When irony is faster than its literal control: The role of mindreading during irony comprehension

Camilo R. Ronderos[1,2], John M. Tomlinson. Jr.[3], & Ira Noveck[4]

[1] University of Oslo, Department of Philosophy, Classics, History of Art and Ideas

[2] Humboldt-Universität zu Berlin, Institut für deutsche Sprache und Linguistik

[4] Laboratoire de Linguistique Formelle, UMR 7110 -Université de Paris & CNRS

[3] Leibniz-ZAS Berlin

Author Note

The authors declare no competing interests. All data, pre-registration forms and analysis scripts are available on the associated OSF project: https://osf.io/vgkst/

Correspondence concerning this article should be addressed to Camilo R. Ronderos, Blindernveien 31 Georg Morgenstiernes hus, room 533, Oslo. E-mail: camilorr@uio.no

Abstract

Irony is a heavily context-dependent pragmatic phenomenon. But what is it about context that facilitates or blocks irony comprehension? Based on the echoic account, we suggest that a context facilitates irony comprehension when it makes manifest a speaker's intentions and attitude, i.e., when a context makes it easy for participants to engage their mindreading abilities. In two pre-registered self-paced reading experiments, we investigated the comprehension of sentences in English that could be understood as ironic or literal, according to the story frame that participants read leading to the target sentence. In Experiment 1, we found that when the story frames prevent participants from anticipating the speaker's intention, literal readings of critical sentences are - not surprisingly - faster than ironic ones. Importantly, when the story frames gave access to the speaker's intentions, we find cases in which ironic readings are actually faster than literal ones, resulting in a novel finding for the irony comprehension literature. Further, when the speaker was described as having a sincere attitude towards their utterance, participants tended to understand the utterances literally. They tended to understand them ironically when it was not clear what the speaker's attitude was. In Experiment 2 we investigated whether the findings of Experiment 1 could be linked to individual differences in participants' mindreading abilities. We found that participants who scored higher on a standard Theory of Mind task (the 'Reading the mind in the Eyes' task) were significantly more likely to derive ironic - but not literal - interpretations. We see these results as supporting the echoic account of irony comprehension. This work discusses the relevance of our findings to the long-standing debate on the processing effort of ironic vs. literal sentences.

Keywords: irony comprehension, mindreading, echoic account, Theory of Mind, experimental pragmatics, figurative language

Word count: 8944

When irony is faster than its literal control: The role of mindreading during irony comprehension

Introduction

It is often the case that people mean something very different from what they actually say. Take the conversation in (1):

(1) (A) Chris: Sorry, could my daughter play this guitar?

(B) Music store owner: Yes, this guitar is here for everyone to play with.

When taken literally, (1B) could just be a sincere answer to a polar (yes-no) question. However, if the store owner sees that Chris's daughter is a small child and knows that the guitar is incredibly expensive, he might actually mean something quite different. The store owner might wish to convey that the question is ridiculous and by no means can Chris's daughter play the guitar. This would be an instance of **verbal irony**: a language strategy through which an indirect, evaluative utterance is communicated with a proposition that stands in some type of opposition to the speaker's intentions (see Bryant, 2012; Pexman, 2008).

It is clear that one must go beyond the literal meaning of (1B) to understand it ironically. But what exactly does a comprehender need to do for this to happen? One well known approach to irony interpretation, *the echoic account* (Jorgensen, Miller, & Sperber, 1984; Sperber & Wilson, 1981; Wilson & Sperber, 1992, 2012), sees irony as a type of attributive use of language. When using irony, a speaker does not endorse their own utterance, but instead implicitly attributes it to someone else or to some normative expectation. This amounts to expressing a dissociative attitude towards the belief articulated in the utterance. This analysis suggests that to understand irony, the comprehender must ultimately accomplish two things: (i) gain access to the speaker's informative intention (*what* it is that the speaker wants to convey) and; (ii) detect the speaker's attitude towards their own proposition (i.e., to capture *that* the speaker wants their

audience to get the informative intention through an ironic attitude). These two features combined illustrate that irony comprehension crucially involves a form of reasoning about mental states that allows a comprehender to interpret a speaker's behavior. Such reasoning generally falls under the umbrella term of mindreading (Nichols & Stich, 2003; Spaulding, 2020) or Theory of Mind (Baron-Cohen, Leslie, & Frith, 1985).

Prior tests of the echoic account have shown that listeners can more readily process ironic utterances if there is an explicit echo in the discourse context (Gibbs, 1986; Jorgensen, Miller, & Sperber, 1984; Ţurcan & Filik, 2017). What is not known, however, is how the two previously mentioned types of mindreading skills - considering a speaker's informative intention and the speaker's attitude towards their proposition - affect irony processing. If a context facilitates these types of reasoning, will irony be more readily understood? Besides serving as a test of the echoic account, answering this question puts one in a position to address one of the longest-running debates in the processing literature on irony: Are ironic sentences harder to process than their literal controls?

In this work, we propose that mindreading can have a *variable* effect on ironic readings of utterances with respect to literal readings of one and the same sentence. In what follows, we first review the evidence linking mindreading to irony comprehension. We then discuss the psycholinguistic findings that investigate the processing effort of ironic, relative to literal, utterances. Then, we present our two experiments and discuss them in the light of the issues raised in the Introduction.

Irony and mindreading

According to the echoic account, when an addressee successfully understands an ironic utterance, they understand that the speaker is attributing this utterance or thought to someone else while simultaneously expressing a dissociative attitude towards it. This means that irony comprehension should require the ability to generate second-order metarepresentations (e.g., a thought about a thought, as in the sentence *Miguel thinks that*

Luisa is upset that Paula is leaving). This ability is believed to be an integral part of adult mindreading skills (Allott, 2017; Sperber & Wilson, 2002, i.a.).

Previous studies have shown that difficulties in generating second-order metarepresentations (due to either brain lesions or atypical neurological development) correlate with difficulties in understanding irony, but not with difficulties in understanding literal language (F. G. E. Happé, 1993; McDonald, 2000). It has also been shown that brain regions typically associated with mindreading activity display increased activation during an irony comprehension task relative to literal controls (Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012). These studies are in line with the echoic account's predictions regarding the involvement of mindreading in irony comprehension (for a summary, see Noveck, 2018). However, it is unclear from these studies whether this involvement is binary (you either have it or you don't) or whether mindreading can have a graded effect on comprehension.

Spotorno and Noveck (2014) were the first to demonstrate that engaging in mindreading skills during irony comprehension is arguably a matter of degree: Comprehenders can be shown to progressively anticipate mindreading situations over the course of an experiment and, as a result, understand irony more readily (as measured by reduced reading times for ironic utterances) by the end of an experimental session. It is important to note that, though this finding suggests an involvement of mindreading, it differs from what the echoic account would predict in a critical way. While Spotorno and Noveck (2014) showed that comprehenders' processing effort can be reduced through repeated encounters of irony, the echoic account would state that the processing effort of a single ironic sentence will depend on that sentence's communicative context, not on whether different ironic sentences have been encountered before. In other words, the echoic account does not state that people engage in mindreading to anticipate 'irony' as a trope. Comprehenders do so to anticipate the beliefs and intentions of the speaker that lay behind a single ironic utterance. A different type of contextual manipulation is therefore needed to investigate whether mindreading has a variable effect on processing irony in an individual trial.

Processing effort of ironic vs. literal utterances

If, as we hypothesize, mindreading can have a variable effect on irony comprehension, this should be reflected during online processing. If comprehenders have strong evidence as to the nature of the speaker's intentions and beliefs, it should be easier for them to understand irony compared to comparable cases that do not provide such evidence. Behaviorally, this should reveal that the processing effort of irony could in fact be more efficient than that of literal readings under certain mindreading-related conditions. For example, a sentence understood ironically in a context rich in mindreading-facilitating cues should be more readily understood than a sentence understood literally in a context deprived of evidence pointing to the speaker's intentions and beliefs. In other words, processing effort of ironic (and literal) utterances is constrained by a comprehender's expectations of the speaker's intended meaning (see Degen & Tanenhaus, 2019 for a related argument).

Investigating the variable role of mindreading during irony comprehension therefore bears on one of the central topics in irony research: The debate on processing effort of ironic relative to literal language. Broadly speaking, there are two camps in this debate. On one side sits the *contextualist* camp, which states that context influences processing such that an ironic sentence can be understood just as easily as a literal equivalent. This view has its origins in Gibbs (1986), who claimed that understanding irony could happen 'directly' without first deriving a literal interpretation, in opposition to previous accounts (Grice, 1989; Searle, 1979). Gibbs's approach, known as the Direct Access view (Gibbs, 1994, 2002), is complemented by the Constraint Satisfaction view (Pexman, 2008; Pexman, Ferretti, & Katz, 2000), which states that multiple factors can influence processing of irony in parallel, often resulting in ironic sentences being understood just as fast as their literal counterparts (Ivanko & Pexman, 2003; Katz, Blasko, & Kazmerski, 2004).

On the other side sits the *context-independence* camp, which states that ironic sentences typically require more processing resources than their literal counterparts regardless of context (Giora & Fein, 1999; Giora et al., 2007; Giora, Fein, & Schwartz, 1998;

Schwoebel, Dews, Winner, & Srinivas, 2000). A prominent representative of this view is the Graded Salience Hypothesis (Giora, 2003). It states that for any utterance, the most salient meaning will be processed by default. While salience is determined by the utterance's familiarity, stereotypicality, prototypicality, and frequency (among other factors), the most salient interpretation usually coincides with the utterance's literal meaning. If a salient meaning is found to be incompatible with context, a secondary, non-salient meaning is computed. Irony is normally non-salient (but see Giora, Drucker, Fein, & Mendelson, 2015 for some exceptions), so it is only understood after deriving the literal meaning, resulting in more processing effort compared to that of understanding a literal utterance (Filik, Howman, Ralph-Nearman, & Giora, 2018; Filik & Moxey, 2010; Giora et al., 2007).

Tests of these accounts have usually consisted in looking for contextual cues that may or may not ease comprehension of an ironic relative to a literal utterance. This has resulted in conflicting evidence (e.g. Filik & Moxey, 2010; Ivanko & Pexman, 2003; Katz, Blasko, & Kazmerski, 2004; Schwoebel, Dews, Winner, & Srinivas, 2000). Some elements of context - such as the presence of an 'echoed' antecedent (Ţurcan & Filik, 2017), explicitly introducing a character as sarcastic (Ţurcan, Howman, & Filik, 2020), or the association of one character with sarcasm throughout an entire experiment (Regel, Coulson, & Gunter, 2010) - seem to facilitate processing. While others - such as the presence of a previous sarcastic utterance by the speaker (Giora et al., 2007) or explicit mention of the speaker's expectations (Ţurcan & Filik, 2016) - do not.

Given the current state of the debate, there is no unified account whose predictions adequately explain these empirical findings. We offer a different approach. Instead of focusing on the specific features of a context that might speed up processing, we examine the effect of context only in as much as it can help participants in an experiment anticipate the speaker's informative intention as well as the speaker's attitude. In other words, we suggest that examining two ways in which participants are encouraged to engage in mindreading abilities will help us understand how the processing effort linked to irony varies, relative to

the processing effort linked to literal readings.

The variable effect of mindreading on irony comprehension

Let us revisit example (1). In the event that we know more about the store owner – e.g., that he has no intention of letting a young girl play with a very expensive guitar – the reader will likely expect the store owner's answer to be 'no.' The interpretation – and processing effort – of (1B) will thus be determined by how strong such expectations are. Knowing the store owner's attitude when he speaks is also a cue to an ironic reading. A reader who is further told that the speaker has a tendency to speak insincerely (e.g. jokingly) is more likely to read (1B) ironically. In short, the more strongly that a comprehender believes to know the store owner's intentions and attitude, the easier it should be to interpret (1B) as ironic or not.

This leads to the goal of the current study. Based on the predictions of the echoic account, we investigate how processing effort of irony varies as a function of features related to irony and mindreading. Concretely, we test the following two hypotheses. First, we hypothesize that a context facilitates irony comprehension when it provides comprehenders with a deeper understanding of a speaker's intention as described through a story frame and by giving explicit information about a speaker's attitude. If this hypothesis is on the right track, one should be able to manipulate such anticipations in such a way that ironic (as well as sincere) readings of identical sentences can be equally facilitated. In the event that mindreading-rich contexts do not facilitate the processing of ironic readings compared to mindreading-poor contexts, it would speak against the echoic account and offer support to views that see irony comprehension as a generally more effortful process than understanding literal sentences, regardless of contextual bias (e.g., Giora et al., 2007; Schwoebel, Dews, Winner, & Srinivas, 2000). Second, we hypothesize that, if a facilitatory effect of context is in fact linked to mindreading, it should be more pronounced for comprehenders who are particularly apt at using their mindreading abilities relative to those who are less so.

Alternatively, if there is no connection between individual differences in mindreading and an effect of context, it would suggest that the way in which comprehenders integrate contextual cues with an utterance during irony interpretation does not necessarily require reasoning about a speaker's informative intention and attitude towards their own utterance.

Concretely, we first validated a narrative context that can lead to either an ironic or a literal reading of a sentence while allowing for reading times measures and comprehension questions (Experiment 1). We show that, when context generates strong expectations regarding the speaker's intentions (operationalized as an expected answer to a polar question), understanding irony can be just as fast as - or even faster than - understanding a literal reading of the same sentence. In Experiment 2, we additionally show that individual differences among participants - with respect to their mindreading abilities - can account for differences in irony comprehension. These results provide empirical support for the echoic account and help explain the oft-reported variations in the literature with respect to the processing effort of ironic readings of sentences relative to literal ones.

Experiment 1

With the idea of testing the echoic account of irony comprehension, we set up story frames in such a way that a speaker's intention can be understood by the reader through two channels: (i) by providing information about a speaker's informative intention with respect to their audience (in the story) and; (ii) by providing information to the reader about the speaker's attitude towards his or her own upcoming utterance. When this information is not available, irony comprehension should not be facilitated. Let us consider these two pieces of information in turn.

The first variable concerns the expectations that a reader is induced to have with respect to the eventual speaker through the story situation. This can be illustrated again through our opening example. In the 'strong expectation' context of the Guitar story (see Figure 1), the reader is encouraged to expect the store owner to not agree to Chris's request.

This occurs through various pieces of information in the context, such as i) indicating that the guitar is the most valuable in his shop, ii) explaining that the person who would handle the guitar is a five-year-old, and iii) that the store owner dislikes children. Note that in the Neutral condition, there are no such statements that serve as cues to the eventual speaker's state of mind.

The second variable concerns explicit information about the speaker's attitude, which indicates that the speaker's upcoming utterance is dissociated (insincere in some way) or sincere. In the conditions that encourage dissociated attitudes towards the speaker's upcoming utterance, readers will encounter statements such as *the owner has a reputation for being a jokester, he therefore replies:* just before reading the speaker's actual utterance. In the sincere conditions, which encourage readers to take the upcoming speaker's utterance at face value, readers receive statements such as *the owner has a reputation for being frank, he therefore replies:* as a lead up to the utterance.

As can be seen, the current design ultimately depends on a critical polar question. Polar questions were chosen because they typically allow for two possible answers: 'yes' or 'no.' As far as irony inducing readings go, the polar question is useful because it arrives at a moment in which readers can determine a) that the eventual speaker is likely to not comply with the request (this is the strong expectation context) and that b) the eventual speaker will reply with a dissociated attitude. As far as literal inducing readings go, there is little intention-revealing information provided (this is the neutral context) and the eventual speaker is described as speaking sincerely. With this design, the speed of comprehending the target utterance can conceivably be fastest under conditions that optimize irony understanding. For completeness, these two features are manipulated as part of a 2 x 2 design.

The current manipulation allows us to do two things. First, we can investigate the effect of mindreading on irony comprehension on a trial-by-trial basis. Second, it will put us in a position to directly determine whether the effect of mindreading can account for

differences in processing effort of ironic sentences relative to literal sentences.

Participants and power analysis

We wanted to determine the minimum number of participants that would allow us to detect a true effect (more conservative in size than that found in the pilot, see supplementary materials) with at least 80% power. To do this, we used the model parameters from the analysis of the pilot study (i.e., the linear, mixed-effects model of the log-transformed reaction times). These models had the following maximally-converging random effects structure: The sum-contrast coded model that tested the interaction between both factors included random intercepts by items and by participants. It also included random slopes for both factors and their interaction by items. The sliding-contrast coded model included random intercepts and slopes by items and random slopes by participants. This information can be found in detail in the corresponding R script found on the project's OSF repository.

Crucially, we changed the estimated model coefficients for considerably more conservative ones: We settled on an effect size with a Cohen's d value of 0.2 for all effects. This is a more conservative estimate for every effect found in the pilot study and is commonly used as a benchmark number for a 'small' effect size in psychological research (Cohen, 1992). For the interaction effect, we settled on an effect size of half the size of the effect found in the pilot study. Table 3 below summarizes the size of the relevant effects found in the pilot study, the corresponding effect size used for computing power, the estimated statistical power for finding such an effect with 220 participants, and the actual effect found in Experiment 1. To estimate statistical power, we used a simulations-approach via the R package SimR (Green & MacLeod, 2016). We simulated the results of 1000 experiments (for every relevant effect) assuming the effect size shown in Table 1. We then counted the number of experiments that found a significant effect, and used this number to estimate power.

Table 1

Effect sizes computed for the power analysis of Experiment 1. Effect sizes are given in Cohen's D.

Comparison	**Effect size found in Pilot 1**	**Assumed effect size for simulations**	**Statistical power with 220 participants**	**Effect size found in Experiment 1**
ATTITUDE*BIAS Interaction	0.397	0.2	86.4%	0.269
Sinc./neg. v. Insinc./neg.	0.389	0.2	89.1%	0.42
Insinc./neg. v. Insinc./neutral	0.6	0.2	81.0%	0.228
Sinc./neutral v. Insinc./neutral	0.4	0.2	87.0%	0.2

Participants recruited for the Experiment were right-handed, native speakers of American English between the ages of 18-35. In anticipation that some participants would not meet the exclusion criterion (correctly answering at least 5 out of 7 filler comprehension questions), we recruited a total of 319 participants via the online platform Prolific. Of these 319, 57 (i.e., 17%) did not meet the inclusion criterion and were removed from the analysis, leaving the final number at 262. Participants gave their informed consent and received monetary compensation for their participation.

Design

Experiments 1 and 2 were programmed using the Ibex experimental software (Drummond, 2013) coupled with the PennController (Zehr & Schwarz, 2018) and run via the internet. Experiment 1 had a 2X2 design with the factors SPEAKER INTENTION ('neutral' vs. 'strong expectation') and SPEAKER ATTITUDE ('sincere' vs. 'insincere'). All manipulations refer to the type of contextual information that participants read prior to the target utterance, which was always identical in every condition. Again, Figure 1 shows an example critical item.

There were a total of nine critical items. For every participant, a new list was automatically created showing only one out of the 4 possible versions of each critical item (using Ibex's built-in Latin-square design function). Because we had an odd number of 9 items, each participant saw 2 instances of three of the conditions and 3 instances of one of the conditions. The condition for which participants saw one additional instance was counterbalanced across participants. Participants also saw ten filler items. We settled on this number of items for two reasons. First, since the experiment was to be web-based, it was important to keep the task as short as possible to maintain participants' attention and minimize noise, following Futrell (2012). Second, we wanted to avoid any potential trial effects, which have consistently been found to interact with processing effort of ironic relative to literal sentences (Olkoniemi, Ranta, & Kaakinen, 2016; e.g., Spotorno & Noveck, 2014). Despite this low number of items, our a-priori power analysis showed that Experiment 1 was sufficiently powered to find a true interaction effect (smaller than the one we actually found) between SPEAKER INTENTION and SPEAKER ATTITUDE (see supplementary materials for details on the power analysis).

There were comprehension questions after each critical item and after 7 out of the 10 filler trials. The critical comprehension questions assessed whether participants understood the sentence ironically or literally (see Figure 1). The filler questions determined if participants were included in the analysis or not: they had to answer at least 5 of the 7 filler

questions correctly. Filler and critical trials were pseudo-randomized, so that there would be at least one filler trial between every critical trial.

Factor 1: SPEAKER INTENTION	
Strong expectation	**Neutral**
Chris wants to buy his five-year-old daughter her first guitar. They go to a professional music shop together and she heads for the oldest and most valuable guitar in the store, which was behind a protective glass case. As she comes closer, one can see that the guitar is twice her size. The owner of the store, who really hates children, sees this and anxiously walks towards them. Chris sees him and says: "Sorry, could my daughter play this guitar?	*Chris wants to buy his 15-year-old daughter a new guitar, so they go to a music shop together. She is overwhelmed by all the different types of guitars they have, so she doesn't know which one to pick. They browse around for a while, and finally she finds one that she really likes, even though Chris doesn't understand why. He starts looking for the owner to ask him about it. Chris sees him and says: "Sorry, could my daughter play this guitar?"*
Factor 2: SPEAKER ATTITUDE	
Insincere Attitude	**Sincere Attitude**
The owner has a reputation for being a jokester. He therefore replies:	*The owner has a reputation for being frank. He therefore replies:*
Target Sentence	
"Yes, this guitar is here for everyone to play with"	
Wrap-up Sentence	
There were many other costumers in the store that day.	
Comprehension question and possible answers	
The owner will:	
***(A)** not let her play the guitar*	***(B)** let her play the guitar* ***(C)** Buy a guitar*

Figure 1. Example of a target utterance in Experiment 1 in the four conditions resulting from crossing the factors SPEAKER INTENTION and SPEAKER ATTITUDE. Note that Experiment 2 only had two conditions: 'strong expectation-insincere attitude' and 'neutral-sincere attitude'

Materials

Each critical item consisted of 8 sentences. The first five sentences set up expectations regarding the answer to the upcoming polar question (again, see Figure 1): Participants should strongly expect a 'no' answer, or not expect any particular answer whatsoever. These expectations were normed in a separate rating experiment, which is reported in the supplementary materials (https://osf.io/vgkst/). After these five sentences, participants read three additional ones: (1) a sentence that conveyed the attitude of the speaker and how it relates to the upcoming target sentence ('sincere attitude' or 'insincere attitude' conditions), (2) a target sentence that was always a 'yes' response and was identical across conditions,

and (3) a final wrap-up statement identical across conditions.

After each critical trial, participants chose one of three possible answers from a multiple-choice question regarding the outcome of the situation. Their choice indicated whether they constructed an ironic interpretation, a literal interpretation, or whether they misunderstood the story altogether (i.e., a 'distractor' answer) (answers A, B, and C in Figure 1 respectively). Position of the answers was randomized across trials.

Procedure

At the beginning of the experiment, participants were told that they were going to read normal, every-day conversations and that they should imagine how these conversations would play out in real life. They were not told that any of the exchanges were going to be ironic. After completing two practice trials, the experiment began. Participants read all stories on a sentence-by-sentence basis and hit the space-bar to reveal the next sentence. When doing so, the previous sentence was replaced by dashed lines. For the comprehension questions, participants could either use their keyboard (by pressing the numbers 1-3) or their mouse to select one of the three possible answers. Participants took 9 minutes on average to complete the Experiment.

Predictions

The predictions for both Experiment 1 and Experiment 2 were pre-registered. The pre-registrations - along with all materials from both experiments, data and analysis scripts - can be found on the project's OSF page: https://osf.io/329cs

Comprehension questions. We reasoned that participants should be able to use explicit information about a speaker's attitude to understand whether an utterance is literal or ironic. We therefore predicted that there should be a main effect of SPEAKER ATTITUDE on comprehension, with items in the 'insincere' condition being taken as ironic and those in the 'sincere' conditions as literal.

Reading times. If the type of context we created mediates processing effort of irony, we should see that participants take less time reading ironic sentences (i.e. what we predict to be sentences in the 'insincere' conditions) when there is a strong expectation compared to when there is no expectation in particular. This should translate to a significant difference in reading times between the 'strong expectation-insincere attitude' and the 'neutral context-insincere attitude' conditions. Further, we predicted the opposite pattern for literal sentences (the 'sincere' conditions): When participants expect the speaker to be sincere, they should struggle processing a 'yes' response when they strongly expected a 'no,' whereas they should have no difficulty reading the 'yes' response when they are not expecting any particular answer (or arguably a 'yes' response by default). These differences should result in a significant interaction between the two factors (SPEAKER INTENTION and SPEAKER ATTITUDE).

Analysis and results

As a reminder, data from participants who did not answer at least 5 out of 7 of the filler comprehension questions correctly were discarded, resulting in the exclusion of 57 participants (17%). Trials in which participants selected the distractor response (answer 'C' in Figure 1), were also discarded. This led to removing 3.4% of critical trials.

All remaining data was analyzed using the Lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2020). The data and analysis script for Experiment 1 are available on the project's OSF repository: https://osf.io/vgkst/. Models were fitted following the recommendations of Barr, Levy, Scheepers, and Tily (2013). They included random intercepts and slopes by items and participants for SPEAKER INTENTION, SPEAKER ATTITUDE and their interaction, but excluded the random correlation between intercept and slopes by participants.

Comprehension questions. Panel A of Figure 2 shows the resulting average responses by condition. Target sentences in the insincere conditions were understood mostly

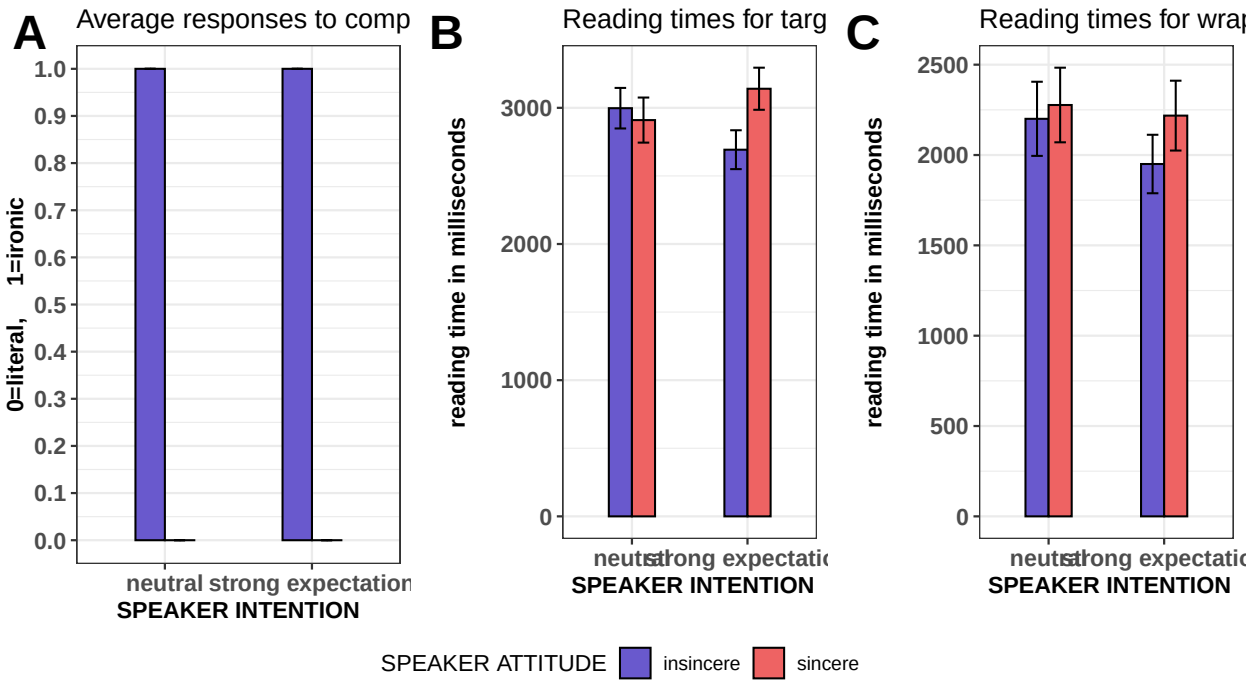

Figure 2. Average responses to comprehension questions (panel A) and raw-reading times of target (panel B) and wrap-up (panel C) sentences for trials with correct responses for Experiment 1. Error bars show confidence intervals.

as ironic (around 70% of the times), particularly in the strong expectation condition (82%). Sentences in the sincere conditions were perceived as literal (around 82% of the times), particularly in the neutral condition (around 93% of the times).

We fitted a mixed-effects logistic regression model to the data (sum-contrast coded). The reference levels for each factor were the neutral condition (factor: SPEAKER ATTITUDE) and the sincere condition (factor: SPEAKER INTENTION). The results confirmed our prediction and showed a main effect of SPEAKER ATTITUDE ($p<0.001$, $z=9.92$). There was an additional effect of SPEAKER INTENTION ($p=0.001$, $z=-3.37$) and no significant interaction ($z=-0.21$, $p=0.837$). The results are shown in Table 2. Overall, both 'insincere' conditions were understood above chance as ironic and both 'sincere' conditions as literal.

Table 2

Summary of model output for accuracy in comprehension questions, Experiment 1

term	$\hat{\beta}$	95% CI	z	p
SPEAKER INTENTION	2.05	[0.86, 3.25]	3.37	.001
SPEAKER ATTITUDE	4.25	[3.41, 5.10]	9.92	$< .001$
ATTITUDE x BIAS interaction	-0.08	[-0.83, 0.68]	-0.21	.837

Note. model used a sum-contrast coding scheme

Table 3

Summary of model output for reading times of target sentence, Experiment 1

term	$\hat{\beta}$	95% CI	t	df	p
SPEAKER INTENTION	0.01	[-0.02, 0.05]	0.79	8.20	.453
SPEAKER ATTITUDE	-0.02	[-0.08, 0.03]	-0.85	7.96	.421
ATTITUDE x BIAS interaction	-0.08	[-0.12, -0.03]	-3.59	8.88	.006

Note. model used a sum-contrast coding scheme

Table 4

Comparison between reading times in individual conditions, Experiment 1

term	$\hat{\beta}$	95% CI	t	df	p
Sincere/strong e. vs. Insincere/strong e.	-0.20	[-0.26, -0.14]	-6.39	64.01	$< .001$
Insincere/strong e. vs. Insincere/neutral	0.11	[0.05, 0.18]	3.63	70.18	.001
Sincere/neutral vs. Insincere/neutral	-0.09	[-0.15, -0.02]	-2.75	67.09	.008

Note. model used a sliding-contrast coding scheme

Table 5

Summary of model output for reading times of the wrap-up sentence, Experiment 1

term	$\hat{\beta}$	95% CI	t	df	p
SPEAKER INTENTION	-0.02	[-0.06, 0.02]	-0.93	6.82	.383
SPEAKER ATTITUDE	-0.04	[-0.07, 0.00]	-2.03	9.03	.073
ATTITUDE x BIAS interaction	-0.04	[-0.07, 0.00]	-1.83	7.35	.108

Note. model used a sum-contrast coding scheme

Reading times. To analyze reading times, we first excluded all incorrect responses. We then fitted a linear mixed-effects regression model to the log-transformed reading data of the target sentence. We settled on a log-transformation following the results of a box-cox test (Box & Cox, 1964). This test was performed because the residuals of a model using raw-reading times were not normally distributed.

The final model had an anova-style sum-contrast coding scheme, which allows us to test for main effects, and more importantly, it allows us the test the pre-registered prediction of their interaction directly. This model showed no main effects of SPEAKER ATTITUDE or of SPEAKER INTENTION. It did, however, show a significant interaction between both terms (p=0.006, t=3.59), in accord with our predictions. This model can be seen in Table 3.

With the thought of comparing reading times of ironic and literal interpretations of the same sentence, we followed up on these results by re-fitting the model using a sliding contrast coding scheme (as per our pre-registration). This form of contrast coding compares neighboring factor levels, which allows us to directly compare each relevant condition to each other. Specifically, we wanted to compare the two 'insincere' conditions ('strong expectation' and 'neutral') to one another, the two 'strong expectation' conditions ('sincere' and 'insincere') to one another, and the two 'neutral' conditions ('sincere' and 'insincere') to one another. This new model showed a significant difference between 'sincere-strong expectation'

and 'insincere-strong expectation' conditions ($p<0.001$, $t=6.39$), a significant difference between 'insincere-strong expectation' and 'insincere-neutral' ($p=0.001$, $t=3.63$), and a significant difference between 'sincere-neutral' and 'insincere-neutral' conditions ($p=0.008$, $t=2.75$). This model can be seen in Table 4. There were no spill-over effects found in the wrap-up sentence (see panel C of Figure 2 and Table 5 for the summarized results).

Discussion

Experiment 1 manipulated two sorts of information put at a participant's disposal prior to hearing a potentially ironic remark. These corresponded with two aspects of mindreading, namely (i) the degree to which information in the context allows a reader to anticipate a speaker's intention and (ii) explicit information about the speaker's attitude towards their own utterance. These two aspects are central to the echoic account of irony comprehension (Wilson & Sperber, 2012). Overall, our results showed that both (i) and (ii) affected comprehension of our target utterances. This pattern played out differently in reading times than it did in interpretation: For the ultimate interpretation of the sentence (quantified as responses to the comprehension question), speaker-specific cues about a speaker's attitude towards their upcoming utterance was the most important factor, with both sincere conditions being mostly understood as literal and both insincere conditions as ironic. However, the degree to which it is possible to anticipate a speaker's intention also influenced participants' irony comprehension: The more a participant expected a 'no' answer, the more they understood a 'yes' answer as ironic. This resulted in two main effects and no interaction.

For reading times, on the other hand, the interaction between both types of cues was crucial: when a sentence was understood as ironic ('insincere' conditions), it was read faster if participants had strong intuitions regarding the informative intention of the speaker ('insincere-strong expectation' condition) compared to when they did not ('insincere-neutral' condition). This finding supports the idea that differences in mindreading engagement (operationalized here as the degree to which a context allows a participant to anticipate the

speaker's upcoming intention as well as attitude towards a proposition) predict ease of processing ironic sentences.

A closer look at this interaction effect has an important bearing on the "ironic vs. literal" debate. First, consider the comparison of the 'insincere-strong expectation' condition to the 'sincere-strong expectation' condition. Among these two in the strong expectation condition, the one encouraging an ironic reading of a sentence is actually faster. Second, consider the 'sincere-neutral' condition as it is compared to the 'insincere-neutral' condition. Here, the ironically understood sentences are read slower than their literal counterparts. These findings therefore suggest two things. First, there is no primacy of the literal meaning regarding the processing speed of an entire sentence: We failed to find a main effect of ATITTUDE, which suggests that literal sentences were not faster to process than ironic ones across the board. Second, the underlying factor that mediates processing effort might not be whether the sentence is ironic or literal, but the degree to which context gives readers access to the speaker's intentions and beliefs. These results provide evidence for a likely rapid engagement of mindreading abilities when understanding irony. It also makes for a very rare finding of irony understanding actually being faster than its explicitly literal control. In Experiment 2 we seek to find further support for our claim by investigating individual differences between participants in comprehending ironic and literal sentences.

Experiment 2

Experiment 1 showed that mindreading considerations mediate irony comprehension effort. We view this as being a consequence of how participants use their mindreading skills: Having access to a speaker's intention can predict the comprehender's ease of irony processing and comprehension accuracy. However, it could be the case that participants in Experiment 1 were not engaging their mindreading skills, but instead learning to associate specific lexical cues in the context with a potential interpretation and used this association as a comprehension strategy. In other words, it could be that participants relied on

contextual cues without considering the speaker's intentions. To support our interpretation of Experiment 1, we need to seek out evidence suggesting that mindreading is at play.

We decided to go about this by taking an individual differences approach. Apperly (2012) argued that there are individual differences with regards to the degree to which people can routinely and appropriately put their mindreading skills to use. This has been shown to have repercussions for pragmatic language comprehension, in as much as people with more developed mindreading skills tend to show a better understanding of various pragmatic phenomena such as irony (Spotorno & Noveck, 2014), scalar implicatures (Fairchild & Papafragou, 2021) and humor (Bischetti, Ceccato, Lecce, Cavallini, & Bambini, 2019). If the differences between conditions in Experiment 1 were linked to differences in mindreading engagement, we should be able to find an association between comprehension of the critical items of Experiment 1 with individual differences in mindreading abilities. This is the goal of Experiment 2.

Participants and power analysis

To calculate power for Experiment 2, we ran a power analysis via simulations, similar to the procedure of Experiment 1. The main difference here is that the effect of interest for Experiment 2 was the interaction in the logistic regression model. The model used for the simulations included a maximally-converging random effects structure of random slopes by items and random intercepts by participants. Since it is not possible to calculate Cohen's D for a logistic regression model, we used a conservative estimate of half of the raw-effect size found in the pilot (i.e., the beta coefficient of the interaction term, see Table 2 of the supplementary materials). After simulating 1000 Experiments using the pilot's parameters and this new - conservative - beta coefficient, we concluded that an Experiment with 220 participants would have over 80% power to detect a true effect of that magnitude. The final effect found in Experiment 2 was larger than this conservative estimate, showing that Experiment 2 was sufficiently powered. The power analysis and pilot data are available on

the project's OSF repository.

We thus recruited 239 participants (who did not participate in Experiment 1), assuming that some might not meet the exclusion criterion: As in Experiment 1, we intend to exclude participants who do not correctly answer at least 5 out of the 7 comprehension questions in the filler items. For Experiment 2, the exclusion criterion led to the exclusion of 16 participants, leaving the final number at 223.

Materials, design and procedure

The materials, design and procedure were similar to that of Experiment 1. There were three differences: First, we kept only the 'insincere - strong expectation' and the 'sincere - neutral' conditions, since these two conditions were the ones that were most typically understood as ironic and literal, respectively. Second, we decided to show participants 8 of the critical items of Experiment 1 in these two conditions (i.e., 4 items in each condition). This was done to balance the number of items in each condition seen by participants relative to Experiment 1. Third (and most importantly), Experiment 2 included a mindreading task, administered to participants after completing the experiment. This task was an abridged version of the 'Reading the Mind in the Eyes' (RME) task (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), meant to measure each participant's ability to deploy their mindreading skills. This abridged version consisted of the first 24 trials of the task. We chose to use an abridged version in order to keep the experiment as a whole as short as possible. Everything else was identical to the original task by Baron-Cohen, Wheelwright, Hill, Raste, and Plumb (2001). We opted for the RME instead of other advanced mindreading measures such as the 'Strange Stories' task (F. G. Happé, 1994) because the former relies less than the latter on pragmatic competence, i.e. on understanding language use in specific contexts (Bosco, Tirassa, & Gabbatore, 2018). As Bosco, Tirassa, and Gabbatore (2018) argue, when tasks explicitly rely on figurative language comprehension and pragmatic inferencing as measures of higher mindreading abilities (such as the 'Strange Stories' does), it is difficult to

estimate the true degree to which mindreading correlates with the comprehension of pragmatic phenomena (such as irony), given that both things are effectively measured with the same task. We address the limitations of using the RME task in the General Discussion.

We computed a mindreading score for each participant based on their results on the RME task. This score was used as a continuous predictor for analyzing the responses to the comprehension questions and the reading times of the target sentence. Together with this continuous predictor (which we refer to as MINDREADING), we coded the 'insincere - strong expectation' and the 'sincere - neutral' conditions as two levels ('ironic' and 'literal,' respectively) of the same factor (SENTENCE TYPE) and included them as predictors of comprehension accuracy and response times. We also included the interaction between MINDREADING and SENTENCE TYPE as a predictor.

Predictions

The landmark study by F. G. E. Happé (1993) showed that irony comprehension correlated with success in a second-order false-belief task, which led her to interpret the results as supporting the echoic account. Wilson and Sperber (2012) (pg. 134) echo this interpretation by stating that Happé's results "confirm the relevance-theoretic account of figurative utterances." We interpret this as an indicator that the echoic account predicts that mindreading scores should correlate necessarily with irony comprehension scores. However, the theory seems to remain vague as to whether mindreading scores should also correlate with irony processing speed. For this reason, our pre-registered predictions refer to sentence comprehension only, as indicated by responses to the comprehension questions after the critical items, and we analyze the reading time data as an exploratory measure only.

We hypothesized that if the context cues used in Experiment 1 (information about the speaker's attitude towards their upcoming proposition and a contextual bias towards expecting a 'no' answer to the polar question) reflect the way in which comprehenders engage in mindreading abilities, there should be a link between an individual comprehender's level

of mindreading skill and their responses in the different conditions (specifically, the 'insincere - strong expectation' and the 'sincere - neutral' conditions of Experiment 1, which are called 'ironic' and 'literal' in Experiment 2). This should result in a significant interaction between MINDREADING and SENTENCE TYPE for responses to the comprehension questions.

Concretely, we predicted that participants with higher mindreading scores should be better at understanding irony in the 'ironic' condition compared to participants with lower mindreading scores. No such effect of MINDREADING should be visible in the 'literal' condition. This prediction reflects that (i) we believe the pattern of results of Experiment 1 to be related to mindreading engagement, and (ii) enhanced mindreading should be particularly advantageous for understanding ironic utterances and not their literal counterparts. These predictions directly motivate how we analyzed our data, which we describe in the following section.

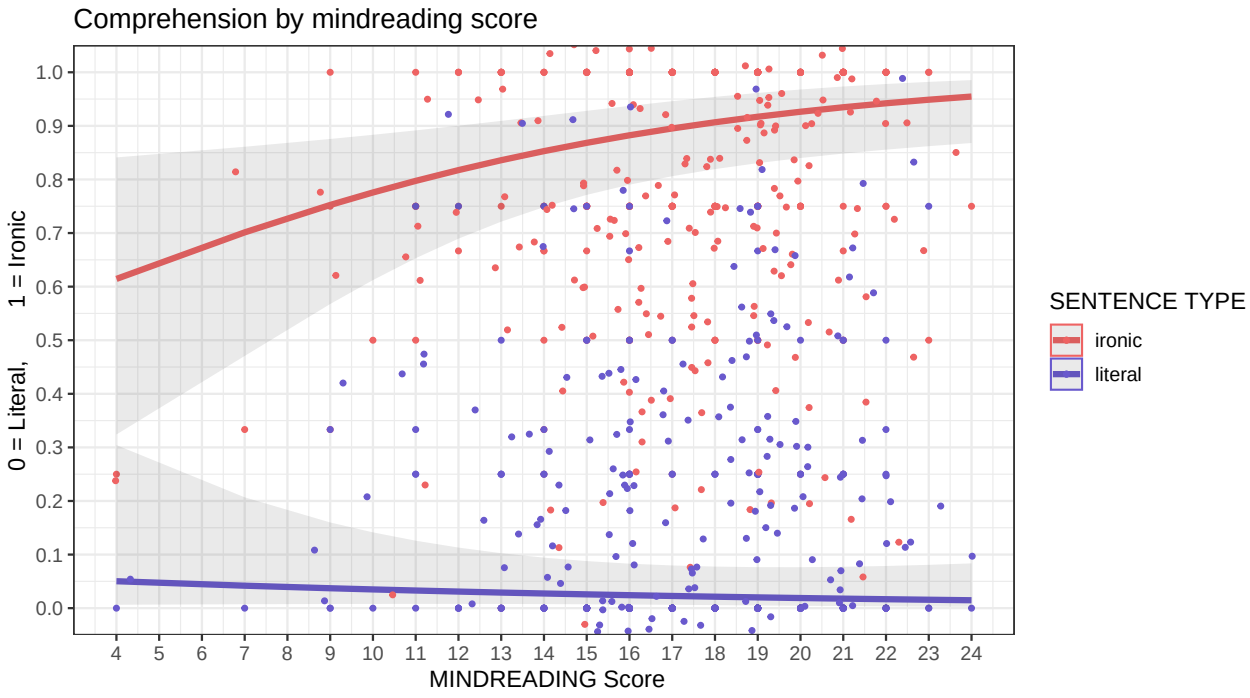

Figure 3. Responses by mindreading scores in Experiment 2. Individual dots show participant averages. Plotted lines and gray ribbons show the predicted values of the logistic regression model and confidence intervals, respectively.

Table 6

Model results for comprehension questions (ironic), Experiment 2

term	$\hat{\beta}$	95% CI	z	p
MINDREADING	0.13	[0.03, 0.23]	2.59	.010
SENTENCE TYPE	-2.64	[-5.59, 0.32]	-1.75	.081
MINDREADING x SENTENCE TYPE interaction	-0.19	[-0.36, -0.03]	-2.28	.023

Note. model used a treatment-contrast coding scheme, ironic condition is coded as the baseline

Analysis and results

RME task. Because we used an abridged version of the RME task, we assessed our version's internal consistency to evaluate its similarity to the original, un-abridged version. We found that our task had a Cronbach's alpha value of 0.61, similar to that found in previously reported uses of the unabridged version (e.g., Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 2005; Vellante et al., 2013; Voracek & Dressler, 2006). We also found a McDonald's Omega value of 0.64. Additionally, an exploratory factor analysis suggests that a unidimensional model was an adequate fit to the data, with an RMSEA index of 0.041 and a BIC of -987. A model with three factors was a better fit to the data (BIC: -898), a phenomenon which has also been observed for the unabridged version (Olderbak et al., 2015). These analyses can be found in the supplementary materials.

Comprehension questions. After excluding trials in which participants selected the distractor response (4% of all trials), we fitted a mixed effects logistic regression model to analyze comprehension data, as we did in Experiment 1. In Experiment 2, the model included the factor SENTENCE TYPE (levels: 'ironic' and 'literal'), the continuous predictor MINDREADING (which was first scaled), and their interaction. Since the goal of our analysis was to test the interaction and see whether MINDREADING affected the two levels of SENTENCE TYPE differently (as per our pre-registered predictions), we fitted the

Table 7

Summary of model output for comprehension questions (literal), Experiment 2

term	$\hat{\beta}$	95% CI	z	p
MINDREADING	-0.05	[-0.14, 0.04]	-1.05	.293
SENTENCE TYPE	1.68	[-0.50, 3.86]	1.51	.131
MINDREADING x SENTENCE TYPE interaction	0.19	[0.07, 0.31]	2.99	.003

Note. model used a treatment-contrast coding scheme, literal condition is coded as the baseline

same model twice using treatment contrast coding: One in which the 'ironic' condition was coded as the baseline, and another in which the 'literal' condition was coded as the baseline. When using a treatment contrast-coding scheme, the coefficients of each predictor represent an effect relative to the baseline condition only. In other words, with treatment contrast we only test the simple effect of MINDREADING on the baseline condition ('ironic' and 'literal,' in each of the two models), instead of the main effect of MINDREADING on responses across conditions. Re-fitting the model thus addresses the prediction that MINDREADING should impact irony comprehension but not the comprehension of literal sentences. Both iterations of the model included random intercepts and slopes for MINDREADING, SENTENCE TYPE and their interaction by items, and a random intercept and slope term for SENTENCE TYPE by participants.

The results of the model are summarized in Table 6, and the results pattern is illustrated in Figure 3. As predicted, there was a significant interaction between MINDREADING and SENTENCE TYPE (z=2.28, p=0.023). This suggests that MINDREADING had a different effect on each of the levels of SENTENCE TYPE: As predicted, higher mindreading scores resulted in significantly more correct interpretations in the 'ironic' condition (z=2.59, p=0.01), and we failed to find an effect of MINDREADING on the 'literal' condition (z=1.05, p=0.293) (see Table 7).

Reading times. As in Experiment 1, we first excluded incorrect responses from the analysis (i.e., ironic answers in the 'literal' condition and 'literal' answers in the ironic condition). This resulted in the removal of 14% of the data. We fitted a mixed-effects linear regression model to the log-transformed reading times of remaining data. We included MINDREADING, SENTENCE TYPE, and their interaction as predictors. The final random-effects structure included random intercepts by participants and items, as well as a random slope term for SENTENCE TYPE by items. We failed to find any significant effects of our predictors on the log-transformed reading times.

Discussion

In Experiment 2, we anticipated that individual participants' scores on the 'Reading the Mind in the Eye' task would be predictive of their accuracy in understanding irony - but not in understanding literal sentences. The results confirmed our predictions: Participants with higher mindreading scores were better at understanding irony than those with lower mindreading scores, but not at understanding literal sentences. This result suggests that the contextual manipulations of Experiment 1 - being aware of the speaker's intention and knowing the speaker's attitude - were in fact tapping into the way in which participants engaged their mindreading abilities during reading comprehension. This is true of at least the 'insincere-strong expectation' and 'sincere-neutral' conditions of Experiment 1, which were the most prototypically ironic and literal, respectively.

The failure to find effects of MINDREADING on reading times could have various explanations. First, irony comprehension is quite low for participants in the bottom-half of the distribution of RME scores. THis means that there are very few instances of successful irony comprehension for which we could measure RTs for these participants. It could be the case that more observations are necessary to detect effects in this regard. However, precisely because comprehension accuracy of irony is low for low-scores on the RME task, it is not clear whether it would even be meaningful to interpret the processing effort of the instances

that these participants do accurately recognize as ironic. These might be either chance occurrences or guided by altogether different comprehension mechanisms.

General Discussion

The goal of the current study was twofold. First, we aimed to investigate whether providing contextual cues that point to a speaker's intention and their attitude towards an upcoming proposition could explain reading times differences with respect to ironic relative to literal readings (Experiment 1). Second, we wanted to examine whether any such irony comprehension differences brought on by context were related to individual differences in participants' engagement in mindreading abilities (Experiment 2).

Our results broadly support our predictions. First, having access to the speaker's beliefs and intentions plus information about a speaker's attitude towards an upcoming proposition provide the means for a rapid interpretation of an ironic response: When participants were told that the speaker might not be committed to the truth of an utterance (because he is 'known to be a jokester,' for example) they overwhelmingly understood the target sentence as ironic compared to when they believed the speaker to be committed to the truth of their utterance (Experiment 1). This finding is in line with previous research showing that speaker-specific information affects the overall rate of interpretation of ironic sentences (Katz & Pexman, 1997; Pexman & Olineck, 2002).

Second, our results show that encouraging participants to engage with a speaker's intentions facilitates processing effort of utterances that are understood ironically. When sentences were understood ironically (based on the responses to comprehension questions), participants read them faster if they were embedded in a context that made manifest a specific intention (prior to a polar question) compared to when the discourse context did not aim to generate any specific expectations. This finding complements the literature on the interaction between mindreading skills and language processing (e.g., Ferguson & Breheny, 2011; Rubio-Fernández, Mollica, Oraa Ali, & Gibson, 2019) by showing a further

phenomenon for which mindreading, when engaged via a linguistic context, has a rapid effect. The finding also supports and extends the work of Spotorno and Noveck (2014) by showing that mindreading is critical for irony comprehension. Here, we showed how mindreading can have an impact within individual trials; it need not rely on trial effects.

Third, we showed that our experimental manipulation correlated with individual differences in mindreading abilities. Participants who scored lower on a mindreading test were less accurate in understanding irony. This was not the case for the comprehension of literal sentences, supporting the claim that mindreading skills are particularly relevant for understanding ironically - but not literally - intended sentences.

Finally, and most importantly, our results contribute to the debate on processing ironic relative to literal sentences. We show that knowledge about a speaker's attitude and expectations about a speaker's intention interact during reading. The result of this interaction is that one and the same sentence can be read faster or slower - and ironically or literally - depending on the degree to which participants think they can anticipate the intention and attitude of the speaker. We suggest that the key to understanding the relationship between context and processing effort of ironic sentences is to focus on how a context helps comprehenders anticipate the intentions of a speaker. With this in mind, we can make sense of previous incongruous empirical findings: whenever a specific cue aids in engaging a comprehender's mindreading abilities, it will ease processing, so that an ironic interpretation can be reached just as fast as a literal one would.

Interpreting the results of the RME task

The RME has often been used as a measure of individual differences in mindreading skills for neurotypical populations (e.g., Domes, Heinrichs, Michel, Berger, & Herpertz, 2007; Kidd & Castano, 2013; Mar, Oatley, Hirsh, Dela Paz, & Peterson, 2006). However, the task's validity and interpretation have been heavily criticized (e.g., Baker, Peterson, Pulos, & Kirkland, 2014; Black, 2019; Oakley, Brewer, Bird, & Catmur, 2016). Oakley, Brewer, Bird,

and Catmur (2016) noted that performance in this task is likely driven by the ability to identify emotional states rather than the ability to attribute mental states to others. The authors argue that the RME task picks up on differences between ASD and neurotypical populations because Autism often co-occurs with Alexithymia, a clinical condition characterized by difficulties in identifying and describing one's emotional states and those of others (see also Bird & Cook, 2013). This co-occurrence is so prevalent that difficulties in the emotional domain have traditionally been considered a common trait of ASD and are even used as diagnostic markers (Lord et al., 2000). This is important to consider given the pattern of results of Experiment 2. Here, we see that the rate of irony comprehension as a function of MINDREADING scores only seems to change for participants who scored in the bottom-half on the RME task. This range of scores is well within the range typically found for ASD individuals (mean=62.7%, according to a meta-analysis by Peñuelas-Calvo, Sareen, Sevilla-Llewellyn-Jones, & Fern'andez-Berrocal, 2019). It is possible that at least some of our participants displayed difficulties both in identifying emotions (as tracked by the RME task) and in attributing mental states to others (since these two difficulties often co-occur for ASD individuals). This would explain the differences in irony comprehension for participants with low RME scores. However, we did not collect information on participants' ASD diagnosis, so this remains speculative.

This interpretation is nonetheless consistent with some of the other criticism that the RME task has received. For example, Black (2019) states that the RME task should be seen as an instrument best suited for detecting strong mindreading differences such as those between ASD and neurotypical populations. This could explain why, for the high performers in the RME task of our Experiment 2, there was no obvious effect of RME scores on irony comprehension: The RME task was likely not sensitive enough to detect any differences in participants with normal to high mindreading abilities. In other words, the RME task might be best seen as a sort of 'blunt' tool that can pick up on the substantial differences between individuals with low- and high-mindreading abilities, but is not ideal for detecting the finer

differences between mindreaders at the upper end of the scale. The goal of Experiment 2 was to determine whether the elicited differences in comprehension found in Experiment 1 could be said to be related to the engagement of mindreading abilities. We interpret the broad differences detected by the RME as sufficient evidence of this. However, future studies interested in fine-grained differences among individuals with high mindreading skills should rely on other measurement tasks better suited to that particular population.

Implications for theories of irony comprehension

The present work derived its predictions from the echoic account of irony (Wilson & Sperber, 2012). We see our study as an extension of this account by way of providing testable linking hypotheses for the theory. Specifically, we provide a principled explanation of what type of context influences irony processing and why: A context that allows one to have access to the speaker's likely informative intention and manifest access to the attitude attached to their proposition will ease irony comprehension. The fact that the contexts we created in Experiment 1 triggered an engagement of these two aspects of mindreading is supported by the results of Experiment 2. Here, we found that individual differences in mindreading abilities critically interact with the experimental conditions, at least in as much as there is a difference between low and high-scorers on the RME task.

The results of the two experiments could also be interpreted as being compatible with other theoretical views. Indeed, our findings can be considered compatible with contextualist accounts such as the Direct Access View (Gibbs, 2002) and the Constraint-Satisfaction Account (Pexman, 2008; Pexman, Ferretti, & Katz, 2000) in the sense that they underline the role played by context during online processing of irony and show that irony can be processed faster than literal language (under certain circumstances). However, a major drawback of these accounts is the absence of a systematic weighing of contextual factors with regards to how they affect processing. This makes it hard for the accounts to predict which contextual aspects will facilitate irony processing and which will not.

Our findings are less compatible with context-independent accounts - such as the Graded Salience Hypothesis (Giora, 2003; Giora et al., 2007) - since such accounts would posit a primacy of salient readings at the sentence-level (which are literal in the majority of cases), regardless of context. In Experiment 1, we showed how one and the same sentence can be read faster or slower as a function of contextually raised expectations of a specific nature. This resulted in irony sometimes being faster and sometimes slower than a literal control, which is at odds with a stronger version of the context-independent view (for example, as formulated by Schwoebel, Dews, Winner, & Srinivas, 2000) that would preclude irony from being read faster than literal equivalents, regardless of a contextual bias. That said, the goal of our study was not to test the predictions of the Graded Salience Hypothesis, but to derive and test predictions from the echoic account in order to investigate the role that mindreading plays with regards to irony comprehension. To test the predictions of the Graded Salience Hypothesis, it would be necessary to carefully norm the target ironic sentences for frequency, prototypicality, familiarity, and other factors that might affect salience. We leave it to future work to study how a mindreading-facilitating context can be used to test the predictions of context-independence views.

An alternative explanation of our results could relate to the granularity of our measures. For example, it could be the case that participants in Experiment 1 first read the initial part of the target sentence (specifically, the word 'yes') faster when intended literally (i.e., the 'sincere' conditions) than when intended ironically (i.e., the insincere conditions) and that a processing advantage for ironic readings only appeared later downstream as the sentence was integrated with context. Though our data cannot rule out this scenario (since we only measured reading times of the entire sentence), we think it is not likely given the size of differences in reading times. For example, the effect size we found for the difference between the Sincere/strong expectations (understood literally) vs. Insincere/strong expectations (understood ironically) in Experiment 1 was of Cohen's $d = 0.42$, or about 450 milliseconds advantage for the ironic condition. This is substantial, and it seems unlikely

that it would be this big if there had first been an effect in the opposite direction at the beginning of the sentence.

Further, such a pattern would not be predicted by theoretical accounts that posit an overall processing advantage for literal over ironic readings (as far as we can tell). For example, the Graded Salience Hypothesis (probably the most prominent 'context-independence' theory of irony) explicitly states that there should be no differences between irony and literal readings of the same expressions at the earliest stages (i.e., at the word-level processing stage), and that a processing speed advantage for salient-based (i.e., literal) readings should appear in a second stage at the earliest (Giora & Fein, 1999; Giora et al., 2007, pg. 141; Giora, Fein, & Schwartz, 1998). Because of this, we see our explanation as being more parsimonious: The degree to which participants can anticipate the speaker's intention is what drives processing speed differences between ironic and literal interpretations of the same sentence.

Conclusion

The current study suggests that differences in mindreading engagement induced via a discourse context result in systematic differences in irony comprehension. These results provide linking hypotheses for the echoic account of irony comprehension as well as contribute to the long-standing debate on the processing effort of verbal irony by showing under which conditions ironic sentences can be read faster than literal equivalents.

References

Allott, N. (2017). Metarepresentation. In *The Routledge Handbook of Pragmatics* (pp. 295–309). Routledge.

Apperly, I. A. (2012). What is "theory of mind?" Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology*, *65*(5), 825–839.

Baker, C. A., Peterson, E., Pulos, S., & Kirkland, R. A. (2014). Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes." *Intelligence*, *44*, 78–92.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*(7), e285–e285.

Bischetti, L., Ceccato, I., Lecce, S., Cavallini, E., & Bambini, V. (2019). Pragmatics and theory of mind in older adults' humor comprehension. *Current Psychology*, 1–17.

Black, J. E. (2019). An IRT analysis of the reading the mind in the eyes test. *Journal of Personality Assessment*, *101*(4), 425–433.

Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018). Why Pragmatics and Theory of Mind Do Not (Completely) Overlap. *Frontiers in Psychology*, *9*.

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Bryant, G. A. (2012). Is Verbal Irony Special?: Is Verbal Irony Special? *Language and Linguistics Compass*, *6*(11), 673–685. https://doi.org/10.1002/lnc3.364

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Degen, J., & Tanenhaus, M. K. (2019). Constraint-based pragmatic processing. *The Oxford Handbook of Experimental Semantics and Pragmatics*, 21–38.

Domes, G., Heinrichs, M., Michel, A., Berger, C., & Herpertz, S. C. (2007). Oxytocin improves "mind-reading" in humans. *Biological Psychiatry*, *61*(6), 731–733.

Drummond, A. (2013). Ibex farm. *Online Server: Http://Spellout. Net/Ibexfarm*.

Fairchild, S., & Papafragou, A. (2021). The Role of Executive Function and Theory of Mind in Pragmatic Computations. *Cognitive Science*, *45*(2), e12938.

797 Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of
798 anticipating behaviour based on complex, conflicting desires. *Cognition*, *119*(2),
799 179–196. https://doi.org/10.1016/j.cognition.2011.01.005

800 Filik, R., Howman, H., Ralph-Nearman, C., & Giora, R. (2018). The role of
801 defaultness and personality factors in sarcasm interpretation: Evidence from
802 eye-tracking during reading. *Metaphor and Symbol*, *33*(3), 148–162.
803 https://doi.org/10.1080/10926488.2018.1481258

804 Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*,
805 *116*(3), 421–436.

806 Futrell, R. (2012). *Processing effects of the expectation of informativity*.

807 Gibbs, R. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental*
808 *Psychology: General*, *115*(1), 3–15. https://doi.org/10.1037/0096-3445.115.1.3

809 Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language, and*
810 *Understanding*. Cambridge University Press.

811 Gibbs, R. (2002). A new look at literal meaning in understanding what is said and
812 implicated. *Journal of Pragmatics*, *34*(4), 457–486.
813 https://doi.org/10.1016/S0378-2166(01)00046-7

814 Giora, R. (2003). *On Our Mind: Salience, Context, and Figurative Language*. Oxford
815 University Press.

816 Giora, R., Drucker, A., Fein, O., & Mendelson, I. (2015). Default sarcastic
817 interpretations: On the priority of nonsalient interpretations. *Discourse Processes*,
818 *52*(3), 173–200.

Giora, R., & Fein, O. (1999). Irony: Context and salience. *Metaphor and Symbol*, *14*(4), 241–257.

Giora, R., Fein, O., Laadan, D., Wolfson, J., Zeituny, M., Kidron, R., ... Shaham, R. (2007). Expecting irony: Context versus salience-based effects. *Metaphor and Symbol*, *22*(2), 119–146.

Giora, R., Fein, O., & Schwartz, T. (1998). Irony: Grade salience and indirect negation. *Metaphor and Symbol*, *13*(2), 83–101.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge: Harvard University Press.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129–154.

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119. https://doi.org/10.1016/0010-0277(93)90026-R

Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition & Emotion*, *19*(7), 999–1025.

841 Ivanko, S. L., & Pexman, P. M. (2003). Context Incongruity and Irony Processing.
842 *Discourse Processes*, *35*(3), 241–279.
843 https://doi.org/10.1207/S15326950DP3503_2

844 Jorgensen, J., Miller, G. A., & Sperber, D. (1984). Test of the mention theory of
845 irony. *Journal of Experimental Psychology: General*, *113*(1), 112.

846 Katz, A. N., Blasko, D. G., & Kazmerski, V. A. (2004). Saying What You Don't
847 Mean: Social Influences on Sarcastic Language Processing. *Current Directions in*
848 *Psychological Science*, *13*(5), 186–189.
849 https://doi.org/10.1111/j.0963-7214.2004.00304.x

850 Katz, A. N., & Pexman, P. M. (1997). Interpreting Figurative Statements: Speaker
851 Occupation Can Change Metaphor to Irony. *Metaphor and Symbol*, *12*(1), 19–41.
852 https://doi.org/10.1207/s15327868ms1201_3

853 Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind.
854 *Science*, *342*(6156), 377–380.

855 Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ...
856 Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A
857 standard measure of social and communication deficits associated with the
858 spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3),
859 205–223.

860 Mar, R. A., Oatley, K., Hirsh, J., Dela Paz, J., & Peterson, J. B. (2006). Bookworms
861 versus nerds: Exposure to fiction versus non-fiction, divergent associations with
862 social ability, and the simulation of fictional social worlds. *Journal of Research in*
863 *Personality*, *40*(5), 694–712.

McDonald, S. (2000). Neuropsychological Studies of Sarcasm. *Metaphor and Symbol*, *15*(1-2), 85–98. https://doi.org/10.1080/10926488.2000.9678866

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds.* Clarendon Press/Oxford University Press.

Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge University Press.

Oakley, B. F., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology*, *125*(6), 818.

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: Toward a brief form for research and applied settings. *Frontiers in Psychology*, *6*, 1503.

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433.

Peñuelas-Calvo, I., Sareen, A., Sevilla-Llewellyn-Jones, J., & Fern'andez-Berrocal, P. (2019). The "Reading the Mind in the Eyes" test in autism-spectrum disorders comparison with healthy controls: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders*, *49*(3), 1048–1061.

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., Ferretti, T. R., & Katz, A. N. (2000). Discourse Factors That Influence Online Reading of Metaphor and Irony. *Discourse Processes*, *29*(3), 201–222. https://doi.org/10.1207/S15326950dp2903_2

Pexman, P. M., & Olineck, K. M. (2002). Understanding irony: How do stereotypes cue speaker intent? *Journal of Language and Social Psychology*, *21*(3), 245–274.

R Core Team. (2020). *R: A language and environment for statistical computing* [Manual]. Vienna, Austria: R Foundation for Statistical Computing.

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121–135. https://doi.org/10.1016/j.brainres.2009.10.077

Rubio-Fernández, P., Mollica, F., Oraa Ali, M., & Gibson, E. (2019). How do you know that? Automatic belief inferences in passing conversation. *Cognition*, *193*, 104011. https://doi.org/10.1016/j.cognition.2019.104011

Schwoebel, J., Dews, S., Winner, E., & Srinivas, K. (2000). Obligatory processing of the literal meaning of ironic utterances: Further evidence. *Metaphor and Symbol*, *15*(1-2), 47–61.

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 76–116). New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511609213.006

Spaulding, S. (2020). What is mindreading? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(3), e1523.

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

Sperber, D., & Wilson, D. (2002). Pragmatics, Modularity and Mind-reading. *Mind & Language*, *17*(1-2), 3–23. https://doi.org/10.1111/1468-0017.00186

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of Experimental Psychology: General*, *143*(4), 1649–1665. https://doi.org/10.1037/a0036630

Ţurcan, A., & Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The roles of familiarity and context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1867–1893. https://doi.org/10.1037/xlm0000285

Ţurcan, A., & Filik, R. (2017). Investigating sarcasm comprehension using eye-tracking during reading. *Irony in Language Use and Communication*, *1*, 255.

Ţurcan, A., Howman, H., & Filik, R. (2020). Examining the role of context in written sarcasm comprehension: Evidence from eye-tracking during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Vellante, M., Baron-Cohen, S., Melis, M., Marrone, M., Petretto, D. R., Masala, C., & Preti, A. (2013). The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy. *Cognitive Neuropsychiatry*, *18*(4), 326–354.

Voracek, M., & Dressler, S. G. (2006). Lack of correlation between digit ratio (2D: 4D) and Baron-Cohen's "Reading the Mind in the Eyes" test, empathy, systemising, and autism-spectrum quotients in a general population sample. *Personality and Individual Differences*, *41*(8), 1481–1491.

Wilson, D., & Sperber, D. (1992). On verbal irony. *Lingua*, *87*(1), 53–76. https://doi.org/10.1016/0024-3841(92)90025-E

Wilson, D., & Sperber, D. (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139028370

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832