., 2021; Regel et al., 2011), would allow to further clarify whether the forward displacement of stress within sentences is a critical cue that places listeners on alert early on in a sentence for a potentially ironic rather than the literal meaning of the still incomplete sentence trajectory.

Although it is now clear that prosodic stress, as examined using the methods of Cole and Shattuck-Hufnagel (2016), contributes to the perception of irony, and that acoustic features are associated with both irony (Table 1) and prosodic stress (e.g., Cole, Mo, & Baek, 2010; Kember et al., 2021), the relation between irony, prosodic stress, and acoustic features remains to be clarified. It is noteworthy that the features found to be relevant for the perception of irony in spoken sentences do not necessarily correspond to the features that significantly differed when comparing the two intended intentions (ironic vs. nonironic). This divergence in the results between Experiments 2–3 and Experiment 4 suggests that even if the intended meaning predicts the perceived meaning, some inconsistencies exist, and that investigations of the relation between irony, prosodic stress, and acoustic features should preferably be based on listeners' perceptions of irony. Regarding the acoustic definition of stress, further examination of the material constructed for the present study (Experiments 1–3) that uses a large set of acoustic features (e.g., the eGeMAPS of Eyben et al., 2016) would plausibly allow for a better understanding of the correspondence between stress and independent or interacting acoustic parameters (e.g., Breen et al., 2010; Cole, Mo, & Hasegawa-Johnson, 2010; Katsika & Tsai, 2021; Mo et al., 2010). We encourage such exploration and provide the material to take this direction (https://doi.org/10.17617/3.IZWNZ1). This could pave the way for a more automatic identification of stress/prominence by means of acoustic analysis.

Finally, this study examined how prosodic stress modulates meaning with a focus on verbal irony. Irony is a fuzzy concept but a particularly suitable test case, since a single sentence can be understood in very different (if not opposite) ways, depending on the differences in its prosodic realization in a given context. Of course, irony can be signaled by cues other than acoustic/stress (e.g., Attardo et al., 2003; Bryant & Fox Tree, 2002; Kunneman et al., 2015; Liebrecht et al., 2013; Rockwell, 2001), and combining these cues would provide interesting insights about their respective roles in irony comprehension. Also, our study can be viewed as a baseline that can be extended to subtypes of irony as well as to other pragmatic functions and thus address the role of prosodic stress in the multiple contexts that make human communication so unique.

Conclusion

By examining listeners' irony ratings (Experiment 4) of recorded sentences (Experiment 2) that were produced following contexts that primed either a literal or an ironic meaning (Experiment 1) and in which prominent words were marked (Experiment 3), we have shown that prosodic stress modulates the detection of an ironic meaning intention. In addition to confirming that acoustic features such as the speaker's rate of speech are associated with ironic meaning, this study thus reveals the crucial role of stress position in predicting ironic understanding. Specifically, a displacement of stress (from the end to an earlier position) within a sentence contributes significantly to listeners' comprehension of irony. In contrast, the individual characteristics of the speakers, listeners, and sentences had only a very limited role in irony detection. Importantly, our constructed material allowed us to conclude that this role of prosodic stress seems to be relatively independent of the semantic content of the stressed words. In this paper, we have shown that prosodic stress, which is typically described in terms of its contrastive or emphatic roles (by accentuating specific words), might have a "warning" role and thus a more general cognitive function for correctly understanding sentences.

Context

This research is part of a larger framework that aims to understand sound–meaning associations. Appreciating music, paying attention to a traffic jam, or inferring a speaker's state of mind may seem to be "easy" cognitive processes, but the core principles that govern all of these, that is, *making sense of sounds*, is altogether poorly understood.

For instance, it is well established that prosodic stress is relevant to verbal communication, but exactly how it contributes to understanding a speaker's intentions is unclear. In this research, we focused on irony, which illustrates perfectly that acoustic properties can have a decisive role in the comprehension of identically worded sentences —that is, whether to interpret them literally or not. To investigate the role of stress in irony perception, we conducted a study that is the fruit of collaborative work between experts from the humanities, social, and cognitive sciences. As a result of this interdisciplinary effort, this research describes the role of prosodic stress in the case of irony comprehension and reveals the characteristics and patterns that modulate meaning, advancing our understanding of sound–meaning associations in human communication.

References

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition*, *6*(2), 81–96. https://doi.org/10.1017/S1366728903001056

Alba-Juez, L., & Attardo, S. (2014). The evaluative palette of verbal irony. In G. Thompson & L. Alba-Juez (Eds.), *Evaluation in context* (pp. 93–116). John Benjamins Publishing.

Anolli, L., Ciceri, R., & Infantino, M. G. (2000). Irony as a game of implicitness: Acoustic profiles of ironic communication. *Journal of Psycholinguistic Research*, *29*(3), 275–311. https://doi.org/10.1023/A:1005100221723

Anolli, L., Ciceri, R., & Infantino, M. G. (2002). From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication. *International Journal of Psychology*, *37*(5), 266–276. https://doi.org/10.1080/00207590244000106

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

Attardo, S. (2002). Humor and irony in interaction: From mode adoption to failure of detection. In L. Anolli, R. Ciceri, & G. Riva (Eds.), *Say not to say: New perspectives on miscommunication* (pp. 159–179). IOS Press.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor: International Journal of Humor Research*, *16*(2), 243–260. https://doi.org/10.1515/humr.2003.012

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Barbe, K. (1995). *Irony in context*. John Benjamins Publishing.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, *30*(6), 644–663. https://doi.org/10.1016/0749-596X(91)90030-N

Belkner, P. (2014). *R128Gain* (Version 1.0.11) [Computer software]. http://r128gain.sourceforge.net/

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*(1), 64–76. https://doi.org/10.3758/BF03197663

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (Version 6.0.39) [Computer software]. http://www.praat.org/

Braga, D., & Marques, M. A. (2004). The pragmatics of prosodic features in the political debate. In *Proceedings of Speech Prosody, Nara, Japan, March 23–26* (pp. 321–324). http://www.isca-speech.org/archive_open/sp2004/sp04_321.pdf

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7–9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., & Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, *20*(2), 116–120. https://doi.org/10.1016/j.cub.2009.11.034

Bruntsch, R., & Ruch, W. (2017). Studying irony detection beyond ironic criticism: Let's include ironic praise. *Frontiers in Psychology*, *8*, Article 606. https://doi.org/10.3389/fpsyg.2017.00606

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, *47*(7), 545–566. https://doi.org/10.1080/01638530903531972

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, *17*(2), 99–119. https://doi.org/10.1207/S15327868MS1702_2

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*(3), 257–277. https://doi.org/10.1177/00238309050480030101

Burgers, C., van Mulken, M., & Schellens, P. J. (2011). Finding irony: An introduction of the Verbal Irony Procedure (VIP). *Metaphor and Symbol*, *26*(3), 186–205. https://doi.org/10.1080/10926488.2011.583194

Burgers, C., van Mulken, M., & Schellens, P. J. (2012). Type of evaluation and marking of irony: The role of perceived complexity and comprehension. *Journal of Pragmatics*, *44*(3), 231–242. https://doi.org/10.1016/j.pragma.2011.11.003

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*(5), 366–381. https://doi.org/10.1016/j.specom.2007.11.003

Cheang, H. S., & Pell, M. D. (2009). Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, *126*(3), 1394–1405. https://doi.org/10.1121/1.3177275

Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language: A cross-linguistic study of English and Cantonese. *Pragmatics & Cognition*, *19*(2), 203–223. https://doi.org/10.1075/pc.19.2.02che

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

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes*, *25*(7–9), 1141–1177. https://doi.org/10.1080/01690960903525507

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2), 425–452. https://doi.org/10.1515/labphon.2010.022

Cole, J., & Shattuck-Hufnagel, S. (2016). New methods for prosodic transcription: Capturing variability as a source of information. *Journal of the Association for Laboratory Phonology*, *7*(1), Article 8. https://doi.org/10.5334/labphon.29

Colston, H. L., & Lee, S. Y. (2004). Gender differences in verbal irony use. *Metaphor and Symbol*, *19*(4), 289–306. https://doi.org/10.1207/s15327868ms1904_3

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139166973

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge University Press.

Cutler, A. (1974). On saying what you mean without meaning what you say. In M. W. Galy, R. A. Fox, & A. Bruck (Eds.), *The tenth regional meeting of the Chicago Linguistic Society* (pp. 117–127). Chicago Linguistic Society.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*(1), 1–10. https://doi.org/10.1177/002383097702000101

Dews, S., & Winner, E. (1995). Muting the meaning: A social function of irony. *Metaphor and Symbolic Activity*, *10*(1), 3–19. https://doi.org/10.1207/s15327868ms1001_2

Dews, S., & Winner, E. (1997). Attributing meaning to deliberately false utterances: The case of irony. *Advances in Psychology*, *122*, 377–414. https://doi.org/10.1016/S0166-4115(97)80142-2

Dropuljić, B., Mršić, L., Kopal, R., Skansi, S., & Brkić, A. (2017). Evaluation of speech perturbation features for measuring authenticity in stress expressions. In N. Nguyen, S. Tojo, L. Nguyen, & B. Trawiński (Eds.), *Intelligent information and database systems* (pp. 685–694). Springer.

Dynel, M. (2018). Deconstructing the myth of positively evaluative irony. In M. Jobert & S. Sorlin (Eds.), *The pragmatics of irony and banter* (pp. 41–57). John Benjamins Publishing.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., & Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, *7*(2), 190–202. https://doi.org/10.1109/taffc.2015.2457417

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*(3), 367–386. https://doi.org/10.1016/j.jml.2010.06.004

Garmendia, J. (2010). Irony is critical. *Pragmatics and Cognition*, *18*(2), 397–421. https://doi.org/10.1075/pc.18.2.07gar

Garmendia, J. (2018). *Irony*. Cambridge University Press.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*(1–2), 5–27. https://doi.org/10.1080/10926488.2000.9678862

Gibbs, R. W., & O'Brien, J. E. (1991). Psychological aspects of irony understanding. *Journal of Pragmatics*, *16*(6), 523–530. https://doi.org/10.1016/0378-2166(91)90101-3

Giora, R. (1995). On irony and negation. *Discourse Processes*, *19*(2), 239–64. https://doi.org/10.1080/01638539509544916

Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol*, *14*(4), 241–257. https://doi.org/10.1207/s15327868ms1404_1

Goupil, L., Ponsot, E., Richardson, D., Reyes, G., & Aucouturier, J.-J. (2021). Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature. *Nature Communication*, *12*(1), Article 861. https://doi.org/10.1038/s41467-020-20649-4

Grice, H. P. (1978). Further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics: Volume 9. Pragmatics* (pp. 113–127). Academic Press.

Grichkovtsova, I., Morel, M., & Lacheret, A. (2012). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*, *54*(3), 414–429. https://doi.org/10.1016/j.specom.2011.10.005

Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford University Press.

Hirano, M. (1981). *Clinical examination of voice*. Springer.

Hirst, D. (2011). The analysis by synthesis of speech melody: From data to models. *Journal of Speech Sciences*, *1*(1), 55–83. https://doi.org/10.20396/joss.v1i1.15011

Holliday, N. R., Bishop, J. B., & Kuo, G. (2020). Prosody and political style: The case of Barack Obama and the L+H* pitch accent. In *Proceedings of the 10th International Conference on Speech Prosody, Tokyo, Japan, May 25–28* (pp. 670–674). ISCA.

Ivanko, S. L., Pexman, P. M., & Olineck, K. M. (2004). How sarcastic are you? Individual differences and verbal irony. *Journal of Language and Social Psychology*, *23*(3), 244–271. https://doi.org/10.1177/0261927X04266809

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*(4), 354–375. https://doi.org/10.1037/a0014173

Jorgensen, J. (1996). The functions of sarcastic irony in speech. *Journal of Pragmatics*, *26*(5), 613–634. https://doi.org/10.1016/0378-2166(95)00067-4

Jun, S. A., & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing* (pp. 217–240). Springer.

Katsika, A., & Tsai, K. (2021). The supralaryngeal articulation of stress and accent in Greek. *Journal of Phonetics*, *88*, Article 101085. https://doi.org/10.1016/j.wocn.2021.101085

Kember, H., Choi, J., Yu, J., & Cutler, A. (2021). The processing of linguistic prominence. *Language and Speech*, *64*(2), 413–436. https://doi.org/10.1177/0023830919880217

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16. https://doi.org/10.1068/v070821

Knox, N. (1961). *The word irony and its context, 1500–1755*. Duke University Press.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General*, *118*(4), 374–386. https://doi.org/10.1037/0096-3445.118.4.374

Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, *21*(2), 127–143. https://doi.org/10.1177/02627X02021002002

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbolic Activity*, *10*(1), 21–31. https://doi.org/10.1207/s15327868ms1001_3

Kumon-Nakamura, S., Glucksberg, S., & Brown, M. (1995). How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, *124*(1), 3–21. https://doi.org/10.1037/0096-3445.124.1.3

Kunneman, F., Liebrecht, C., van Mulken, M., & van den Bosch, A. (2015). Signaling sarcasm: From hyperbole to hashtag. *Information Processing & Management*, *51*(4), 500–509. https://doi.org/10.1016/j.ipm.2014.07.006

Ladd, B. (1980). *The structure of intonational meaning: Evidence from English*. Indiana University Press.

Landgraf, R. (2014, May – 16). *Are you serious? Irony and the perception of emphatic intensification* [Paper presentation]. 4th International Symposium on Tonal Aspects of Languages (TAL), Nijmegen, The Netherlands.

Landis, R. J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Larrouy-Maestri, P., & Morsomme, D. (2014). The effects of stress on singing voice accuracy. *Journal of Voice*, *28*(1), 52–58. https://doi.org/10.1016/j.jvoice.2013.07.008

Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

Lehiste, I. (1976). Suprasegmental features of speech. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225–239). Academic Press. https://doi.org/10.1016/B978-0-12-437150-7.50013-0

Leykum, H. (2019). Acoustic characteristics of verbal irony in standard Austrian German. In *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia, August 5–9* (pp. 3396–3402). Australasian Speech Science and Technology Association.

Liebrecht, C., Kunneman, F., & van den Bosch, A. (2013, June 14). *The perfect solution for detecting sarcasm in tweets #not* [Paper presentation]. 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, GA, United States.

Lœvenbruck, H., Jannet, M. A. B., D'Imperio, M., Spini, M., & Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in French: Slower, higher, wider. In C. C. F. Bimbot, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, & P. Perrier (Eds.), *14th Annual Interspeech Conference* (pp. 3537–3541). International Speech Communication Association.

MATLAB. (2017). *Version 9.2 (R2017a)*. MathWorks.

Mauchand, M., Caballero, J. A., Jiang, X., & Pell, M. D. (2021). Immediate online use of prosody reveals the ironic intentions of a speaker:

Neurophysiological evidence. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(1), 74–92. https://doi.org/10.3758/s13415-020-00849-7

Mauchand, M., Vergis, N., & Pell, M. D. (2018). Ironic tones of voices. In *Proceedings of the 9th International Conference on Speech Prosody, Poznań, Poland, June 13–16* (pp. 443–447). https://doi.org/10.21437/SpeechProsody.2018-90

Mauchand, M., Vergis, N., & Pell, M. D. (2020). Irony, prosody, and social impressions of affective stance. *Discourse Processes*, *57*(2), 141–157. https://doi.org/10.1080/0163853X.2019.1581588

Mo, Y., Cole, J., & Hasegawa-Johnson, M. (2010). Prosodic effects on temporal structure of monosyllabic CVC words in American English. In *The Fifth International Conference of Speech Prosody, Chicago, IL, United States, May 10–14* (pp. 1–4). International Speech Communication Association.

Muecke, D. C. (1970). *Irony*. Methuen.

Murry, T., Brown, W. S., & Rothman, H. (1987). Judgments of voice quality and preference: Acoustic interpretations. *Journal of Voice*, *1*(3), 252–257. https://doi.org/10.1016/S0892-1997(87)80008-5

Niebuhr, O. (2014). "A little more ironic": Voice quality and segmental reduction differences between sarcastic and neutral utterances. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Social and Linguistic Speech Prosody: Proceedings of the 7th International Conference on Speech Prosody* (pp. 608–612). Trinity College Dublin.

Partington, A. (2007). Irony and reversal of evaluation. *Journal of Pragmatics*, *39*(9), 1547–1569. https://doi.org/10.1016/j.pragma.2007.04.009

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE*, *6*(11), Article e27256. https://doi.org/10.1371/journal.pone.0027256

Pell, M. D., & Kotz, S. A. (2021). Comment: The next frontier: Prosody research gets interpersonal. *Emotion Review*, *13*(1), 51–56. https://doi.org/10.1177/1754073920954288

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, *17*(4), 286–290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

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

Pexman, P. M., & Zvaigzne, M. T. (2004). Does irony go better with friends? *Metaphor and Symbol*, *19*(2), 143–163. https://doi.org/10.1207/s15327868ms1902_3

Phillips, L. H., Allen, R., Bull, R., Hering, A., Kliegel, M., & Channon, S. (2015). Older adults have difficulty in decoding sarcasm. *Developmental Psychology*, *51*(12), 1840–1852. https://doi.org/10.1037/dev0000063

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J. J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), 3972–3977. https://doi.org/10.1073/pnas.1716090115

Powell, D., Liossi, C., Schlotz, W., & Moss-Morris, R. (2017). Tracking daily fatigue fluctuations in multiple sclerosis: Ecological momentary assessment provides unique insights. *Journal of Behavioral Medicine*, *40*(5), 772–783. https://doi.org/10.1007/s10865-017-9840-4

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

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

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, *29*(5), 483–495. https://doi.org/10.1023/A:1005120109296

Rockwell, P. (2001). Facial expression and sarcasm. *Perceptual and Motor Skills*, *93*(1), 47–50. https://doi.org/10.2466/pms.2001.93.1.47

Rodero, E. (2011). Intonation and emotion: Influence of pitch levels and contour type on creating emotions. *Journal of Voice*, *25*(1), e25–e34. https://doi.org/10.1016/j.jvoice.2010.02.002

Roy, J., Cole, J., & Mahrt, T. (2017). Individual differences and patterns of convergence in prosody perception. *Journal of the Association for Laboratory Phonology*, *8*(1), Article 22. https://doi.org/10.5334/labphon.108

Scharrer, L., & Christmann, U. (2011). Voice modulations in German ironic speech. *Language and Speech*, *54*(4), 435–465. https://doi.org/10.1177/0023830911402608

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Schwarz-Friesel, M. (2009) Ironie als indirekter expressiver Sprechakt: Zur Funktion emotionsbasierter Implikaturen bei kognitiver Simulation. In A. Bachmann-Stein, S. Merten, & C. Roth (Eds.), *Perspektiven auf Wort, Satz und Text: Semantisierungsprozesse auf unterschiedlichen Ebenen des Sprachsystems* (Festschrift für Inge Pohl, Kola 3, pp. 223–232). Wissenschaftlicher Verlag.

Scott-Phillips, T. (2014). *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Palgrave Macmillan.

Sperber, D., & Wilson, D. (1981). Irony and the use–mention distinction. In P. Cole (Ed.), *Radical pragmatics* (pp. 295–318). Academic Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Harvard University Press.

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

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*(4), 932–943. https://doi.org/10.1093/brain/awl032

Wheeldon, L. (2000). Generating prosodic structure. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 249–274). Psychology Press.

Woodland, J., & Voyer, D. (2011). Context and intonation in the perception of sarcasm. *Metaphor and Symbol*, *26*(3), 227–239. https://doi.org/10.1080/10926488.2011.583197

(*Appendices follow*)

Appendix A

Definitions of Irony

Irony has been discussed in the context of (often openly) competing theories, as *echoic mention* (Sperber & Wilson, 1981), as *pretense* (Clark & Gerrig, 1984), as *allusional pretense* (Kumon-Nakamura et al., 1995), as *indirect negation* (Giora, 1995), and as *relevant inappropriateness* (Attardo, 2000). Attardo (2002, p. 166) claims that the term "irony" is a concept "with fuzzy boundaries, if any," thus questioning the possibility of an all-encompassing theory of irony. Gibbs (2000) suggests that the different theories might not be mutually exclusive and might all have merits for distinct irony phenomena. While this is not the place to provide another review of the different theories (for an overview, see, e.g., Attardo, 2000; for a historical review of the concept of irony, see Knox, 1961; Muecke, 1970), we would like to briefly discuss two aspects on which there has been relatively broad agreement.

First, the existence of a *discrepancy* between what is said, on the one hand, and what is actually meant by an ironic utterance, on the other hand, seems to be a component of all theories of irony (see Burgers et al., 2011). The fact that the ironist wants this discrepancy to be noticed separates irony from lies. Whereas narrower definitions talk of an opposition between the literal and the intended meaning, many authors assume a more complex discrepancy (e.g., Giora, 1995; Sperber & Wilson, 1981). The notion of opposition is convenient for more stereotyped cases of irony such as the sentence "What lovely weather!" said during a downpour, intended to express something like "What awful weather!" Other cases, such as those involving hyperbole or understatement, seem to call for a more nuanced/scalar approach—for example, the sentence "It seems to be raining" uttered during a heavy downpour (both examples are from Sperber & Wilson, 1981).[1]

Second, another important component of most theoretical accounts of irony, which has often been oversimplified in research on the acoustic characteristics of ironic speech, is the *evaluative component* (see Burgers et al., 2011; Garmendia, 2018), which is closely related to speaker feelings. As Grice (1978, p. 124) notes, irony is "intimately connected with the expression of a feeling, attitude, or evaluation." To illustrate this view, he offers the example of two people walking down the street; upon seeing a car with a shattered window, person B says to person A: "Look, that car has all its windows intact"—which was intended to be ironic, but obviously only results in an absurd counterfactual statement. The reason why this statement fails as an instance of irony is, according to Grice, the lack of an evaluative component.

Regarding the *valence* of the evaluation, there has been some debate. On the one hand, irony has often been connected with some kind of negative or critical attitude. In Roberts and Kreuz's (1994) study, 94% of the participants listed *to show negative emotions* as a discursive goal of irony. Grice (1978, p. 124) writes: "I cannot say something ironically unless what I say is intended to reflect a hostile or derogatory judgment or a feeling such as indignation or contempt." Similarly strong views are held by Sperber and Wilson (1986), Garmendia (2010), and Dynel (2018), among others. On the other hand, humor and jocularity have also been discussed as important functions of ironic speech; *to be humorous* was listed as a discursive goal of irony by 65% of the participants in Roberts and Kreuz's (1994) study. Gibbs (2000) and Schwarz-Friesel (2009) express their belief that the critical function of irony has generally been overestimated, whereas its function for displaying conversational humor has been largely neglected. Alba-Juez and Attardo (2014, p. 95) state that irony can "occupy any point within the evaluative continuum."

A number of studies on irony have represented the evaluative component through a distinction between two subtypes of irony, one of which is seen as a more critical and the other one as a kinder form of irony (e.g., Anolli et al., 2000, 2002; Bruntsch & Ruch, 2017; Dews & Winner, 1995; Knox, 1961; Pexman & Olineck, 2002; Pexman & Zvaigzne, 2004). The first type of irony has gone by the name "blame-by-praise"; another common term, especially among English-speaking authors, is "sarcastic irony" or "sarcasm" (or "ironic criticism").[2] Blame-by-praise is assumed to use a positive or "praising" literal sentence meaning to convey a negative or "blaming"/critical evaluation. The second type, "praise-by-blame" (also "kind irony" or "ironic compliment"), works exactly the other way around: a "blaming" literal sentence meaning is used to convey a praising evaluation. Examples used by Anolli et al. (2002; their translation from the Italian) in their study on the acoustics of irony illustrate the difference. As an instance of blame-by-praise, the critical type of irony, they offer "You're a real genius," said to "an insolent, superficial, and presumptuous companion, who took an exam unwisely and failed it" (p. 269). As an instance of praise-by-blame, the kind type, they offer "You're a real donkey," said after the successful performance of a fellow student who was well prepared for the exam but had been doubtful about his abilities beforehand (p. 269). Another example of praise-by-blame, or kind irony, can be found in Bruntsch and Ruch (2017, p. 15): "You are right, the food is inedible!" said to a friend who cooks dinner and apologizes "that he could not salt the food to taste as he has a cold and cannot taste properly."

Praise-by-blame is generally considered to be a nonstandard case of irony (Clark & Gerrig, 1984; Kreuz & Link, 2002; Sperber &

[1] Certain speech acts such as requesting or thanking are another problem for definitions that take the ironic meaning to be the opposite of the literal sentence meaning. For example, Kumon-Nakamura et al. (1995) ask what the opposite of "Thanks for holding the door!" would be. A problem of a different kind involves certain types of implicatures. As will be elaborated in what follows, the object of the discrepancy between the literal and the ironic meaning is the *evaluation* of the person/object/situation at hand. Sometimes, however, this evaluation is not conveyed directly, but only *implied* (cf. Burgers et al., 2012; Partington, 2007). In such cases, the discrepancy between the literal and the ironic meaning cannot (or cannot only) be placed at the level of the explicit sentence meaning. Consider the sentence "I love children who keep their rooms clean," said by a mother upon entering her child's messy room (example from Gibbs & O'Brien, 1991). In the literal version of the sentence (i.e., in a context where the child's room is perfectly tidy), the evaluation—namely, that the child is one of those children and therefore deserves "praise"—is only implied. When intended ironically, this implied evaluation is turned into its opposite, so that the sentence actually conveys "blame," or criticism, regarding the messy state of the room. Note that the explicit sentence meaning, namely that the mother loves children who keep their rooms clean, does not change from the literal to the ironic interpretation—it is the implied evaluation of the situation at hand that does.

[2] The equation of critical forms of irony with sarcasm has become a matter of convention; the question whether all sarcasm is necessarily ironic is outside of the scope of this study.

(*Appendices continue*)

Wilson, 1981). Dews and Winner (1997) report that only 6% of the ironic instances they found in popular television shows belonged to this type (the rest were instances of blame-by-praise, though not all of these were actually critical, and some were more humorous). In addition, Pexman and Zvaigzne (2004) report that these items (they call them "ironic compliments") had the lowest ratings—even lower than literal insults—for "production likelihood."

While the categories of praise-by-blame and blame-by-praise seem to provide a handy distinction, it can be argued that even instances of kind irony (praise-by-blame) contain a critical component (e.g., Barbe, 1995; Dynel, 2018; Garmendia, 2010). In Anolli et al.'s (2002) example, one might suspect that a critical evaluation of the unfounded (and maybe notorious) self-doubts of the well prepared and, in the end, successful student, is expressed along with the potential "praise" for the passed exam. The same applies to Bruntsch and Ruch's (2017) example: Although the speaker probably wants to express that the food tastes good and thus to give praise by words of blame, there might also be criticism of the cooks' rushing ahead with unnecessary and maybe even annoying apologies (for further examples of supposedly "kind irony" that contain a critical component, see Dynel, 2018).

Whereas it has already been argued that praise-by-blame always seems to contain a critical component alongside the intended praise, certain examples of blame-by-praise might be devoid of actual criticism. Consider the following situation: A is visiting her best friend B; when A rings the bell and B opens the door wearing a cucumber mask, A says, "Fancy!" "Fancy" has a positive literal meaning that is apparently not the intended meaning of the ironic utterance; rather, there is a certain evaluative discrepancy between the positive literal meaning and the intended evaluation of how the cucumber mask looks. This implication could well be classified as blame-by-praise. However, it does not seem to be very convincing to suppose that A actually wants to *criticize* B with the remark or to express a negative evaluation of the way B looks. It seems more likely that A is simply making a humorous remark that will be taken as a demonstration of A's and B's closeness, or more generally as a demonstration of conversational humor (Schwarz-Friesel, 2009), potentially with a teasing undertone ("friendly teasing"; Partington, 2007). The valence of an ironic remark is thus dependent on the context. If B wears a strange new dress that she apparently seems to like, an ironic "Fancy!" has the potential to be quite hurtful, whereas in the case of a cucumber mask, there is not much harm to be done. In Alba-Juez and Attardo's (2014, p. 95) words, "it is important to distinguish … between the evaluative polarity expressed by the words or expressions used in the ironic utterance and the resulting evaluation of a given person, thing or situation."

These considerations make a strictly dichotomous a priori categorization of ironic sentences into a negatively evaluative and a positively evaluative subtype seem questionable. More generally, research on the effect that positive and negative emotions exert on the voice as well as the role that perceived valence plays in auditory emotion recognition has a long history (Banse & Scherer, 1996; see Scherer, 2003 for an overview). It seems thus important to document (or control for) the degree to which speech material is connected to *positive* and *negative feelings* for the speaker, given the situational context, in addition to the *humorous*, *teasing*, or *critical* components that the ironic sentences carry.

Appendix B

The Material That was Created and Selected

Thirty sentences were created and presented with their respective scenarios in Experiment 1. The sentences were designed to fit both ironic and nonironic contexts, but we suspected that some of them might be associated with irony more strongly than others. Indeed, sentences with high irony ratings might correspond to conventional/idiomatic/familiar irony, where the ironic meaning is habitually attached to the wording, as is the case, for example, for sentences like "Very funny" or "Tell me about it" (Giora & Fein, 1999). To examine the degree to which the wording of the sentences was itself associated with irony, we presented the 30 target sentences without their ironic and nonironic scenario contexts to 35 native German-language speakers (22 self-reported as females, 13 self-reported as males, aged between 18 and 33, $M = 23.14$, $SD = 3.06$) in the form of an online survey. Each participant was asked to rate each sentence (on the same items that were used to rate the contextualized sentences; see Experiment 1) without having read a text that provides the sentence with a context.

Across the participants, the mean irony rating (on a 6-point scale) of the 30 sentences presented in isolated written form was highly variable, ranging from 0.51 to 3.77 ($M = 2.00$, $SD = 0.92$). As illustrated in Figure B1, the ratings of most of the sentences included the entire scale. For instance, the sentence "Das haben wir toll hingekriegt" was rated as 0 (*not at all ironic*) by seven participants and as 5 (*very ironic*) by seven participants. The scores 1, 2, 3, and 4 were selected by six, five, seven, and three participants, respectively. Thus, even the sentences with irony ratings that were above the average were not rated as "very ironic" by the majority of the participants; rather, the ratings were spread across the scale. We, therefore, decided not to exclude these sentences at this point, but to see whether they might still work as nonironic utterances as soon as they were embedded in their respective scenarios (and also be equally likely to occur in both cases)—which was tested through the online survey in Experiment 1.

(*Appendices continue*)

Table B1
Final Set of 28 Context–Sentence Pairs, Selected on the Basis of the Results of Experiment 1 and Used in Experiment 2

Sentence number	Ironic condition	Nonironic condition	Sentence
1	Matthias und Christine sind zusammen im Supermarkt. Vor einem der Regale sehen sie eine Frau mit einem kleinen Jungen, der Süßigkeiten haben möchte. Doch die Mutter sagt nein, weil zu viel Süßes ungesund sei. Da lässt das Kind seiner Wut mitten im Laden freien Lauf. Es schreit und weint, um die Süßigkeiten doch noch zu bekommen. Christine sagt zu ihrem Freund Matthias:	*Christine ist mit ihrem fünfjährigen Sohn Matthias im Supermarkt. Matthias möchte Süßigkeiten haben, aber seine Mutter sagt nein, weil zu viel Süßes ungesund sei. Matthias fängt an zu schreien und zu weinen. In dem Moment geht ein etwa gleichaltriges Kind brav an der Hand seiner Mutter an ihnen vorbei, und Christine sagt zu Matthias:	"Das ist aber ein gut erzogenes Kind."
	Matthias and Christine are together in the supermarket. In front of one of the shelves, they see a woman with a little boy who wants to have sweets. But the mother says no, because too many sweets are unhealthy. The child lets his anger run wild in the middle of the store. He screams and cries in order to get the sweets after all. Christine says to her friend Matthias:	*Christine is in the supermarket with her five-year-old son Matthias. Matthias wants to have sweets, but his mother says no, because too many sweets are unhealthy. Matthias starts to scream and cry. At this moment, a well-behaved child of about the same age passes by them holding hands with his mother, and Christine says to Matthias:*	"*That is a well-behaved child.*"
4	Maria kocht für sich und ihren Bruder ein Risotto. Sie isst gerne Pilze, und obwohl sie weiß, dass ihr Bruder keine Pilze mag, macht sie diesmal besonders viele ans Essen. In dem Moment geht ihr Bruder hinter ihr vorbei und sagt:	Maria kocht für sich und ihren Bruder ein Risotto. Sie isst gerne Pilze, und weil sie weiß, dass auch ihr Bruder Pilze liebt, macht sie ein paar mit ans Essen. In dem Moment geht ihr Bruder hinter ihr vorbei und sagt:	"Nimm doch ruhig noch ein paar mehr."
	Maria cooks a risotto for herself and her brother. She likes to eat mushrooms, and although she knows that her brother doesn't like mushrooms, this time she adds a lot of them to the meal. At that moment, her brother passes behind her and says:	*Maria cooks a risotto for herself and her brother. She likes to eat mushrooms, and because she knows that her brother likes mushrooms as well, she adds some to the meal. At that moment, her brother passes behind her and says:*	"*Why don't you grab a few more.*"
5	Sylvia und ihre Freundin haben sich für Sonntag zum Barbecue verabredet. Als sie gerade den Grill angezündet haben, ziehen Wolken auf und es fängt an zu regnen. Sylvia sagt zu ihrer Freundin:	Sylvia und ihre Freundin haben sich für Sonntag zum Barbecue verabredet. Als sie gerade den Grill angezündet haben, kommt die Sonne durch und es wird angenehm warm. Sylvia sagt zu ihrer Freundin:	"Der perfekte Tag zum Grillen."
	Sylvia and her friend have arranged for a barbecue on Sunday. Just after they light the barbecue, clouds gather and it begins to rain. Sylvia says to her friend:	*Sylvia and her friend have arranged for a barbecue on Sunday. Just after they light the barbecue, the sun comes out and it becomes pleasantly warm. Sylvia says to her friend:*	"*The perfect day for a barbecue.*"
7	Christoph hatte ein anstrengendes Wochenende und hat sich vorgenommen, Sonntag früh ins Bett zu gehen. Aber als ein Freund vorschlägt, in eine Bar zu gehen, lässt er sich doch überreden. Er kommt erst mitten in der Nacht nach Hause. Als er am Montagmorgen völlig übernächtigt einem Kollegen über den Weg läuft, sagt dieser:	Christoph hat sich nach einer anstrengenden Woche ein Wochenende Erholung vorgenommen. Er hat lange geschlafen, ist spazieren gegangen und hat sich viel Ruhe gegönnt. Als er am Montagmorgen frisch und erholt einem Kollegen über den Weg läuft, sagt dieser:	"Du siehst ja ausgeschlafen aus."
	Christoph had a exhausting weekend and planned to go to bed early on Sunday evening. But when a friend suggested going to a bar, he let himself be persuaded. He did not return home until late at night. When he passes a colleague on Monday morning looking bleary-eyed, the latter says:	*After a exhausting week, Christoph decided on a weekend of relaxation. He slept late, went for walks, and allowed himself some rest. When he passes by a colleague on Monday morning looking refreshed and recovered, the latter says:*	"*You look well rested.*"
11	Philipp versucht, für einen Test zu lernen. Aber weil sein Bruder nebenan wieder einmal die Musik aufgedreht hat, kann er sich nicht konzentrieren. Er geht zu seinem Bruder ins Zimmer und sagt:	*Philipp und Bianka gucken zusammen Fernsehen, und wie immer hat Bianka den Ton sehr leise gestellt. Philipp hat sie schon mehrfach gebeten, lauter zu machen, aber Bianka hat die Lautstärke nur unwesentlich erhöht. Philipp versucht es noch einmal:	"Mach doch noch ein bisschen lauter."
	Philipp is trying to study for a test. But because his brother has once again turned up the music in the next room, he is unable to concentrate. He walks into his brother's room and says:	*Philipp and Bianka are watching television together, and as usual, Bianka has turned the volume very low. Philipp has repeatedly asked her to turn it up, but Bianka has only increased the volume slightly. Philipp tries once again:*	"*Why don't you turn the volume up some more.*"
12	Jörg hat heute keine Lust zu arbeiten, deshalb trinkt er einen Kaffee nach dem anderen und löst dabei Kreuzworträtsel. Als seine Kollegin Nadine zu ihm ins Büro kommt, sagt sie:	Jörg sitzt jeden Tag von früh bis spät im Büro und gönnt sich kaum eine Pause. Als er den vierten Tag in Folge noch spät abends an seinem Schreibtisch sitzt, sagt seine Kollegin Nadine zu ihm:	"Überarbeite dich mal nicht."
	Jörg does not feel like working today, so he drinks one coffee after another and solves crossword puzzles. When his colleague Nadine comes into his office, she says:	*Day after day, Jörg sits in his office from morning to night and rarely allows himself a break. When he still sits in at his desk late in the evening for the fourth day in a row, his colleague Nadine says to him:*	"*Don't overwork yourself.*"
16	Sebastian kocht selten, hat sich heute aber vorgenommen, einen Auflauf zu machen. Während der Auflauf im Ofen ist, telefoniert er und vergisst dabei die Zeit. Als er den Auflauf herausholt, ist dieser komplett verbrannt. In dem Moment kommt seine Mitbewohnerin in die Küche und sagt:	*Sebastian und Andrea sind zusammen im Restaurant. Sie haben Thunfischsteak bestellt, der Fisch ist jedoch nicht sehr gelungen. In den Moment wird am Nachbartisch ein duftendes Gericht serviert. Andrea sagt:	"Der Auflauf riecht aber lecker."
	Sebastian rarely cooks, but decided on making a casserole today. While the casserole is in the oven, he talks on the telephone and loses track of the time. When he takes the casserole out of the oven, it is completely burned. At this moment, his roommate enters the kitchen and says:	*Sebastian and Andrea are together in a restaurant. They have ordered tuna steak, but the fish is not prepared very well. At this moment, an aromatic dish is served at the neighboring table. Andrea says:*	"*The casserole smells so good.*"

(*table continues*)

(*Appendices continue*)

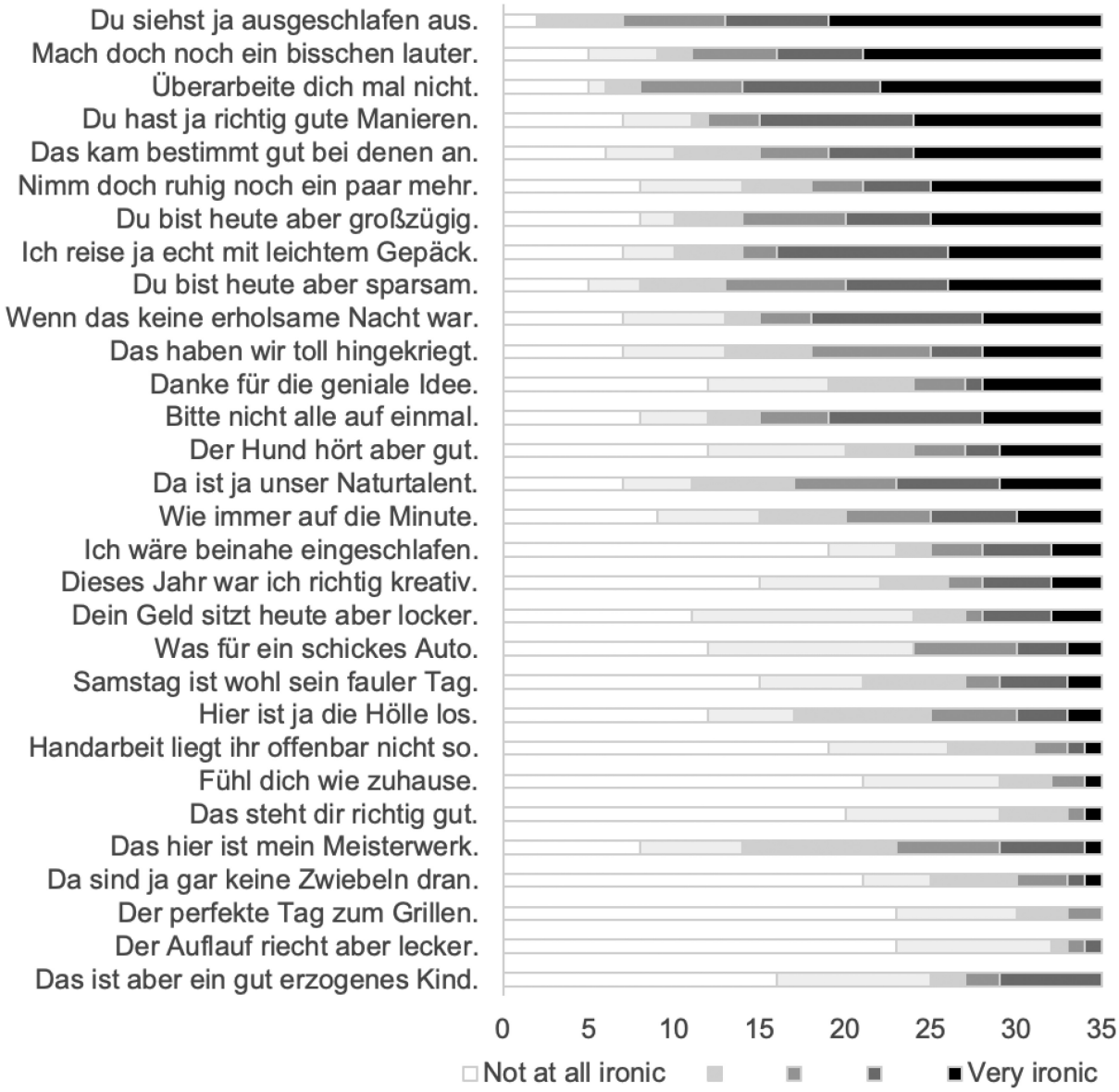
Table B1 (*continued*)

Sentence number	Ironic condition	Nonironic condition	Sentence
17	Jan hat einem Bekannten für wenig Geld ein ziemlich verrostetes Auto abgekauft. Als er seinen Freund Thomas zum ersten Mal mit dem Wagen abholt, sagt Thomas:	Jan hat sich ein tolles neues Auto gekauft. Als er seinen Freund Thomas zum ersten Mal mit dem Wagen abholt, sagt Thomas:	"Was für ein schickes Auto."
	Jan has bought a rather rusty car from an acquaintance for a small amount of money. The first time he picks up his friend Thomas with this car, Thomas says:	*Jan has bought himself a great new car. The first time he picks up his friend Thomas with this car, Thomas says:*	"*What a fancy car.*"
19	Roberts neue Freundin will ihn an Weihnachten ihrer Familie vorstellen. Alles ist sehr förmlich, und weil Robert so nervös ist, trinkt er die ganze Zeit Sekt. Irgendwann ist er betrunken und seine Freundin muss ihn nach Hause bringen. Als Robert am nächsten Tag einem guten Freund davon erzählt, sagt dieser:	Roberts neue Freundin will ihn an Weihnachten ihrer Familie vorstellen. Robert hat seine Freundin vorher gefragt, wer alles kommt, und für jeden ein kleines Geschenk gekauft. Als er am Tag nach Weihnachten einem guten Freund davon erzählt, sagt dieser:	"Das kam bestimmt gut bei denen an."
	Robert's new girlfriend wants to introduce him to her family at Christmas. Everything is very formal, and because Robert is quite nervous, he drinks champagne the whole time. At some point, he is drunk and his girlfriend has to bring him home. When Robert tells a good friend about this the next day, the latter says:	*Robert's new girlfriend wants to introduce him to her family at Christmas. Beforehand, Robert asked his girlfriend who was coming and bought a small present for everyone. When he tells a good friend about this the day after Christmas, the latter says:*	"*That must have been received well.*"
21	Alice schenkt Hanno jedes Jahr Kinokarten zum Geburtstag. Und weil sie auch dieses Jahr keine bessere Idee hat, schenkt sie ihm wieder zwei Karten. Als sie Hanno den Umschlag überreicht, sagt sie:	Alice schenkt Hanno jedes Jahr Kinokarten zum Geburtstag. Dieses Jahr hat sie sich aber überlegt, ihm eine kleine Reise nach Brügge zu schenken. Sie hat Zugtickets und die Unterkunft gebucht und auf einem Stadtplan eine Tour eingezeichnet. Als sie Hanno den Umschlag überreicht, sagt sie:	"Dieses Jahr war ich richtig kreativ."
	Every year, Alice gives Hanno movie tickets for his birthday. As she does not have a better idea this year, she once again gives him two tickets as a present. When handing the envelope to Hanno, she says:	*Every year, Alice gives Hanno movie tickets for his birthday. This year, however, she had the idea of inviting him on a little trip to Bruges. She booked the train tickets and accomodation and marked a tour on a city map. When handing the envelope to Hanno, she says:*	"*This year I was really creative.*"
23	Robin und Daniel haben beschlossen, sich das neue Café anzusehen, das bei ihnen in der Nähe eröffnet hat. Als sie dort ankommen, müssen sie jedoch feststellen, dass sie fast die einzigen Gäste sind. Daniel sagt:	*Robin und Daniel haben beschlossen, sich das neue Café anzusehen, das bei ihnen in der Nähe eröffnet hat. Sie freuen sich auf einen ruhigen Nachmittag. Als sie dort ankommen, ist es jedoch brechend voll und es drängeln sich immer mehr Leute hinein. Daniel sagt:	"Hier ist ja die Hölle los."
	Robin and Daniel decided to have a look at the new café that recently opened nearby. Upon arriving, however, they find that they are almost the only guests. Daniel says:	*Robin and Daniel decided to have a look at the new café that recently opened nearby. They were looking forward to a quiet afternoon. Upon arriving, however, they found the place was jammed, with still more people squeezing in. Daniel says:*	"*This place is packed as hell.*"
24	Anne und Paula sind alte Freundinnen und haben sich am Wochenende bei Paula verabredet. Als Anne klingelt, ist Paula noch nicht fertig und öffnet ihr mit einer Gurkenmaske im Gesicht die Tür. Anne sagt:	*Nächste Woche ist Saschas Abschlussfeier und seine Mutter freut sich über die Gelegenheit, ihn zum ersten Mal im Anzug zu sehen. Sie nimmt Sascha und seine Freundin am Wochenende mit zum Einkaufen. Als Sascha widerwillig im Anzug aus der Umkleidekabine kommt, stößt ihn seine Freundin mit dem Ellenbogen in die Seite und sagt:	"Das steht dir richtig gut."
	Anne and Paula are old friends and have arranged to meet this weekend at Paula's. When Anne rings the door bell, Paula is not ready and opens the door wearing a cucumber mask. Anne says:	*Sascha's graduation ceremony is next week, and his mother is looking forward to the occasion because she will see him in a suit for the first time. On the weekend, she takes Sascha and his girlfriend shopping. When Sascha reluctantly comes out of the fitting room in a suit, his girlfriend pushes her elbow into his side and says:*	"*That looks really good on you.*"
26	Clemens und Hannah sitzen auf einer Bank, als ein Hund vor ihnen stehen bleibt. Der Besitzer ruft den Hund mehrfach aus einiger Entfernung, aber der Hund betrachtet weiterhin Clemens und Hannah. Hannah sagt:	Clemens und Hannah sitzen auf einer Bank, als ein Hund vor ihnen stehen bleibt. Der Besitzer ruft den Hund aus einiger Entfernung, und der Hund rennt sofort zu ihm hin. Hannah sagt:	"Der Hund hört aber gut."
	Clemens and Hannah are sitting on a bench when a dog stops in front of them. The owner calls for the dog from a distance, but the dog keeps looking at Clemens and Hannah. Hannah says:	*Clemens and Hannah are siting on a bench when a dog stops in front of them. The owner calls for the dog from a distance, and the dog immediately runs back to the owner. Hannah says:*	"*The dog listens really well.*"
27	Frank ist Fotograf und arbeitet schon seit Monaten an einer neuen Serie von Bildern. Als er die Fotos einem Freund zeigt, ist versehentlich auch ein komplett überbelichtetes Bild dazwischen geraten. Frank sagt:	Frank ist Fotograf und arbeitet schon seit Monaten an einer neuen Serie von Bildern. Vor allem ein Motiv hat ihn lange beschäftigt. Als er das Foto einem Freund zeigt, sagt er zu diesem:	"Das hier ist mein Meisterwerk."
	Frank is a photographer and has been working for several months on a new series of pictures. When he shows the photographs to a friend, a completely overexposed picture has accidentally been included. Frank says:	*Frank is a photographer and has been working for several months on a new series of pictures. One particular motif has occupied him for a while. When he shows the photograph to a friend, he says to him:*	"*This is my masterpiece.*"

Note. Two contrasting scenarios (i.e., ironic and nonironic conditions) are associated with each of the 14 selected sentences (English translations given in italics). The complete material presented in Experiment 1 is available following the link: https://doi.org/10.17617/3.IZWNZ1. *For the final material of the Non-ironic Condition, five contexts (1, 11, 16, 23, and 24) were selected from the alternative set.

(*Appendices continue*)

Figure B1
The 30 Sentences and Their Irony Ratings



Note. Different shades of gray depict the scores selected on the 6-point scale, ranging from 0 (*not at all ironic*) to 5 (*very ironic*). The bottom axis represents the cumulative number of participants who rated each sentence. Based on the results of Experiment 1, fourteen of the thirty sentences were selected to be used for Experiment 2, in association with two contexts associated with an ironic and a non-ironic condition (see Table B1). The fourteen sentences (in spoken form, and without an associated context) were also used in Experiments 3 and 4.

Received September 22, 2021
Revision received November 30, 2022
Accepted January 11, 2023 ■