Context, facial expression and prosody in irony processing

Gaétane Deliens[a], Kyriakos Antoniou[a,c,d], Elise Clin[a,b], Ekaterina Ostashchenko[a], Mikhail Kissine[a,*]

[a]*ACTE at LaDisco and Neuroscience Institute, Université libre de Bruxelles, CP 175, 50 avenue F.D. Roosevelt, 1050 Brussels, Belgium*
[b]*F.R.S.-FNRS, Fund for Scientific Research, Wallonia-Brussels Federation of Belgium*
[c]*Current affiliation: Department of Theoretical and Applied Linguistics, University of Cambridge, English Faculty Building, 9 West Road, Cambridge, CB3 9DP, United-Kingdom*
[d]*Current affiliation: School of Humanities, Hellenic Open University, 18 Parodos Aristotelous, Patra, 26335, Greece*

Abstract

While incongruence with the background context is a powerful cue for irony, in spoken conversation ironic utterances often bear non-contextual cues, such as marked tone of voice and/or facial expression. In Experiment 1, we show that ironic prosody and facial expression can be correctly discriminated as such in a categorization task, even though the boundaries between ironic and non-ironic cues are somewhat fuzzy. However, an act-out task (Experiments 2 & 3) reveals that prosody and facial expression are considerably less reliable cues for irony comprehension than contextual incongruence. Reaction time and eye-tracking data indicate that these non-contextual cues

*Corresponding author
Email addresses: `gaetane.deliens@ulb.ac.be` (Gaétane Deliens), `ka353@cam.ac.uk` (Kyriakos Antoniou), `elisclin@ulb.ac.be` (Elise Clin), `ekaterina.ostashchenko@ulb.ac.be` (Ekaterina Ostashchenko), `mkissine@ulb.ac.be` (Mikhail Kissine)

entail a trade-off between accuracy and processing speed. These results suggest that interpreters privilege frugal, albeit less reliable pragmatic heuristics over costlier, but more reliable, contextual processing.
Keywords: irony; figurative language; prosody; facial expression; context; eye-tracking

Introduction

Imagine that, as you announce that you will not attend a crisis meeting because of a party, your boss replies 'I love your sense of responsibility!'. Most likely, the incongruity of her comment with the conversational context – broadly understood as shared background knowledge or beliefs (in the classic sense of Stalnaker, 2002) – will (correctly) prompt you to interpret it as ironic. While such ironic utterances pervade our daily conversations, irony is notoriously difficult to define in precise terms (e.g. Gibbs, 2000; Gibbs & Colston, 2012, p. 52) and surfaces under many different guises (such as sarcasm, jocularity, hyperbole, rhetorical question, and understatement). Nevertheless, in one sense or another, all ironically intended messages deliberately mismatch the utterance literal content.[1]

Incongruence with the background context, of the kind just illustrated, is a powerful cue for irony (Gerrig & Goldvarg, 2000; Kreuz & Link, 2002). However, there are indications that a statement may still be interpreted as ironic in the absence of such contextual incongruity, provided that other cues

[1] Of course, contextual incongruence does not necessarily boil down to manifest falsity; for instance, hyperbolic, but nevertheless literally true statements may be ironic (Sperber & Wilson, 1981; Kreuz & Roberts, 1995).

are available (e.g. Kowatch et al., 2013; Jacob et al., 2016). In particular, spoken ironic utterances are often associated with a specific facial expression and a distinctive prosody (e.g. Attardo et al., 2003; Rankin et al., 2009). To the extent that such cues to irony do not directly rely on background context, in what follows we will dub them 'non-contextual', as opposed to contextual incongruity.

The precise role non-contextual cues play in irony processing remains ill understood. On one hand, there is some evidence that a global ironic prosody can be correctly discriminated from a non-ironic one (Bryant & Fox Tree, 2005), provided that the statement is uttered in a familiar language (Cheang & Pell, 2011). And, in fact, many experimental designs implicitly presuppose that ironic prosody is efficient, as they use a distinctive prosody to contrast between ironic and literal stimuli (e.g. Chevallier et al., 2011; Colich et al., 2012; Kowatch et al., 2013). On the other hand, Bryant & Fox Tree (2005) report that a prosodic contour that is successfully discriminated as ironic is also perceptually associated with other dimensions, such as anger or inquisitiveness. Furthermore, the perception of a given prosodic contour as ironic or not may be influenced by the contextual availability of an ironic interpretation (Voyer et al., 2016).

We submit that while ironic tone of voice and/or ironic facial expression may be correctly *discriminated*, these cues are not necessarily efficient in a genuine process of irony *comprehension*. Arguably, successful social interactions do not reduce to tagging statements as literal or not (viz. discrimination), but require the identification of the speaker's discourse goals, and the selection of an appropriate reaction (viz. comprehension; see Kreuz 2000).

Studies in brain-damaged patients suggest a dissociation between these two processes: some patients fail to understand the speaker's intent when contextual and prosody cues are available, even though they are able to identify the tone of voice as sarcastic (McDonald, 2000; McDonald & Pearce, 1996). Yet, irony processing is usually investigated through tasks in which participants have to judge as quickly as possible if statements are ironic or not, thus measuring only the discrimination component. For instance, Bryant & Fox Tree (2002) found that participants successfully discriminate ironic vs. non-ironic utterances based on their prosody.[2] However, making decisions in a binary, forced-choice task is very different from interpreting a message in the same way as would its actual addressee. The precise role of prosody within irony comprehension is further blurred by the fact that Bryant & Fox Tree (2002) found context to be a more powerful cue for ironic judgements than prosody.

A notable exception to such metalinguistic decision paradigms is the study by Kowatch et al. (2013), who designed an innovative 'shopping task' that positions participants as active interpreters. In this experimental design, a puppet faces food items (e.g. an apple and an orange) and utters literal or ironic statements about what it wants to buy (e.g. 'I just love apples'). Only the puppet's tone of voice allows to disentangle ironic criticisms (e.g. 'I just love apples'), literal criticisms (e.g. 'I just hate apples') and literal praise (e.g. 'I just love oranges'). Participants are asked to put in a shopping cart the food item the puppet really wants. In this way, participants' response

[2] There are many other experimental studies that approach irony exclusively through discrimination; see, for instance, Kreuz & Roberts (1995); Climie & Pexman (2008); Epley et al. (2004); Chevallier et al. (2011); Colich et al. (2012).

mirrors their interpretation of the discourse goals of the speaker. The results of Kowatch et al. (2013) display an interesting asymmetry between accuracy and reaction time. The rate of correct responses for ironic items is low (less than 60%), and significantly so relative to literal items. At the same time, the authors report no difference in processing time or in frequencies of first looks to the correct object for ironic and literal criticisms. It could be the case, then, that while ironic prosody and/or facial expression are not very reliable for accurately grasping an ironic communicative intention, they still prompt a rapid, cognitively shallow attribution of ironic intentions to the speaker.

Importantly, Kowatch et al. (2013) did not compare ironic prosody relative to the role of context, so it is unclear whether interpreters still use prosody when context is available, and if yes, whether non-contextual cues merely complement context-based processing or whether they may take precedence over it. There is ample evidence that mastery of irony presupposes complex mental-state attribution skills (e.g. Akimoto et al., 2012; Bryant, 2012; Spotorno & Noveck, 2014). Such mentalising processes require inferring the speaker's intention by assessing the utterance content against the background context. Some theorists hold that any type of pragmatic processing involves complex, context-based inferences about the speaker's communicative intentions (Sperber & Wilson, 2002). Consistently with this idea, in 5- to 7-year-old children, it is the capacity to attribute multilayered mental states, and not ironic prosody, that predicts correct discrimination between irony and white lies (Wimmer & Leekam, 1991; see also Filippova & Astington, 2010).

However, it is also plausible that conversationally experienced interpreters sometimes rely more on salient non-contextual cues than on context. For instance, Deliens et al. (2017) recently found that in the presence of salient ironic prosody, participants do not engage in context-based perspective-shifting to gauge the sarcastic nature of a message. According to the parallel-constraint-satisfaction account (Katz, 2005; Pexman, 2008), all cues are processed in parallel and activate a certain – possibly ironic – interpretation. However, as acknowledged by Pexman (2008) herself, this model does not currently provide any indication as to the relative weight of different cues. A more radical idea, to which we subscribe, is that the presence of salient, albeit perhaps less reliable, non-contextual cues prompts interpreters to disregard costlier contextual processing. This hypothesis is consistent with the Direct Access view (e.g. Gibbs, 2002), which predicts that interpreters do not always need to analyse literal meaning in full to form a hypothesis about the meaning communicated by the speaker. It is also in line with a model of pragmatics according to which interpreters are driven by considerations of cognitive economy, and do not necessarily engage in extensive context-driven reasoning about speaker's intentions (Kissine, 2016; see also Ferreira & Patson, 2007).

By contrast, Giora's Graded Salience theory (Giora, 2003; Giora et al., 2015) holds that, unless the sentence form bears a conventional or by-default association with irony,[3] utterance literal, compositional meaning will nec-

[3] So far, evidence for such by-default ironic meanings, outside conventionally ironic constructions, is limited to negative statements of the form 'X is not the most Y' (Giora et al., 2015).

essarily be activated first before being rejected in favor of a contextually computed ironic interpretation. On different grounds, authors like Sperber & Wilson (2002), who hold that any pragmatic processing involves context-based inference of speaker's intentions, would also have to predict that non-contextual cues can supplement, but not replace context in irony comprehension.

Summing up, two related research questions clearly emerge from the current state of the art: one about the reliability of non-contextual cues, and the other about the relative processing roles of contextual and non-contextual cues. In Experiment 1 of this paper we assess the discrimination of ironic prosody relative to neutral prosody, as well as to positive or negative literal prosody; we also test, in the exact same way, the discriminability of ironic facial expression. (While the discrimination of ironic prosody has been previously investigated, to the best of our knowledge no such evidence is available for ironic facial expression.) In Experiments 2 and 3 we assess how the same prosody and facial cues, as well as contextual information impact irony comprehension, using an act-out task inspired by Kowatch et al. (2013). Our Hypothesis 1 is that in a categorization task ironic prosody and ironic facial expression should allow correct discrimination of ironic items. In line with the model put forward by Kissine (2016), as well as with the Direct Access view (Gibbs, 2002), we predict that in the act-out tasks of Experiments 2 and 3 the presence of salient – albeit potentially less reliable – non-contextual cues should prompt interpreters to bypass costlier contextual processing. That is, our Hypothesis 2 is that ironic prosody and facial expression are privileged in irony comprehension at the expense of costlier, but more accurate assessment of the utterance literal content relative to the context. Accordingly, one should expect non-contextual cues to be associated with faster responses; furthermore, if, as we predict, the processing of ironic prosody or facial expression does not supplement context-based assessment of the compositional meaning, non-contextual cues should not entail any accuracy gain relative to contextual incongruence.

Our Hypothesis 2 may also be seen as one possible implementation of the parallel-constraint-satisfaction model of irony interpretation (Katz, 2005; Pexman, 2008). As we already mentioned, this model predicts that contextual and non-contextual cues are processed in parallel. If parallel processing of all cues must be completed before the outputs are weighted and the final interpretation reached, then, contrary to our predictions, the presence of non-contextual cues along with contextual incongruence should lead to longer reaction times. However, the parallel-constraint-satisfaction model could also assign greater weight to non-contextual cues, in such a way that costlier processes terminate before completion in the presence of salient ironic prosody or facial expression. In that case, this model would be entirely consistent with our predictions.

By contrast, the predictions generated by our Hypothesis 2 are incompatible with accounts that posit obligatory processing of contextual cues in irony derivation. According to Relevance theory (Sperber & Wilson, 2002) or the Graded Salience theory (Giora, 2003), ironic interpretation necessarily involves the assessment of the utterance content against the background context.[4] If so, non-contextual cues would merely add a supplementary source of

[4] Again, with the possible proviso concerning conventional or by-default ironic meanings

evidence for (or against) ironic interpretation, the integration of which should lead to increased response times – especially if these cues are not entirely congruent with context-based processing. Furthermore, if non-contextual cues supplement obligatory context-based processing, accuracy levels should increase when ironic prosody or facial expression combine with contextual incongruence.

Experimental stimuli

All experiments reported in this paper use (part of) thirty-six videos (and two practice trials videos) in French, in which two individuals discuss two items placed on the table in front of them. Each video can be subdivided in (*a*) a context segment, (*b*) a labeling and question segment, (*c*) a pause segment and (*d*) the target utterance. First, the character (A) at the right of the screen mentions her/his knowledge about the second character's (B) preferences regarding the two items placed on the table (e.g. 'George, I know that you like chemistry and that you really hate physics. But reading a physics book could be interesting.'). This part contains contextual background information useful for detecting potential sarcasm. Second, A labels the two items on the table to ensure participants could identify them (e.g. 'Here is a chemistry book and here is a physics book.') and then asks B if s/he wants one of the two items (e.g. 'Would you like the physics book as a gift, now?'). Third, a black screen appears and participants are asked to press the space bar to hear B's reply. Fourth, a video appears with B's

(Giora et al., 2015); see footnote 3.

reply, viz. the target (e.g. 'No, you know how much I hate physics!'). In Experiments 2 and 3 the video freezes until the participant selects the item she believes B really wants. Clicking the right mouse button corresponds to the object at the right of the screen and clicking the left mouse button to the object at the left of the screen. Time course of a video stimulus is illustrated in Figure 1.

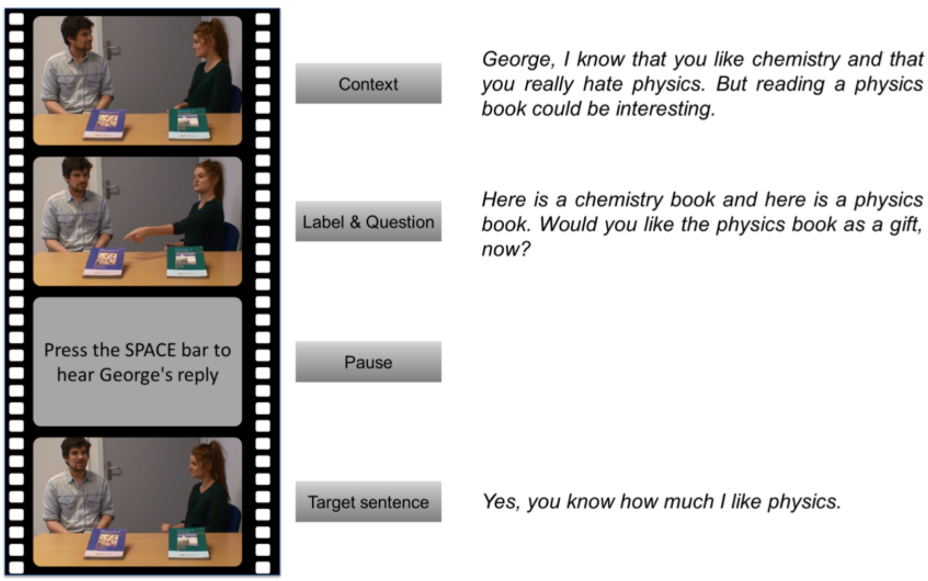

Figure 1: Time course of a full stimulus

The thirty-six videos are drawn on twelve scenarios (see Figure 2). There are three versions of each scenario based on the meaning of the target: an *Ironic*, a *Literal Yes* and a *Literal No* version (see Table 1 for a translated example of the three versions of a scenario). The meaning of the target was manipulated by modifying the contextual information (B's preferences are congruent vs. incongruent with B's reply) and the beginning of the target

(‘Yes, you know how much I like X’ vs. ‘No, you know how much I hate X’). Three professional actors formed three pairs (actor 1 and 2, actor 1 and 3, actor 2 and 3). Each actor performed in 12 videos, 6 as individual A and 6 as individual B.

<table>
<tr><th></th><th>LITERAL NO</th><th>IRONIC</th><th>LITERAL YES</th></tr>
<tr><td>Context</td><td colspan="2">George, I know that you like drinking milk and that you really don’t like drinking tea for breakfast.
But some kinds of tea are really nice.</td><td>George, I know that you like tea for breakfast and you have said this to me many times.</td></tr>
<tr><td>Label & Question</td><td colspan="3">George, here is a glass of milk and here is a cup of green tea. Would you like the cup of green tea with your breakfast, now?</td></tr>
<tr><td>Target</td><td>No, you know how much I hate tea for breakfast!</td><td colspan="2">Yes, you know how much I like tea for breakfast!</td></tr>
</table>

Table 1: Example of three versions (Literal No, Ironic and Literal Yes) of a scenario

Each video vignette was assigned to one of 6 conditions defined in function of the presence or the absence of specific cues: Context only, Prosody only, Context & Prosody, Context & Facial expression, Prosody & Facial expression, and Context, Prosody & Facial expression. In conditions where contextual cues were not available, the context segment of the video was removed. In the conditions where the prosodic cues were not available, the actor was asked to utter the target sentence on a monotonous tone of voice. By contrast, when prosody cues were present, the actor was instructed to utter the target sentence with the corresponding prosody (literal positive for

Literal Yes items, literal negative for Literal No items and ironic for Ironic items). The same applies to facial expression cues. Each target is thus associated with one of the four following prosody contours and one of four following facial expressions: *Ironic*, *Literal Yes*, *Literal No* and *Neutral*.

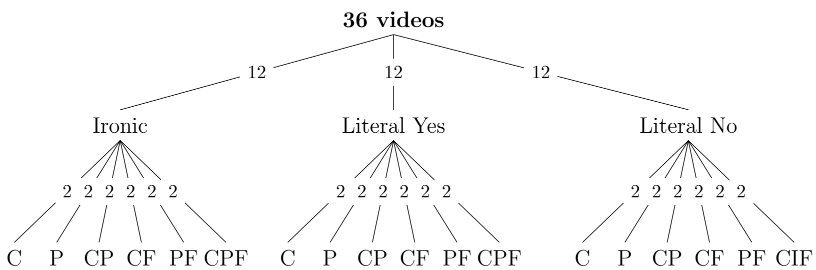

Figure 2: Assignment of the 36 experimental videos across conditions. Contextual cue (C), Prosody cue (P), Facial expression cue (F)

Recall that our overarching objective is to disentangle potentially differential roles of contextual and non-contextual cues in irony processing. In order to do so, one should avoid using stimuli whose ironic character is inherently difficult to grasp, as this would entail a markedly low accuracy on a sub-set of items. This is why, as in Kowatch et al. (2013), our ironic stimuli always consist of negative meanings associated with literally positive sentences ('Yes' sentences; see Table 1); utterances of the opposite valence — literally negative sentences with a positive ironic meaning — are much less canonical forms of irony, and have been found to be particularly difficult to grasp (e.g. Kreuz & Link, 2002; Climie & Pexman, 2008; Filippova & Astington, 2010). For a similar reason, no item is associated with Facial expression as the only cue towards the (non-)ironic meaning. A marked facial expression with no context support and combined with a neutral prosody would be too an unnatural and ambiguous cue, and would increase the risk of chance performance.

It is possible that the most salient acoustic correlates of ironic prosody are not intrinsic, but rather relative to the surrounding discourse (Bryant, 2010). However, ironic prosody has also been reported to have inherent acoustic correlates at the level of fundamental frequency (F0), intensity and delivery rate (e.g. Rockwell, 2000; Bryant, 2010; Anolli et al., 2000; Lœvenbruck et al., 2013). For all target segments, F0 (in Hz, every 3 ms), intensity (in dB, every 11 ms) and syllable duration (in ms) were measured using Praat (Boersma & Weenink, 2017). A linear regression implemented in the lmer package (Bates et al., 2015) in R (R Core Team, 2016) reveals that only mean intensity predicts the type of prosody ($F(3,32) = 6.84$; $p = 0.001$; all other ps> 0.3). Because the first word of the target is always a monosyllabic *yes* (*oui*) or *no* (*non*), it made sense to assess whether ironic prosody is reflected in its acoustic properties. For first syllables of target utterances, we found an effect of prosody type on the mean intensity ($F(3,32) = 4.472$, $p = 0.001$) and of syllable length ($F(3,32) = 8.586$, $p < 0.001$). As can be seen from Table 2, the ironic prosody of our stimuli is associated with a significantly higher mean intensity in the whole sentence and in the first syllable in comparison with all other prosody types, and with significantly longer first syllable relative to Neutral prosody.

In an attempt to objectify differences in facial expressions, five components have been analyzed during the target sentence: eyes, mouth, eyebrows, head and upper body. If there was at least one movement in a component (e.g., raising eyebrows), it was scored as 1; if the component remained still,

	Mean intensity (sentence)	Mean intensity (first syllable)	Length (first syllable)
Intercept (Ironic)	73.48***	71.54***	553.50***
	(0.79)	(0.96)	(41.34)
Literal Yes	−3.35**	−4.22**	−106.37
	(1.11)	(1.36)	(58.47)
Literal No	−3.68**	−0.31	−102.50
	(1.11)	(1.36)	(58.47)
Neutral	−4.43***	−2.64*	−261.33***
	(1.01)	(1.24)	(53.38)
R^2	0.39	0.30	0.45
Adj. R^2	0.33	0.23	0.39
Num. obs.	36	36	36
F statistic	6.84	4.47	8.59

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Coefficients (and standard errors) of linear regressions of mean intensity (whole sentence and first syllable) and of syllable length (first syllable) on prosody Type

it was scored as 0. A total facial expression score (ranging from 0 to 5) for an item is then the sum of the five component scores. Actors have been consistent in their way to display emotions. For neutral facial expressions ($n = 18$), they all kept still ($m = 0.28, sd = 0.57$). For marked facial expressions (ironic or literal, $n = 18$), actors were all using combinations of different movements ($m = 2.94, sd = 1.26$). An ordinal regression implemented in the clm function from the ordinal package (Christensen, 2015) in R (R Core Team, 2016) reveals a significant effect of video category (Marked vs. Neutral facial expression; $z = 34.98, p < 0.001$) with a higher number of component movements in the Marked facial expression condition, but no effect of Type (Ironic vs. Literal Yes vs. Literal No; $p > 0.19$).

Further qualitative inspection reveals a difference between Marked facial expressions across Types. In the Literal Yes condition, actors look enthusiastic or happy: they display sincere static smile (mouth and eyes), as well as little and brief eyebrows upward movements. In the Literal No condition, actors look upset or angry: they visibly accentuate plosive consonants, shrug and slightly project their torsos, display accentuated and long eyebrow raising and frowning, produce head negation movements at the syllable rhythm, as well as increased blinking, half-closure and a wide-open eyes. In the Irony condition actors look sarcastic: they produce many wide eyebrow upward movements, eyebrows are also often arched, they display huge false frozen smiles ending in a cold expression, sway their body, and produce many repeated wide and slow head movements.

Experiment 1

The aim of Experiment 1 is to assess whether prosody or facial expressions can be correctly discriminated as ironic against a sincere or neutral counterpart. To this end, we isolated these cues from target segments of our material, and ran a first experiment were participants had to rate these isolated cues from sincere to ironic on a 7-point Likert scale.

Participants

One hundred thirty-nine volunteers participated in this first experiment. Data from 12 participants were discarded from the analyses because they were not French native speakers (N=9) or due to a problem with the network connection during the task (N=3). The remaining 127 volunteers (63 males) ranged in age from 16 to 53 years ($m = 26.68, sd = 7.50$).

Procedure

To investigate the discrimination of ironic prosody and facial expression, we isolated the targets of the 36 videos described in the previous section, and next, split them into a stand-alone audio file and a video file with muted sound. The resulting audio and video files all belong to one of the following four types: *Literal Yes*, *Literal No*, *Neutral* and *Ironic*. The 36 audio files and 36 video files were presented in the same randomised order across participants, using the online survey application LimeSurvey, with implemented audio and video players. The experiment was composed of two parts: the scoring of the facial expression and the scoring of the prosody. One group of participants (Group *Prosody-Facial Expression*: N=66, 43 males, age 16-49 years, $m = 26.06, sd = 7.04$) scored first the prosody and then the facial expression, while the second group performed the task in the opposite order (Group *Facial Expression-Prosody*: N= 61, 20 males, age 19-53 years, $m = 27.34, sd = 7.98$). In the prosody part, participants were asked to listen to each sound excerpt and to rate the speaker's tone of voice on a 7-point Likert scale ranging from (1)-'completely sincere' to (7)-'completely ironic'. They were instructed to focus on the prosody independently from the content of the sentence. In the facial expression part, participants were asked to watch each video and to rate the (left side of the screen) speaker's facial expression on the same 7-point Likert scale. We pointed out that the sound of the videos was removed to allow them to focus on the speaker's facial expression.

Results

Participants' ratings of prosody and facial expression were analysed with cumulative link mixed models, using the clmm function from the ordinal package in R (Christensen, 2015). Here and in Experiments 2 and 3, the significance of the fixed effects was assessed by performing likelihood ratio tests in which a model containing the fixed effect is compared to another model without it, but with an otherwise identical structure (Baayen et al., 2008). Post-hoc comparisons of least square-means, with Tukey adjustment for multiple comparisons and Satterthwaite method for estimating degrees of freedom, were performed using the lsmeans package (Lenth, 2016).

Prosody (audio data). Figure 3 displays irony rating on a (1-7) Likert scale per type of prosody. Cumulative link multilevel logit regressions with by-subject random intercepts and random slopes for the Type factor (Ironic vs. Literal Yes vs. Literal No vs. Neutral) revealed a significant effect of Type ($\chi^2(3) = 233.18$, $p < 0.001$), but not of Group (Prosody-Facial Expression vs. Facial Expression-Prosody) ($p = 0.8$). The model, displayed in Table 3, shows that Ironic prosody leads to significantly higher irony scores than all other types of prosody. Since, however, other levels of Type do not seem equivalent (see Figure 3), we conducted post-hoc comparisons, which confirmed that Literal No type was rated as significantly less ironic than Literal Yes ($z = -16.48$, $p < 0.001$) and Neutral ($z = -17.37$, $p < 0.001$), while there was no difference between Literal Yes and Neutral types ($p = 0.92$).

Facial expression (Video data). Figure 4 displays irony rating on a (1-7) Likert scale per type of facial expression. Cumulative link multilevel logit

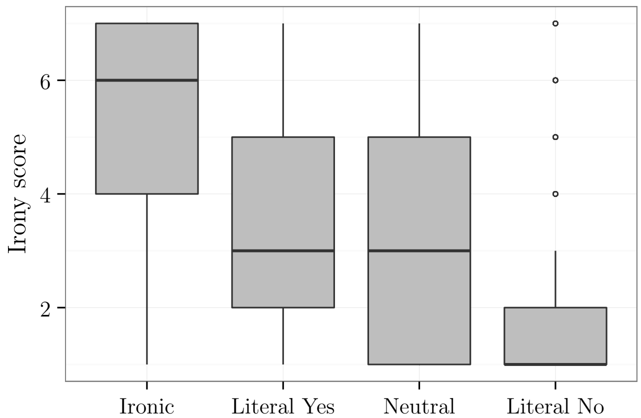

Figure 3: Tukey box-plots for ratings of audio-files per Prosody type

regressions with by-subject random intercepts and random slopes for the Type factor (Ironic vs. Literal Yes vs. Neutral vs. Literal No) revealed a significant effect of Type ($\chi^2(3) = 176.58$, $p < 0.001$), but not of Group (Prosody-Facial Expression vs. Facial Expression-Prosody; $p = 0.38$). As shown in Table 4, Ironic facial expression prompts significantly higher irony scores relative to other types of facial expression. Again, Figure 4 suggests that not all non-ironic levels are equivalent; post-hoc comparisons show that Literal Yes type is judged more ironic than Neutral ($z = 7.9$, $p < 0.001$) and Literal No ($z = 11.81$, $p < 0.001$), and that Neutral type is judged more ironic than Literal No ($z = 5.17$, $p < 0.001$).

Discussion

Experiment 1 confirms that in a rating task that explicitly opposes ironic to literal stimuli, ironic prosody and facial expression can be correctly discriminated against literal (positive or negative) or neutral prosody and facial

	Irony ratings of prosody
Literal Yes	$-1.38\ (0.09)^{***}$
Neutral	$-1.62\ (0.09)^{***}$
Literal No	$-3.61\ (0.15)^{***}$
1—2	$-2.78\ (0.08)^{***}$
2—3	$-1.84\ (0.07)^{***}$
3—4	$-1.32\ (0.07)^{***}$
4—5	$-0.76\ (0.07)^{***}$
5—6	$-0.16\ (0.06)^{*}$
6—7	$0.68\ (0.06)^{***}$
Num. obs.	5040

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Coefficients (and standard errors) of fixed effects of the cumulative link multilevel logit regressions of irony ratings on Prosody Type on irony ratings (Ironic prosody is the reference level). The three first lines of the table report the coefficients for the predictors included in the CLMM. Lines 4 to 9 report the coefficients for each transition of the 7-point Likert scale.

expression. However, using a gradual Likert scale instead of a forced 'ironic vs. literal' binary choice (unlike, e.g. Bryant & Fox Tree, 2002; Voyer et al., 2016) allows a finer-grained insight into the identification of prosody and facial expression as ironic or not. Of course, ironic prosody and facial expression led to unambiguously higher scores on an irony scale. (In the case of ironic prosody, such perceptual judgements thus reflect the acoustic salience of the stimuli intensity.) Recall, however, that our non-ironic audio and video stimuli fall into three different types: positive prosody/expression, negative prosody/expression and neutral prosody/expression. If ironic cues were completely unambiguous, one should expect the remaining three types to receive the same ironic scores. And yet, literal positive prosody contours and facial

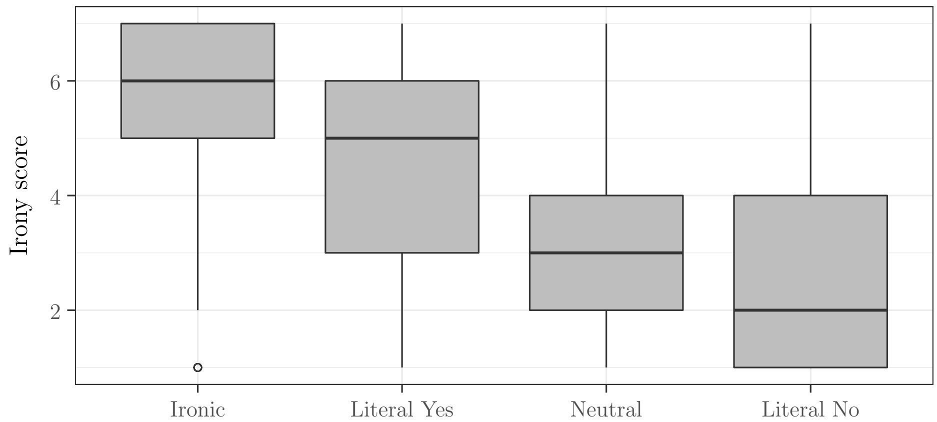

Figure 4: Tukey box-plots for ratings of video-files per facial expression type

expressions were judged more ironic than their literal negative counterparts. In other words, positive or neutral prosody and facial expression are more ambiguous as to the ironic vs. sincere meaning of an utterance. Furthermore, while ironic cues were accurately discriminated against the other ones as more ironic, the distinction could have been artificially boosted up by the fact that participants' task is reduced to merely rating the ironic dimension of various stimuli. In real life, however, interpreters have to decide on speaker's intentions rather than classify the utterance along an ironic-literal continuum. It is therefore likely that, in such settings, the actual reliability of non-contextual cues to irony is considerably lower than what the results of Experiment 1 might suggest.

Experiment 2

The second experiment uses an act-out irony comprehension task in order to compare the impact of ironic prosody and facial expression relative to

Irony ratings of facial expressions	
Literal Yes	-1.42 (0.14)***
Neutral	-2.64 (0.15)***
Literal No	-3.28 (0.18)***
1—2	-4.21 (0.13)***
2—3	-3.10 (0.13)***
3—4	-2.42 (0.12)***
4—5	-1.58 (0.12)***
5—6	-0.71 (0.12)***
6—7	0.28 (0.12)*
Num. obs.	5292

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Coefficients (and standard errors) of fixed effects of the cumulative link multilevel logit regressions of irony ratings on Facial expression Type (Ironic facial expression is the reference level).The three first lines of the table report the coefficients for the predictors included in the CLMM. Lines 4 to 9 report the coefficients for each transition of the 7-point Likert scale.

that of contextual incongruence. First, we expect that, in spite of being correctly discriminated in Experiment 1 ironic prosody and facial expression, should not improve accuracy in irony comprehension relative to contextual incongruence. Second, in spite of not being associated with an accuracy gain relative to context, we expect these non-contextual cues to lead to shorter response times.

Participants

Fifty-six students gave their written informed consent to participate in this study approved by the Faculty Ethics Committee at the Université libre de Bruxelles. Participants were recruited according to the following criteria: native French speakers, normal or corrected-to-normal vision, no hearing difficulties, no history of neurological disorders. Ten participants were excluded from statistical analyses because they were not native speakers of French. Two other participants were excluded because they reported a history of attention-deficit/hyperactivity disorder, a neurodevelopmental disorder associated with pragmatic deficits, including difficulties understanding irony (e.g. Caillies et al., 2014; Staikova et al., 2013; Bignell & Cain, 2007). The age of the 44 remaining participants (15 males) ranged between 18 and 26 years ($m = 20.43$; $sd = 1.47$).

Procedure

The task was run in 64-bit Windows 7 using Tobii Studio^{TM} 3.2.1 software, which controlled the stimuli presentation in a random order and recorded participant's response and reaction times. A Tobii pro X2-60(Hz) screen-based eye-tracker device (Tobii Technology, Inc. Stockholm, Sweden) was used to record participants' eye movements during the target sentence. A five-point calibration procedure designed by Tobii Studio was used before the irony task.

Each participant was seated at a distance of $\pm$ 60 cm in front of a 16.5-inch monitor (resolution: 1920 x 1080 pixels) wearing headphones. The stimuli were presented at a comfortable sound pressure level (65dB $\pm$ 5dB). Following eye-tracker calibration, participants were presented with the following instructions on screen:

> In each trial of this task you will watch videos with short conversations between two individuals. One person will ask a second person questions about two items on a table. After each question, you will watch a video with the second person's reply. Listen

> carefully to what the first person says and to the second person's reply! At the end of the second person's answer you will have to give him/her the item you believe he/she really wants. You should press the left mouse button if you think that he/she really wants the item at the left or the right mouse button if you believe that he/she really wants the item at the right. Next trial will start automatically after your answer.

Participants first completed two training trials (one Literal Yes and one Literal No item) before the experimental phase began.

Results

In order to assess the relative impact of context, prosody and facial expression, each stimulus was associated with a binomial variable Context, Prosody and Facial expression, depending on which cue(s) were associated with the target. Note that our Sincere No items were unambiguously literal. It is possible, then, that participants realize, in the course of the experiment, that any answer starting with 'No' would lead to a non-ironic interpretation. In order to assess this possibility, we examined the effect of the linear Order of the stimuli.

Accuracy. A correct interpretation of a target corresponds to a trial where the participant accurately selects the object the second character (B) in the video really wants (see the description of the stimuli). For Literal Yes items, the correct choice was the object named in the target (e.g. 'Yes, you know how much I like physics!'), whereas in Literal No (e.g. 'No, you know how much I hate physics!') and Ironic items (e.g. 'Yes, you know how much I like physics!'), it was the other object displayed in the video. As can be seen from Figure 5, while accuracy rate is generally high, it is lower for Ironic targets.

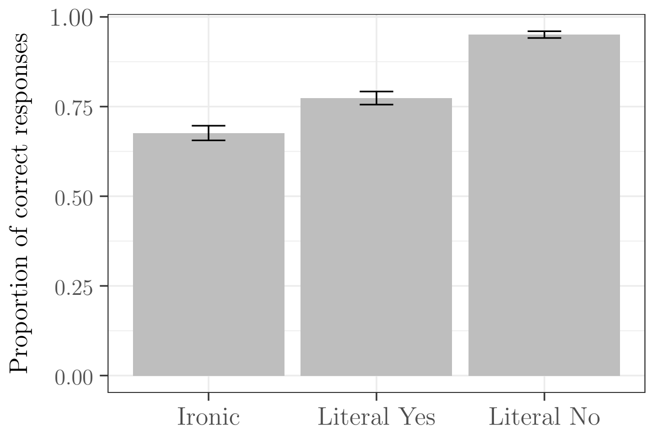

Figure 5: Proportions of correct responses by utterance Type (vertical bars represent standard errors)

Binomial logistic multilevel models, with the by-participant random intercepts were implemented using the glmer function of the lme4 package (Bates et al., 2015). They revealed a significant effect of Type (Ironic vs. Literal Yes vs. Literal No; $\chi^2(2) = 149.62$, $p < 0.001$), as well as of Context ($\chi^2(1) = 8.75$, $p = 0.003$) and of Facial Expression ($\chi^2(1) = 814.85$, $p = 0.001$); there was no effect of Prosody ($p = 0.14$) and of Order ($p = 0.21$). Interactions between Type and Context ($\chi^2(2) = 10.4$, $p < 0.007$) and Type and Face ($\chi^2(3) = 77.85$, $p < 0.001$) were also significant. The model in Table 5 shows that Ironic targets elicit less correct responses than Literal No ones. The presence of Context strongly increases the accuracy on Ironic items. As for Facial expression, it has a detrimental effect on accuracy of

Ironic items; on the contrary, it increases the accuracy of Literal Yes relative to Ironic items.

	Accuracy
Intercept (Ironic)	0.45 (0.20)*
Literal No	2.68 (0.46)***
Literal Yes	0.05 (0.28)
Context	
Ironic X Context	0.86 (0.20)***
Literal No X Context	-0.33 (0.45)
Literal Yes X Context	0.07 (0.24)
Facial expression	
Ironic X Facial expression	-0.44 (0.20)*
Literal No X Facial expression	0.32 (0.41)
Literal Yes X Facial expression	1.98 (0.26)***
Num. obs.	1578

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Coefficients (and standard errors) of fixed effects for the multilevel logistical regression of correct responses on target Type, Type X Context & Type X Facial expression

Reaction times. In each target, the speaker's preference was entirely determined once the word was uttered referring to the object s/he wanted or pretended to want ('Yes, you know how much I like **physics**!'). Reaction times were recorded from the onset of the word referring to an object in the target until participant's response. A negative reaction time means that participant responded before the onset of the target word. Boxplots in Figure 6 suggest longer reaction times for Ironic targets. Linear multilevel regression models, with by-participant random intercepts and random slopes for the Type factor, were implemented with the lmer function of the lme4 package (Bates et al., 2015). They revealed a significant effect of Type ($\chi^2(2) = 31.36$, $p < 0.001$), Prosody ($\chi^2(1) = 11.85$, $p < 0.001$), and Facial Expression ($\chi^2(1) = 5.3$, $p = 0.021$), but not of Context ($p = 0.48$). Interactions between Type and Prosody ($\chi^2(1) = 8.45$, $p = 0.15$) and between Type and Facial expression ($\chi^2(2) = 21.166$, $p < 0.001$) were also significant. Finally, there was also an effect of Order ($\chi^2(1) = 146.87$, $p < 0.001$), but no interaction between Order and Type ($p = 0.61$). The model, displayed in Table 6 reveals that responding to Ironic targets takes longer than for the other two types and that reaction times decrease along experimental trials. Both Prosody and Facial Expression decrease reaction times for Ironic items. (Prosody also decreases the processing of Literal Yes, but increases that of Literal No relative to Ironic items.)

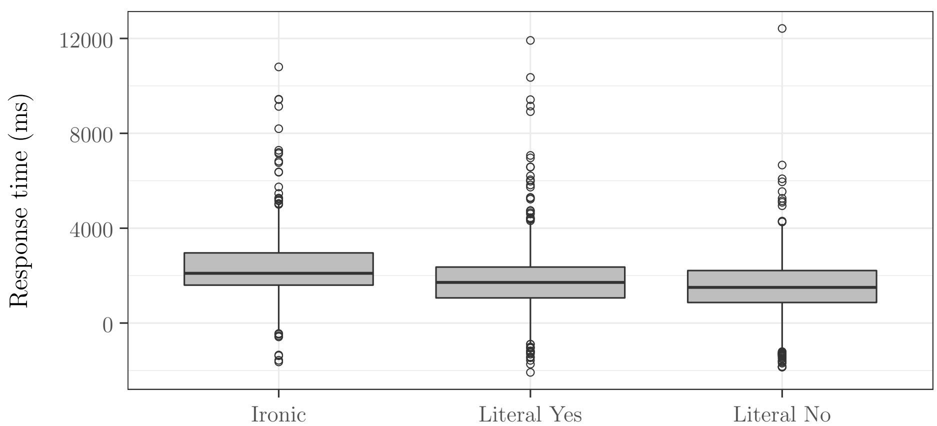

Figure 6: Tukey box-plots for reaction time per target type

Eye-Tracking data. We identified three areas of interest (AOI) for the target sentence segment using Tobii Studio software (version 3.2.1): the speaker's face (eyes plus lips regions), the correct and the incorrect objects (see Figure

	Reaction time
Intercept (Ironic)	3576.93 (220.77)***
Literal No	−1442.15 (174.05)***
Literal Yes	−754.39 (192.42)***
Order	−38.62 (3.07)***
Prosody	
Ironic X Prosody	−564.82 (121.77)***
Literal No X Prosody	216.16 (104.35)*
Literal Yes X Prosody	−321.43 (118.52)**
Facial expression	
Ironic X Facial expression	−457.33 (117.80)***
Literal No X Facial expression	−143.85 (98.26)
Literal Yes X Facial expression	14.45 (112.48)
Num. groups: Subject	44

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Coefficients (and standard errors) of fixed effects of the multilevel linear regression of reaction time on target Type, Half , Type X Prosody & Type X Facial expression

7). The position of the AOI was manually adapted to the movements of the two actors in real time. Eye movement data for the target sentence segment was exported from Tobii using the I-VT fixation filter in the default setting. From the onset of the word referring to the object in the target, and for each AOI, we calculated the total fixation duration (i.e. the sum of the duration for all fixations within an AOI) and the fixation count (i.e. the number of times the participant fixates on an AOI). The fixation count and the total fixation duration were normalised according to participants' reaction times. For instance, number of fixations for the AOI 'correct object' = ([number of fixations on the correct object / time between the beginning of the target word until participant's response] $*$ 1000). We also coded whether participants' first three fixations went to the correct object, both from the onset of the target utterance and from the onset of the word referring to the object in the target.

Figure 7: Areas of interest (AOIs) for the target sentence segment: the speaker's face (eyes plus lips regions), the correct and incorrect objects.

A linear mixed model on proportion of fixation durations with random by-participant intercepts revealed a significant effect of AOI ($\chi^2(2) = 112.77$, $p < 0.001$), as well as a significant AOI X Type interaction ($\chi^2(6) = 24.87$, $p < 0.001$). There was no interaction between AOI and Context, AOI and Prosody, and AOI and Facial expression (all $ps > 0.14$). A linear mixed model on proportion of fixation counts with random by-participant intercepts also revealed a significant effect of AOI ($\chi^2(2) = 69.12$, $p < 0.001$) and a significant AOI and Type ($\chi^2(6) = 23.76$, $p < 0.001$). Additionally, there was a weak interaction between AOI and Context ($\chi^2(3) = 8.69$, $p = 0.034$), but no interaction between AOI and Prosody ($p = 0.88$), between AOI and Facial expression ($p = 0.086$), and between AOI, Type and Context ($p = 0.4$). As can be seen from the model summaries in Table 7, the most relevant result – which is also fairly consistent with accuracy data – from fixation duration and counts is that Literal No targets attracted longer and more numerous fixations to the correct object than Ironic ones.

	Proportions of fixation durations	Proportions of fixation counts
Intercept(Correct Object)	4.96 (2.64)	9.19 (3.04)**
Incorrect Object	−0.51 (2.63)	−3.34 (3.42)
Speaker's face	11.37 (2.28)***	8.94 (2.96)**
Type		
Correct Object X Literal Yes	2.70 (2.64)	2.94 (2.64)
Incorrect Object X Literal Yes	−1.23 (2.64)	−1.19 (2.64)
Speaker's face X Literal Yes	2.06 (1.87)	1.55 (1.87)
Correct Object X Literal No	11.76 (2.66)***	11.92 (2.66)***
Incorrect Object X Literal No	1.21 (2.66)	1.43 (2.66)
Speaker's face X Literal No	3.03 (1.88)	1.73 (1.88)
Correct Object X Context		−3.85 (2.30)
Incorrect Object X Context		−0.46 (2.30)
Speaker's face X Context		−3.93 (1.63)*
Num. obs.	5609	5608

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: Coefficients (and standard errors) of fixed effects of the multilevel linear regression of proportions of fixation durations and fixation counts on AOI, AOI X target Type, & AOI X Context

Turning to the first three fixations on the correct object from the beginning of the target utterance, binomial multilevel models, with the by-participant random intercepts revealed an effect of Type ($\chi^2(2) = 35.56$,

$p < 0.001$), but no effect of Context, Prosody or Facial expression (all $ps > 0.06$). There was also an effect of the Fixation Number (first, second or third; $\chi^2(1) = 32.93$, $p < 0.001$), as well as interaction between Type and Fixation Number ($\chi^2(3) = 36.16$, $p < 0.001$).

For the first three fixations on the correct object, computed from the beginning of the word referring to the object, binomial multilevel models, with the by-participant random intercepts also revealed an effect of Type ($\chi^2(2) = 19.17$, $p < 0.001$) on the fixation on the correct object, but no effect of Context, Prosody or Facial expression (all $ps > 0.21$). Here too, there was an effect of Fixation Number ($\chi^2(1) = 123.02$, $p < 0.001$), as well as interaction between Type and Fixation Number ($\chi^2(3) = 125.03$, $p < 0.001$). As can be seen from Table 8, from the start of the target utterance, Literal Yes – but not Literal No – items attract more anticipatory fixations to the correct object than Ironic. Towards the end of the target utterance, however, Literal No items – but not Literal Yes – are more likely to attract anticipatory looks towards the correct object than Ironic items. For both measures (i.e. target utterance and object mention), the probability to look at the correct object increases from the first to the third fixation for all item types.

Discussion

Correct identification of the speaker's goals is significantly lower for ironic utterances; in that respect, our results parallel findings by Kowatch et al. (2013), who used analogous utterance Types. Equally consistent with the literature (e.g. Kreuz & Roberts, 1995; Gerrig & Goldvarg, 2000) is our result that incongruence with the preceding context increases the correct interpretation of ironic utterances. Interestingly, ironic prosody does not facilitate

	First fixations (target utterance)	First fixations (object mention)
Intercept (Ironic)	$-5.30\ (0.54)^{***}$	$-3.90\ (0.32)^{***}$
Literal Yes	$1.42\ (0.59)^{*}$	$0.39\ (0.41)$
Literal No	$0.03\ (0.67)$	$0.78\ (0.38)^{*}$
Fixation number		
Ironic X Fixation number	$0.58\ (0.22)^{**}$	$0.80\ (0.13)^{***}$
Literal Yes X Fixation number	$0.45\ (0.14)^{***}$	$0.72\ (0.12)^{***}$
Literal No X Fixation number	$0.76\ (0.18)^{***}$	$0.71\ (0.11)^{***}$
Num. obs.	3905	3402
Num. groups: Subject	43	43

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 8: Fixed effects of the multilevel logistic regression of first fixations on the correct object (beginning from start the target utterance and from the start of the word referring to an object) target Type & Type X Fixation Number (first, second or third).

interpretation, and ironic facial expression actually hampers it, confirming that non-contextual cues for irony are not very reliable in a comprehension task.

We also found that ironic items elicit slower reaction times relative to the literal ones. Slower processing of ironic items, to a certain extent at least, is probably linked to the contextual assessment and rejection of the compositional, literal meanings (Giora, 2003). However, Experiment 2 also strongly suggests that context-based processing of irony may be aborted in the presence of a distinctive prosody and/or facial expression. A striking result, which is consistent with our predictions, is that both ironic prosody and ironic facial cues dramatically decrease reaction times. Together, accuracy and reaction times results indicate that non-contextual cues entail a trade-off in irony interpretation. On the one hand, they are less reliable than context, but, on the other hand, their presence prompts a faster processing. One explanation of this effect, in line with the model proposed by Kissine (2016), is that, in the presence of distinctive prosody and/or facial expression participants by-pass contextual interpretation of the literal meaning. Another, very similar interpretation is that both contextual and non-contextual cues are processed in parallel, but that the latter lead to faster decision, thus terminating the former. This finding is coherent with the parallel-constraint-satisfaction account (Katz, 2005; Pexman, 2008), but allows to go a step further in evidencing the relative weight of the different cues in irony comprehension. Note that, unlike us, Kowatch et al. (2013) found no difference in reaction times between ironic and literal items. Recall, however, that their stimuli were all associated with ironic prosody and no contextual cues. To the extent that prosody prompts faster (but less accurate) processing, this feature may explain the difference between their and our results.

Another clear-cut result of Experiment 2 is the advantage in processing for literal ('No') negative sentences. These items led to strikingly higher accuracy rates and faster response times. Likewise, these items were associated with longer and more numerous fixations on the correct object, reflecting lesser hesitation as to the response. Further evidence for the advantage of Literal No items comes from first fixations. At the beginning of the target utterance, more looks go to the correct object in literal positive items, which is certainly due to the spill-over from the mention of the correct object in the preceding question (see Figure 1 and Table 1). However, by the time the object is mentioned in the target, literal negative utterances trigger more anticipatory looks towards the correct object. One reason why Literal No items stand apart could be that in our design all ironic items were associated with positive ('Yes') sentences. However, neither accuracy nor reaction times for negative sentences change across the experiment, as revealed by the absence of the interaction between Type and Order. There is another reason why participants implicitly grasped the unambiguously literal nature of Literal No items. Irony is usually associated with negatively oriented readings of positive literal counterparts, whereas the opposite, ironic positive / literal negative valence is highly atypical (Kreuz & Link, 2002). One may surmise, then, that interpreters' language experience makes them sensitive to irony's negative valence. If so, our participants could rapidly associate negative sentences with a non-ironic interpretation, without necessarily assessing the literal content relative to the context, or, for that matter, processing any other cue. In line with this idea, even though in Experiment 1 negative prosody was the most clearly distinguished from ironic, in Experiment 2 it tended to slow down reaction times. That is, information provided by literal negative prosody is made redundant by the negative valence of the sentence.

Returning to our main research questions, Experiment 2 strongly suggests that, as hypothesised in the Introduction, ironic prosody and facial expression are less reliable cues for irony comprehension than contextual incongruence, but that they entail an accuracy-processing speed trade-off. However, three methodological choices we made could be taken to somehow mitigate these results. First, recall that we decided not to include a condition with Facial expression, but no Prosody and Context. While the results of Experiment 2 indicate that ironic Facial expression is not a fully reliable cue for the interpretation of ironic utterances, this cue was always associated with at least another one. Second, our decision not to include Ironic No items (viz. ironic compliments), albeit fully justified from a methodological point of view (see above), entails an unbalanced experimental design, with more positive (Ironic and Literal Yes) than negative (Literal No) items. Third, we compared the reliability of a prosodic or contextual cue delivered alone or in combination with one or two other cues, but at no time these conditions were compared to a complete absence of cues. Adding a control condition without any cues would allow to appreciate more closely the reliability of isolated cues. Even though none of these three features is likely to impact the differential processing roles of contextual and non-contextual cues uncovered in Experiment 2, we seek to determine, in Experiment 3, whether these effects are robust enough to show up in a fully balanced design.

Experiment 3

In Experiment 3 we seek to replicate the effects uncovered in Experiment 2 using a perfectly balanced, between-subject design, while keeping exactly the same target sentences segments as in Experiments 1 and 2. In order to rule out any potential bias due to the presence of literal negative items, we kept only the Literal Yes and Ironic items of Experiment 2.[5] Using these

[5] In theory, we could have balanced our design by adding negative ironic items. To begin with, such a design would have considerably increased the task duration, and hence the risk of biases due to cognitive fatigue. More importantly, and as already discussed above, negative ironic items are highly atypical (Kreuz & Link, 2002), and poorly comprehended even in discrimination tasks, which arguably are easier than our act-out paradigm (Climie & Pexman, 2008; Filippova & Astington, 2010). Interpretation of this less common type

items, we created two sets of stimuli – Context vs. No Context – presented to two different groups of participants. Stimuli presented to the Context Group included all possible combinations of Context with non-contextual cues: Context; Context and Facial expression; Context and Prosody; and Context, Prosody and Facial expression. Stimuli presented to the No Context Group pooled all combinations of non-contextual cues in the absence of Context: No cue; Prosody; Facial expression; Prosody and Facial expression. In this way, Experiment 3 provides a balanced design suited to isolate the impact of all three cues, and includes a condition with only a facial-expression cue and a control condition without any cues.

This between-subject design also allows to further test our Hypothesis 2. In line with Experiment 2, we expect an overall effect of Group (Context vs. No Context) on accuracy: in the absence of contextual cues, participants should be more error prone in gauging the speaker's ironic intent. If, as we hypothesise, processing of non-contextual cues is privileged at the expense of context-based assessment of the utterance content, the presence of ironic prosody and/or facial expression should lead to comparable accuracy in both Context and No Context groups. Furthermore, we also predict contextual processing not to be completed in the presence of these non-contextual cues. Accordingly, the presence of the non-contextual cues lead to shorter response times in both groups, which would indicate, in the Context group, that contextual processing has been aborted. If, by contrast, non-contextual cues supplement full processing of contextual cues, one should expect ironic

of irony is thus a topic orthogonal to the relative roles of contextual and non-contextual cues, and clearly falls out of the scope of this paper.

prosody and facial expression to increase accuracy in the Context group, and to lead to longer response times.

Participants

Fourty-seven undergraduate students, none of whom took part in Experiments 1 and 2 , participated for monetary reward in Experiment 3. Inclusion criteria were similar to Experiment 2. One participant was excluded from statistical analyses for technical reasons. The Context group (N=23) consisted of 13 women and 10 men between 19 and 29 years ($m = 22.70$; $sd = 2.60$), and the No Context group consisted of 15 women and 8 men between 20 and 28 years ($m = 22.83$; $sd = 2.50$).

Procedure

Two groups of sixteen videos from the previous set of videos were used to form the Context and No Context group. Target sentences in the Context group were always associated with a Contextual cue and were subdivided in 4 categories depending on the presence (+) or the absence (-) of prosody and/or facial expression cues: Context only (C+P-F-), Context & Prosody (C+P+F-), Context & Facial expression (C+P-F+), Context, Prosody & Facial expression (C+P+F+) conditions. In the No Context group, the target sentence was never associated with a Contextual cue; stimuli were also subdivided in 4 categories depending on the presence (+) or absence (-) of prosody and facial expression cues: No cues (C-P-F-), Prosody only (C-P+F-), Facial expression only (C-P-F+), Prosody & Facial expression (C-P+F+) conditions. Each category is composed of 2 Ironic and 2 Literal Yes items (see Table 9).

Group	Context			No Context		
	Context	Prosody	Facial expression	Context	Prosody	Facial expression
	+	-	-	-	-	-
	+	+	-	-	+	-
	+	-	+	-	-	+
	+	+	+	-	+	+

Table 9: Design of Experiment 3

To obtain the items in the No cues (C-P-F-) and in the Facial expression only (C-P-F+) conditions, we used the same videos as in the Context only and Context & Facial expression conditions from the Context group, removing the context segment from the videos.

Results

In order to uncover the roles of ironic prosody and facial expression, as in Experiment 2, we associated each item with binomial Prosody and Facial Expression factor, depending on which cue(s) were associated with the target.

Accuracy. Figure 8 displays the proportions of correct responses by Group, Type and non-contextual cue. Accuracy was analyzed building hierarchical binomial logistic multilevel models, with the by-participant random intercepts, using the glmer function of the lme4 package (Bates et al., 2015). As predicted, there was a significant effect of Group (Context vs. No Context; $\chi^2(1) = 7.94$, $p < 0.005$). There was also an effect of Type (Literal vs. Ironic; $\chi^2(1) = 5.78$, $p < 0.001$), but no Group X Type interaction ($p = 0.42$). There was no effect of Prosody ($p = 0.17$), but an effect of

Facial expression ($\chi^2(1) = 36.2$, $p < 0.001$). The Type X Facial expression interaction was significant ($\chi^2(1) = 8.32$, $p < 0.004$), but not the Group X Facial expression ($p = 0.49$). That is, ironic prosody and facial expression have a comparable effect in both groups, indicating that their presence does not have a cumulative effect on accuracy in the Context group.

In order to assess further our predictions, we conducted post-hoc comparisons of least square-means on the final model. As predicted, overall accuracy is significantly lower in the No Context group ($\beta = -0.53, se = 0.18, p = 0.0036$). Overall accuracy was higher on Ironic than on Literal items ($\beta = 0.6, se = 17, p = 0.0005$). In other terms, in this irony comprehension task, the rate of false alarms exceeds misses.[6]

Reaction times. As in Experiment 2, reaction times were recorded from the onset of the word referring to an object in the target until participant's response. Response times, per Group, Type and non-contextual cue are summarised in Figure 9. Linear multilevel regressions, with by-participant random intercepts revealed no effect of Group or of Type (ps> 0.7). However, there was an effect of Prosody ($\chi^2(1) = 5.38$, $p < 0.02$) and Facial expression ($\chi^2(1) = 5.58$, $p = 0.018$). As predicted, reaction times were shorter in the presence of Prosody ($\beta = -335.1, se = 144, p < 0.02$) and Facial expression ($\beta = -344, se = 145.4, p = 0.018$), across Groups and Type.[7]

[6]The Type X Facial expression interaction it was due to the fact that literal marked Facial expression improved accuracy on Literal items (viz. reduces the rate of false alarms; $\beta = 1.46, se = 0.24, p < 0.0001$).

[7]Using the same method as in Experiment 2 and the same AOIs (see Figure 7), we also analysed total fixation durations and fixation counts. Hierarchical multilevel linear

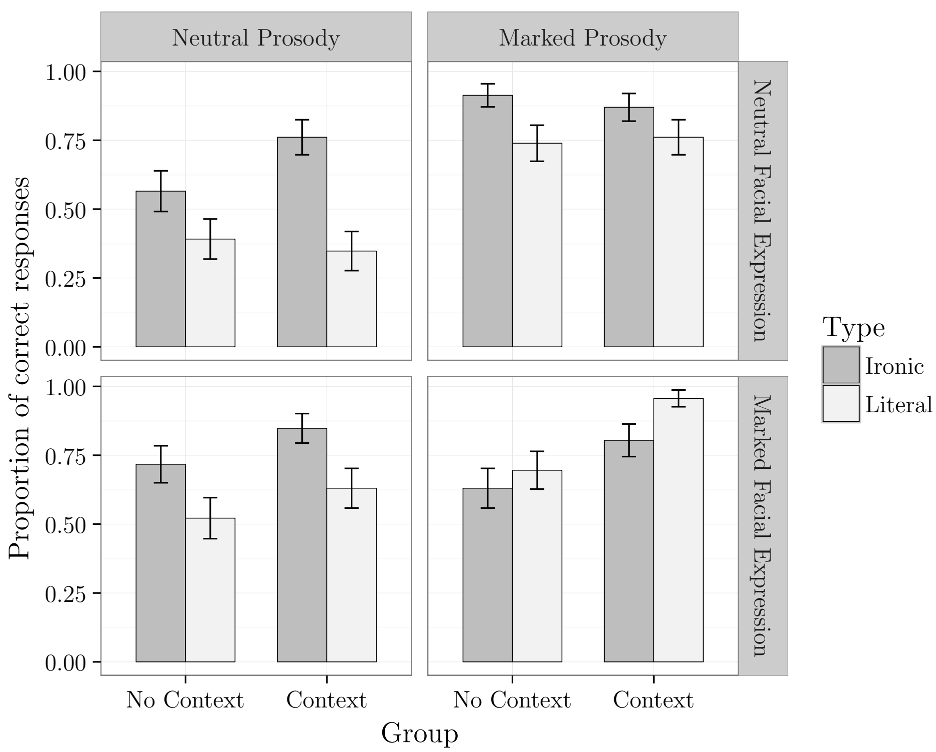

Figure 8: Proportions of correct responses per group, target type and non-contextual cue (vertical bars represent standard errors)

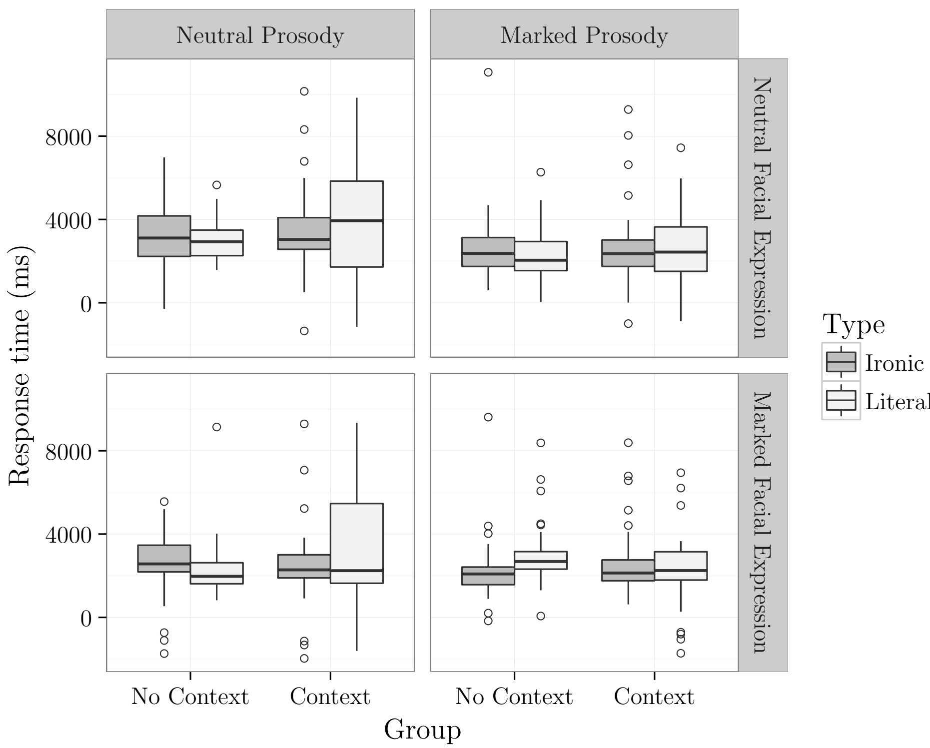

Figure 9: Tukey box-plots for reaction time per group, target type and non-contextual cue

Discussion

Results of our Experiment 3 are entirely consistent with those of Experiment 2, and provide supplementary confirmation for our hypotheses. To begin with, we confirm that contextual incongruence is a much more reliable cue for irony than ironic intonation and facial expression. That is, in line with Experiment 2, high discriminability of these cues, evidenced in Experiment 1, does not translate into comparable reliability in an act-out comprehension task.

Furthermore, in Experiment 3 ironic prosody and/or facial expression do not have a cumulative effect with contextual incongruence; if they did, their presence should have entailed higher accuracy in the context group. This result suggests that, as predicted by our Hypothesis 2, intonation and/or facial expression are salient cues that prompt interpreters to terminate costlier context-based processing. This interpretation is reinforced by the fact that, as in Experiment 2, ironic prosody and facial expression are associated with a processing speed-accuracy trade-off. That is, in spite of being less reliable than contextual incongruence for irony comprehension, the presence of prosody and/or facial expression is associated, in both Context and No Context groups, with shorter response times.

---

regressions with by-participant random intercept revealed an effect of AOI on total fixation durations ($\chi^2(2) = 113.58$, $p < 0.001$). However, there is no interaction with Group, Type, Prosody and Facial expression (all ps> 0.38). As for fixation counts, there was no effect of AOI ($p = 0.16$). These data are orthogonal to the main point of Experiment 3 and will not be discussed further on.

General discussion

While ironic prosody and facial cues can be accurately categorized in a discrimination task (Experiment 1), they do not lead to better grasp of irony in a task where participants must make a decision about the speaker's communicative goals (Experiments 2 and 3). An obvious upshot of our paper, then, is methodological. Researchers should be wary of drawing conclusions about figurative language comprehension based on forced-choice categorization tasks. The asymmetry between discrimination and use is probably due to the fact that perceptual thresholds between ironic vs. non-ironic prosody and facial expression are not entirely clear-cut. This was made clear by the irony ratings in Experiment 1, which showed that literal positive and neutral cues are perceived as more ironic than their negative counterparts. The relative fuzziness of these boundaries has probably less importance in a task where participants have to focus exclusively on locating audio or video stimuli on an irony scale, but they can lead to more incorrect responses when participants have to make decisions on speaker's goals. In other terms, categorization of ironic prosody and facial cues can be carried out off-line, but is much more difficult on-line.

A potential limitation here could be our use of professional actors in the video stimuli. Although it is a standard practice in the literature on irony (e.g. Rockwell, 2000; Anolli et al., 2000; Attardo et al., 2003; Rankin et al., 2009), there is a risk that prosody and facial expression may have been overplayed. Recall, however, that results of Experiment 1 did not show any ceiling effect in rating score of prosody and facial expression, and that they led to far from perfect detection of irony in Experiments 2 and 3. To the best of our knowledge, no study compares prosody and facial expression associated with ironic statements in actors, untrained confederates or in spontaneous speech. One study of acoustic correlates of spontaneous verbal irony reports slower delivery rate as the only robust prosodic characteristic of ironic utterances (Bryant, 2010). A slowdown in speech rate has also been reported in studies using actors (Rockwell, 2000; Anolli et al., 2000), as well as in the current paper. It would be interesting to replicate our findings using recordings of verbal irony in real life situations. However, studying the interplay between ironic cues requires to tightly control the structure of the context segment and the target sentence, which is extremely difficult to achieve in real situations.

The trade-off between accuracy and reaction times, which emerged from Experiments 2 and 3, might look very much like a conceptual conundrum. On the one hand, it seems clear that neither ironic prosody nor ironic facial expression form natural kinds (in line with Bryant & Fox Tree, 2005); on the other hand, participants do seem to privilege such cues, at the expense of accuracy, whenever these are available. On second thought, however, the contradiction is only apparent. Any definition of irony, be it framed in terms of echo or pretence, includes the incompatibility between the context and a literal interpretation of the utterance (Kumon-Nakamura et al., 1995; Wilson, 2006). This is why the capacity to distinguish lies from jokes is operational only if one can make context-based hypotheses about what the speaker wanted the hearer to believe (Wimmer & Leekam, 1991; Martin & McDonald, 2004). It is also for this reason that, as shown by Experiments 2 and 3, assessing the utterance content relative to the background context remains the most reliable route to grasp ironic meanings. In that sense, contextual assessment of the literal meaning is, indeed, an essential part of irony processing, as predicted, for instance, by Giora (2003) and Sperber & Wilson (2002). Yet, even though our results vindicate the central role of context in irony comprehension, they also indicate, consistently with the second prediction made in the Introduction, that irony processing is not always context-based. One may speculate that along with our communicative experience, grows implicit knowledge that ironic utterances are often accompanied by distinctive prosody or facial expression. Mature communicators may then privilege (what they perceive as) ironic prosody and/or facial expression to speed up the comprehension process. That is, unreliable as they are, these non-contextual cues lead to an activation of ironic meanings without the full-fledged, compositional interpretation being completed.

Such a processing route is fully compatible with the Direct Access model (Gibbs, 2002). It can also be implemented within the parallel-constraint-satisfaction model (Katz, 2005; Pexman, 2008), provided that this model is amended in way to allow salient non-contextual cues to terminate context-based processing before it is complete. In a way, then, our findings lay ground for reconciling these two models with more context-based theories of irony. It is generally plausible that frugal heuristics are privileged by interpreters whenever possible (in line with, for instance, Ferreira & Patson 2007; Shintel & Keysar 2009; Epley et al. 2004; Kissine 2016). Assessing the utterance content to the context is a relatively complex, and arguably costly process, so it is not entirely surprising that interpreters forgo it in the presence of more salient cues (as also evidenced by Deliens et al., 2017).

Participants' failure to see that prosody and facial expression are not as reliable as context-based assessment can be profitably conceived of as an instance of meta-cognitive error (in the sense of, e.g., Koriat, 2000; Proust, 2013). According to Kissine (2016), context – understood, this time, as the entire interactional frame of the utterance, including intonation and/or facial expression – plays a two-pronged role in pragmatic processing. On the one hand, it determines the interpretative goal, including, for instance, the level of the specificity of the interpretation output. On the other hand, it is used to monitor and control the interpretation process that has been determined by this goal. For instance, the interpretative goal in Experiment 1 consists in mere discrimination of an ironic or not character of a stimulus, which is less complex than genuinely accessing the speaker's intention, as in Experiments 2 and 3. Such a superficial ironic interpretation may thus be reached without attempting to assess the speaker's intentions. However, non-contextual processes are less reliable to achieve the more complex interpretation goals mandated by the tasks in Experiments 2 and 3. Relying on ironic prosody and/or facial expression in these cases, at the expense of context-based processing, thus reflects a meta-cognitive bias, driven by cognitive economy principles, which leads participants to select an interpretation process less than optimally suited for the interpretative goal at hand.

Acknowledgements

We are grateful to the journal editor and to the three anonymous reviewers for their detailed and constructive remarks on a previous version of this paper. We also thank the *Jean Rage* theatre company for acting in the video stimuli, and Alba Leszczynski for her help with data collection. At the time of writing, Gaétane Deliens's research was supported by a grant from the Fondation Jean-François Peterbroeck, Kyriakos Antoniou was a Wiener-Anspach Foundation post-doctoral fellow at the Université libre de Bruxelles. and Ekaterina Ostaschenko and Fanny Papastamou were supported by the Research Incenstive F.R.S.-FNRS Grant F.4502.15 to Mikhail Kissine. Elise Clin is a FRESH doctoral fellow at the F.R.S.-FNRS.

References

Akimoto, Y., Miyazawa, S., & Muramoto, T. (2012). Comprehension processes of verbal irony: The effects of salience, egocentric context, and allocentric Theory of Mind. *Metaphor and Symbol*, *27*, 217–242.

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

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, *16*, 243–260.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bignell, S., & Cain, K. (2007). Pragmatic aspects of communication and language comprehension in groups of children differentiated by teacher ratings of inattention and hyperactivity. *British Journal of Developmental Psychology*, *25*, 499–512.

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer. Version 6.0.28*. http://www.praat.org/.

Bryant, G. A. (2010). Prosodic contrasts in ironic speech. *Discourse Processes*, *47*, 545–566.

Bryant, G. A. (2012). Is verbal irony special? *Language and Linguistics Compass*, *6*, 673–685.

Bryant, G. A., & Fox Tree, J. E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, *17*, 99–119.

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and speech*, *48*, 257–277.

Caillies, S., Bertot, V., Motte, J., Raynaud, C., & Abely, M. (2014). Social cognition in ADHD: irony understanding and recursive theory of mind. *Research in Developmental Disabilities*, *35*, 3191–3198.

Cheang, H. S., & Pell, M. D. (2011). Recognizing sarcasm without language: A cross-linguistic study of english and cantonese. *Pragmatics & Cognition*, *19*, 203–223.

Chevallier, C., Noveck, I., Happé, F. G. E., & Wilson, D. (2011). What's in a voice? prosody as a test case for the theory of mind account of autism. *Neuropsychologia*, *49*, 507–517.

Christensen, R. H. B. (2015). ordinal—regression models for ordinal data. R package version 2015.6-28. http://www.cran.r-project.org/package=ordinal/.

Climie, E. A., & Pexman, P. M. (2008). Eye gaze provides a window on children's understanding of verbal irony. *Journal of Cognition and Development*, *9*, 257–285.

Colich, N. L., Wang, A.-T., Rudie, J. D., Hernandez, L. M., Bookheimer, S. Y., & Dapretto, M. (2012). Atypical neural processing of ironic and sincere remarks in children and adolescent with autism spectrum disorders. *Metaphor and Symbol*, *27*, 70–92.

Deliens, G., Antoniu, K., Clin, E., & Kissine, M. (2017). Perspective-taking and frugal strategies: Evidence from sarcasm detection. *Journal of Pragmatics*, *119*, 33–45.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327–39.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*, 71–83.

Filippova, E., & Astington, J. W. (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Development*, *81*, 913–928.

Gerrig, R. J., & Goldvarg, Y. (2000). Additive effects in the perception of sarcasm: Situational disparity and echoic mention. *Metaphor and Symbol*, *15*, 197–208.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*, 5–27.

Gibbs, R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*, 457 – 486.

Gibbs, R. W., & Colston, H. L. (2012). *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

Giora, R. (2003). *On our Mind. Salience, context, and figurative language*. Oxford: Oxford University Press.

Giora, R., Givoni, S., & Fein, O. (2015). Defaultness reigns: The case of sarcasm. *Metaphor and Symbol*, *30*, 290–313.

Jacob, H., Kreifelts, B., Nizielski, S., Schutz, A., & Wildgruber, D. (2016). Effects of emotional intelligence on the impression of irony created by the mismatch between verbal and nonverbal cues. *PLoS ONE*, *11*, e0163211.

Katz, A. N. (2005). Discourse and sociocultural factors in understanding nonliteral language. In H. Colston, & A. N. Katz (Eds.) *Figurative language comprehension: Social and cultural influences*, (pp. 183–207). Mahwah, NJ: Erlbaum.

Kissine, M. (2016). Pragmatics as metacognitive control. *Frontiers in Psychology*, *6*.

Koriat, A. (2000). The Feeling of Knowing: Some Metatheoretical Implications for Consciousness and Control. *Consciousness and Cognition*, *9*, 149–171.

Kowatch, K., Whalen, J. M., & Pexman, P. M. (2013). Irony comprehension in action: A new test of processing for verbal irony. *Discourse Processes*, *50*, 301–315.

Kreuz, R. J. (2000). The production and processing of verbal irony. *Metaphor and Symbol*, *15*, 99–107.

Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, *21*, 127–143.

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and Symbol*, *10*, 21–31.

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*, 3–21.

Lenth, R. (2016). Least-squares means: The r package lsmeans. *Journal of Statistical Software*, *69*, 1–33.

Lœvenbruck, H., Ben Jannet, M. A., D'Imperio, M., Spini, M., & Champagne-Lavau, M. (2013). Prosodic cues of sarcastic speech in french: slower, higher, wider. In *14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, (pp. 3537–3541).

Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*, 311–328.

McDonald, S. (2000). Neuropsychological studies of sarcasm. *Metaphor and Symbol*, *15*, 85–98.

McDonald, S., & Pearce, S. (1996). Clinical insights into pragmatic theory: Frontal lobe deficits and sarcasm. *Brain Lang*, *53*, 81–104.

Pexman, P. M. (2008). It's fascinating research the cognition of verbal irony. *Current Directions in Psychological Science*, *17*, 286–290.

Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. Oxford: Oxford University Press.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/

Rankin, K. P., Salazar, A., Gorno-Tempini, M. L., Sollberger, M., Wilson, S. M., Pavlic, D., Stanley, C. M., Glenn, S., Weiner, M. W., & Miller, B. L. (2009). Detecting sarcasm from paralinguistic cues: anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage*, *47*, 2005–15.

Rockwell, P. (2000). Lower, slower, louder. vocal cues of sarcasm. *Journal of Psycholinguistic Research*, *29*, 483–495.

Shintel, H., & Keysar, B. (2009). Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science*, *1*, 260–273.

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

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language*, *17*, 3–23.

Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of Experimental Psychology. General*, *143*, 1649–1665.

Staikova, E., Gomes, H., Tartter, V., McCabe, A., & Halperin, J. M. (2013). Pragmatic deficits and social impairment in children with adhd. *Journal of Child Psychology and Psychiatry*, *54*, 1275–1283.

Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, *25*, 701–721.

Voyer, D., Thibodeau, S.-H., & Delong, B. J. (2016). Context, contrast, and tone of voice in auditory sarcasm perception. *Journal of Psycholinguistic Research*, *45*, 29–53.

Wilson, D. (2006). The pragmatics of verbal irony: echo or pretence? *Lingua*, *116*, 1722–1743.

Wimmer, E., & Leekam, S. R. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, *9*, 257–270.