

# Children's processing of written irony: An eye-tracking study<sup>☆</sup>

Henri Olkonieni<sup>a,b,\*</sup>, Sohvi Halonen<sup>b</sup>, Penny M. Pexman<sup>c</sup>, Tuomo Häikiö<sup>b</sup>

<sup>a</sup> Division of Psychology, Faculty of Education and Psychology, University of Oulu, Finland

<sup>b</sup> Department of Psychology and Speech-Language Pathology, University of Turku, Finland

<sup>c</sup> Department of Psychology, University of Calgary, Canada

## ARTICLE INFO

### Keywords:

Language development  
Irony  
Reading  
Eye movements  
Children

## ABSTRACT

Ironic language is challenging for many people to understand, and particularly for children. Comprehending irony is considered a major milestone in children's development, as it requires inferring the intentions of the person who is being ironic. However, the theories of irony comprehension generally do not address developmental changes, and there are limited data on children's processing of verbal irony. In the present pre-registered study, we examined, for the first time, how children process and comprehend written irony in comparison to adults. Seventy participants took part in the study (35 10-year-old children and 35 adults). In the experiment, participants read ironic and literal sentences embedded in story contexts while their eye movements were recorded. They also responded to a text memory question and an inference question after each story, and children's levels of reading skills were measured. Results showed that for both children and adults comprehending written irony was more difficult than for literal texts (the "irony effect") and was more challenging for children than for adults. Moreover, although children showed longer overall reading times than adults, processing of ironic stories was largely similar between children and adults. One group difference was that for children, more accurate irony comprehension was qualified by faster reading times whereas for adults more accurate irony comprehension involved slower reading times. Interestingly, both age groups were able to adapt to task context and improve their irony processing across trials. These results provide new insights about the costs of irony and development of the ability to overcome them.

## 1. Introduction

In the novel *Harry Potter and the Half Blood Prince* (Rowling, 2005, pp. 327–328), Harry is leaving from the Weasley house and Mrs. Weasley says to him: "Promise me you will look after yourself... stay out of trouble..." Harry responds to her: "I always do Mrs. Weasley. I like a quiet life, you know me." Mrs. Weasley chuckles to his answer, and anyone familiar with Harry Potter knows that his life is far from quiet, and he is not really meaning what he is saying. In fact, Harry is being ironic.

Children, like adults, encounter ironic language in their daily lives. Ironic language is present in many contexts, from family conversations (Pexman, Zdrzilova, McConnachie, Deater-Deckard, & Petrill, 2009; Recchia, Howe, Ross, & Alexander, 2010) to literature, as in the example above from the Harry Potter book series. However, previous research has demonstrated that irony is challenging for children to understand (e.g., Capelli, Nakagawa, & Madden, 1990; Dews et al., 1996; Hancock,

Dunham, & Purdy, 2000; Harris & Pexman, 2003). Although challenging, comprehension of irony is an important skill for children. For example, comprehending irony is considered a major milestone in the development of children's social cognition, as it goes hand in hand with the development of understanding others' beliefs, intentions, and attitudes (Peterson, Wellman, & Slaughter, 2012; see also Bosco, Tirasa, & Gabbatore, 2018; Del Sette, Bambini, Bischetti, & Lecce, 2020). Thus, it is not surprising that deficits in irony comprehension have been shown to be related, for example, to feelings of social exclusion (Kim & Lantolf, 2018).

Most of the previous studies on children's irony comprehension have presented ironic and literal language examples using illustrated stories (e.g., Banasik-Jemielniak & Bokus, 2019; Filippova & Astington, 2008; Köder & Falkum, 2021; Winner & Leekam, 1991) and puppet shows (e.g., Climie & Pexman, 2008; Harris & Pexman, 2003). These studies have helped map the development of children's irony comprehension, but the issue of how children process and comprehend written irony remains

<sup>☆</sup> This paper is a part of special issue "20 Years of XPrag".

\* Corresponding author at: P.O. Box 2000, 90014, University of Oulu, Finland.  
E-mail address: [henri.olkonieni@oulu.fi](mailto:henri.olkonieni@oulu.fi) (H. Olkonieni).

underexplored. Children's processing has been examined in a few studies using variations on the visual world paradigm, where children's eye gaze and/or reaching to response objects is tracked (e.g., [Climie & Pexman, 2008](#); [Köder & Falkum, 2021](#); [Nicholson, Whalen, & Pexman, 2013](#); [Whalen, Doyle, & Pexman, 2020](#)). This paradigm, however, does not capture the very earliest moments of children's irony processing, and does not afford the fine-grained processing information that has been acquired in reading time studies with adults (see e.g., [Olkonieni & Kaakinen, 2021](#) for review). As such, the issue addressed by [Noveck, Bianco, and Castry \(2001\)](#), of the costs (and benefits) of figurative language processing compared to nonfigurative language, has been underexplored in the case of children's irony appreciation. The current paper represents an important step toward addressing that issue, investigating children's ironic vs literal language processing with greater precision than has previously been achieved.

It has been stated that, since children are not proficient readers, it would not be feasible to use reading time methodologies to study children's written irony comprehension ([Nicholson et al., 2013](#); [Pexman, 2008](#)). In the current study, however, we challenged this assumption, using eye-tracking methodology to investigate comprehension and processing of written irony by 10-year-old-children and comparing their performance to that of adults. With eye-tracking methodology it is possible to tap into the detailed time-course of processing written language during reading ([Rayner, 1998](#)). We further capitalized on the facts that: Finnish children have high reading literacy ([Mullis, Martin, Foy, & Hooper, 2017](#)), the reading performance of 10-year-olds is expected to be similar to that of adults for the reading of literal language (see [Blythe & Joseph, 2011](#), for a review on reading development), and 10-year-olds are still developing irony comprehension skills ([Glenwright & Pexman, 2010](#); [Nicholson et al., 2013](#); [Pexman, Glenwright, Krol, & James, 2005](#)).

### 1.1. Children's irony comprehension

What is currently known about children's irony comprehension is that it is a developing ability that typically first emerges when children are around 5–6 years old (e.g., [Dews et al., 1996](#); [Filippova & Astington, 2008](#); [Hancock et al., 2000](#)). By this age children have developed sufficient language and perspective-taking skills (or theory of mind) to understand something of the intended indirect meaning of ironic phrases (e.g., [Filippova & Astington, 2008](#); c.f., [Bosco & Gabbatore, 2017](#)). For some children, comprehension ability may start even earlier, when they are 3–4 years of age (e.g., [Angeleri & Airenti, 2014](#); [Loukusa & Leinonen, 2008](#); [Recchia et al., 2010](#)). However, this early comprehension is typically associated with higher use of ironic language in family environments ([Angeleri & Airenti, 2014](#); [Banasik-Jemielniak et al., 2020](#); [Recchia et al., 2010](#)). At the early stage of comprehension, children first learn to understand that with irony speakers say something other than what they mean. Nonetheless, at this early stage children's comprehension accuracy is low (e.g., [Hancock et al., 2000](#); c.f., [Banasik-Jemielniak & Bokus, 2019](#)). Children also tend to misinterpret ironic phrases as literal language (e.g., [Demorest, Christine, Phelps, Gardner, & Winner, 1984](#); [Loukusa & Leinonen, 2008](#)), or misunderstand them as lies ([Demorest et al., 1984](#); [Winner & Leekam, 1991](#)).

Around age 7 to 8, children start to understand the ironic speaker's intent ([Glenwright & Pexman, 2010](#); [Harris & Pexman, 2003](#); [Pexman et al., 2005](#)) and children's irony comprehension accuracy usually exceeds guessing-level (e.g., [Banasik-Jemielniak et al., 2020](#); [Dews et al., 1996](#); [Filippova & Astington, 2008](#); [Köder & Falkum, 2021](#); [Loukusa & Leinonen, 2008](#); [Pexman et al., 2005](#)). For example, [Dews et al. \(1996\)](#) showed short clips from children's favourite television cartoon shows (e.g., *Teenage Mutant Ninja Turtles*, and *The Little Mermaid*) containing irony to 5–6-year-olds, 8–9-year-olds, and adults. It was shown that 5–6-year-olds correctly comprehended 65% of the ironic items, 8–9-year-olds 82%, and adults 93%. When children reach adolescence, they comprehend irony more accurately and their own use of irony also

increases ([Aguert, Laval, Gauducheau, Atifi, & Marccoccia, 2016](#)). However, more sophisticated understanding of irony's humour and teasing functions continues to develop until early adulthood ([Capelli et al., 1990](#); [Glenwright & Pexman, 2010](#); [Glenwright, Tapley, Rano, & Pexman, 2017](#); [Whalen & Pexman, 2010](#)).

Children's increasingly accurate irony comprehension across middle childhood has been linked to a number of cognitive and linguistic skills. [Filippova and Astington \(2008\)](#) characterized irony comprehension as a form of advanced theory of mind reasoning. While several studies have found that second-order theory of mind is important to irony appreciation ([Filippova & Astington, 2008](#); [Happé, 1993](#); [Hayashi & Ban, 2021](#); [Massaro, Valle, & Marchetti, 2013](#); [Nilsen, Glenwright, & Huyder, 2011](#); [Sullivan, Winner, & Hopfield, 1995](#); [Winner & Leekam, 1991](#)), other studies have failed to find evidence for this relationship suggesting that second-order theory of mind is not a necessary condition for irony comprehension ([Angeleri & Airenti, 2014](#); [Bosco & Gabbatore, 2017](#); [Massaro, Valle, & Marchetti, 2014](#); [Panzeri, Giustolisi, & Zampini, 2020](#)). Children's irony comprehension has also been linked to executive function skills including cognitive flexibility ([Zajączkowska & Abbot-Smith, 2020](#)), inhibitory control ([Caillies, Bertot, Motte, Raynaud, & Abely, 2014](#)), and working memory ([Godbee & Porter, 2013](#)). Recently, [Mazzarella and Pouscoulous \(2020\)](#) suggested that children's irony comprehension depends on their emerging vigilance toward deception, or epistemic vigilance. Epistemic vigilance affords children the capacity to evaluate a speaker's honesty and distinguish lies from ironies. Certainly, these various cognitive skills are related to each other and to children's developing language abilities and their relationships (and directionality of those relationships) are yet to be determined.

In the present study we investigated irony comprehension and processing of 10-year-old children, and based on previous studies we expected children of this age to be proficient in their irony comprehension, but still less accurate in comparison to adults. However, the previous studies have not presented children with *written* irony, but rather irony in other modalities, and manner of presentation could influence the results. One could argue that narrated stories, which are used often in the experiments with children, would be easier for children to understand than written stories as the latter lack ironic tone of voice to aid the comprehension process. Indeed, some studies have shown that intonation can facilitate children's comprehension of ironic phrases ([Ackerman, 1983](#); [Capelli et al., 1990](#)). However, the findings related to ironic tone of voice are contradictory, as some studies have found no benefit of intonation for children's irony comprehension ([Köder & Falkum, 2021](#); [Winner & Leekam, 1991](#)) or for that of adults ([Bryant & Fox Tree, 2005](#); see also [Rivière, Klein, & Champagne-Lavau, 2018](#)). Furthermore, children have shown high levels of comprehension of ironic materials presented only with a neutral tone of voice ([Banasik-Jemielniak & Bokus, 2019](#)). On balance, these results suggest that irony comprehension accuracies for written irony may be comparable with those found in previous studies.

### 1.2. Development of children's reading skill

Children's processing of written irony will necessarily be influenced by their reading skills. Reading skills can be thought to develop through stages, in which children first process words via smaller units (e.g., [Ehri, 1995, 2005](#)). This is referred to as decoding. With practice, children then start using larger units, which gives way to more comprehensive reading skills (e.g., [Perfetti, 2007](#)). It has been shown that Finnish children develop their decoding skills early on ([Seymour, Aro, & Erskine, 2003](#)) and that their literacy skills are high (e.g., [Mullis et al., 2017](#)) compared to children in many other linguistic communities. Generally, the reading performance of 10-year-olds is expected to be similar to that of adults for the reading of literal, non-ironic language, although children are expected to be generally slower due to making more and longer fixations, shorter saccades, and more regressions (see [Blythe & Joseph, 2011](#), for a review of eye movement research in developing readers).

With regard to reading comprehension, Zargar, Adams, and Connor (2020) showed that while 3rd grade children (approx. 8–9 years of age) did not show differences in reading times of plausible and implausible words in sentence context, 4th graders (approx. 9–10 years of age) spent more time reading implausible words. This finding is indicative of active comprehension monitoring which emerges around 9 years of age. Indeed, several studies have shown that around 10 years of age, readers can detect inconsistencies in a text, although this varies between the tasks and readers (e.g., Connor et al., 2015; Oakhill, Hartt, & Samols, 2005; van der Schoot, Reijntjes, & van Lieshout, 2012). This is also seen in eye movement patterns, as Vorstius, Radach, Mayer, and Lonigan (2013) reported that 5th graders (on average 11 years of age) had larger rereading times when a sentence was more difficult. Despite this active comprehension monitoring, however, children may still not correctly answer explicit comprehension questions, such as whether the sentence is plausible.

As for strategic reading, Häikiö, Heikkilä, and Kaakinen (2018) showed that 7- to 8-year-old Finnish children did not tend to reread past sentences when reading for comprehension. In contrast, Kaakinen, Lehtola, and Paattilampi (2015) observed that 7–8-year-old children's first-pass reading was slower than that of 9–10-year-old children, but both age groups showed similar reading times for later rereading when reading for comprehension. When compared to adults, 9–10-year-olds spent more time during first-pass reading but there were no differences in probabilities or durations of later look-backs. The reading patterns of younger and older readers were modulated by the task. When reading to answer a “why” question (as opposed to reading for comprehension), younger children focused more on the first-pass reading while older children and adults made more look-backs. The task effect indicates that children were able to modify their reading behavior on the basis of task demands. Furthermore, the difference between younger and older children with regard to the task effect suggests a shift toward more strategic reading. Häikiö et al. (2018) hypothesized that this is due to younger children still being in the process of perfecting their decoding skills while the older children already use higher order comprehension strategies (see also Perfetti, 2007). Taken together, while reading development does continue after 10 years of age (witnessed e.g., in reading speed, Häikiö, Bertram, Hyönä, & Niemi, 2009; Kaakinen et al., 2015), the previous findings imply that around 10 years of age, many readers have developed comprehension and strategic reading skills required for resolving the irony in short stories.

### 1.3. Processing of written irony

Currently we do not know how children resolve written irony. In general, it is assumed that there may be costs associated with processing of figurative expressions, as contrast between expectations and literal meaning needs to be resolved (e.g., Grice, 1975; Noveck et al., 2001). For irony, theories assume that when the ironic phrase is not more familiar than its literal counterpart (e.g., the familiar irony “Yeah right”; see Giora, 2003) and is not supported by preceding context (i.e., there are no cues about forthcoming irony) the ironic statement should be harder to comprehend and take more time to process (Gibbs, 1994). Eye-tracking studies of processing of verbal irony with adult participants have shown this to be the case (see e.g., Olkonieni & Kaakinen, 2021 for review). That is, previous eye-tracking studies measuring comprehension accuracies of written ironies have systematically shown ironic statements being harder to comprehend than literal statements (e.g., Au-Yeung, Kaakinen, Liversedge, & Benson, 2015; Kaakinen, Olkonieni, Kinnari, & Hyönä, 2014; Olkonieni, Ranta, & Kaakinen, 2016). The irony comprehension accuracies in these studies range between 75% (Kaakinen et al., 2014, Exp 1) and 90% (Olkonieni et al., 2016). Moreover, processing of irony in text has been shown to take longer than for its literal counterpart (e.g., Filik & Moxey, 2010). This slowdown is typically seen as increased looking back to ironic than literal target phrase when the phrase has been already read once, and increased

returns (i.e., look-backs) to context from the ironic than literal target phrase (e.g., Olkonieni & Kaakinen, 2021). Based on the theories of irony comprehension the slowdown would reflect interpretation process, in which ironic meaning is integrated to the context (e.g., Grice, 1975; Spotorno & Noveck, 2019). This process is suggested to be dependent on a reader recognizing the discrepancy between literal meaning and the context and/or realizing that a protagonist criticizes a previous expectation or belief (e.g., Spotorno & Noveck, 2019; Wilson, 2013). However, not all of the rereading is necessarily related to comprehension (Olkonieni, Johander, & Kaakinen, 2019). Olkonieni, Johander, and Kaakinen (2019) showed that higher probability to look-back from the ironic target phrase to context was associated with poorer irony comprehension. However, these results should be interpreted with caution, as Olkonieni et al. used a masking paradigm in their study that disrupted normal reading, which might affect generalizability of the results to normal reading.

Moreover, previous studies have shown trial effects: evidence that the processing of ironic phrases changes across an experimental session (Olkonieni et al., 2016; Olkonieni, Johander, & Kaakinen, 2019; Olkonieni, Strömberg, & Kaakinen, 2019; Spotorno & Noveck, 2014). For example, in the study by Spotorno and Noveck (2014) participants read short stories containing ironic and literal sentences at their own pace, one sentence at a time. They found that participants showed higher reading times for ironic than literal target sentences in the beginning of the experimental session, but the effect wore off toward the end of the experiment (they call this Early-Late effect). They also found the same effect for the sentence following an ironic target phrase (this is typically called the spillover region in eye tracking studies). These results suggest that when the reader repeatedly encounters ironic statements, there will be an expectation of forthcoming irony, making the processing of ironic statements easier and hence more similar to that of literal statements. Thus, readers adjust to the experimental context and the costs of irony are attenuated.

Currently, we do not know how the costs of irony that have been observed for adult processing times (overall and across trials) might translate to children. Most theories of irony comprehension do not address developmental changes. However, theories that take individual differences into account, such as the *parallel constraint-satisfaction framework* (Pexman, 2008) and the *predictive coding theory* of irony (Fabry, 2021), could be extended to make the prediction that as irony comprehension ability is still developing in children, they should show more difficulty in processing and comprehending irony than adults. The increased processing effort could be seen as extra effort that children would need to invest in integrating ironic meaning to text context, in comparison to adults. This prediction is admittedly not very precise, but it is consistent with previous findings from studies on children's irony comprehension.

### 1.4. Present study

We recruited 4th grade Finnish elementary school children (10-year-olds) and compared their comprehension and processing of written irony to that of adults. Previously, it has been suggested that processing of written irony cannot be studied with children, because their reading skill is not sufficiently developed (Nicholson et al., 2013). However, Finnish children develop decoding skills early (Seymour et al., 2003) and have high reading literacy (Mullis et al., 2017). In addition, as the reading performance of 10-year-olds is similar, albeit slower, to adults for the reading of literal, non-ironic language, it is likely that effects of written irony could be observed in a sample of Finnish 10-year-old children.

We expected that children would show similar reading patterns, and similar comprehension of the literal target statements, to that of adults. At the same time, we expected that children would struggle with their irony comprehension, as indexed by lower comprehension scores for ironic than literal statements. We expected that children would show

above guessing-level irony comprehension accuracy, whereas adults were expected to show near-ceiling performance (e.g., Capelli et al., 1990; Olkonieni & Kaakinen, 2021). Successful comprehension should be seen as increased rereading of ironic statements in comparison to their literal counterparts. For children, the increase in rereading time for ironic statements should be more pronounced than for adults. Alternatively, it could also be that children do not recognize the ironic intent at all; if so, there should be no difference in processing ironic statements and their literal counterparts.

Finally, it should be noted that the materials used in the present experiment were designed for children. Thus, the materials might be too simple for the adults (see e.g., Schroeder et al., 2022), which might affect the results for adults. In the context of the present experiment this might mean that the irony effects in comprehension and processing times are not particularly large and can only be seen in the beginning of the experiment (i.e., trial effects, or Early-Late effect).

## 2. Method

### 2.1. Preregistration

This study’s sample size, materials, hypotheses, and planned analyses were preregistered on Open Science Framework (<https://osf.io/zbjn>) prior to data collection. In addition, analysis scripts and data are available via OSF: <https://osf.io/e2ysx/>

### 2.2. Participants

A total of 70 children and young adults participated in the study. The children were 35 4th grade children (15 female,  $M = 10:4$  years,  $SD = 0:3$ , range 9:8–10:8) recruited from two classrooms in two different Finnish elementary schools in Varsinais-Suomi region. At the time of testing, they had received approximately three years and three months of formal reading instruction. All the children were native Finnish speakers, had normal or corrected-to-normal vision, and no known reading difficulties.

We assessed children’s abilities that have been shown to affect children’s irony comprehension: level of language skill (i.e., reading comprehension and technical reading skill), working memory capacity, and empathy skill (e.g., Filippova & Astington, 2008; Godbee & Porter, 2013; Nicholson et al., 2013). We measured these to establish that children participating into the study were within typical range for these skills and the results would reflect normative development, and to facilitate comparison with samples in future developmental studies. Technical reading skill was measured using the word fluency subtest of Lukilasse II (Häyrinen, Serenius-Sirve, & Korkman, 2013), in which the children had to read correctly as many words as possible from a list of 105 words within 120-s time-limit. The average score was 81.80 ( $SD = 12.74$ , range 54–104). Reading comprehension was measured using the Maze task (Ronimus, Tolvanen, & Hautala, 2022). The task was comprised of 16 texts with 4 words missing from each text. Based on the textual cues, participants had to choose a correct word for each missing word from four options. The average score was 41.53 ( $SD = 7.76$ , range 27–56). Children’s working memory capacity was assessed with the Digit Span subtest of WISC-IV (Wechsler, 2010). The average raw score was 13.03 ( $SD = 2.04$ , range 10–21). Last, children’s empathy skill was assessed using the Index of Empathy for Children and Adolescents (Bryant, 1982) that was translated in Finnish for this experiment (Cronbach’s  $\alpha = 0.64$ , 95% CI [0.48, 0.80]). The average score was 12.94 ( $SD = 2.79$ , range 5–20). Children’s parents signed a written informed consent form prior to the experiment, and verbal assent was asked from each child upon arrival to the experiment. The children received candy or stickers for their participation.

In addition, 35 University of Oulu students (27 women,  $M = 24$ ,  $SD = 4.52$ , range 19–38) participated and received a cafe voucher for their participation. Adults only completed the written irony reading task. All

the adult participants were also native Finnish speakers and had normal or corrected-to-normal vision. Each participant gave written consent.

The study was conducted in accordance with the Declaration of Helsinki. The Ethics Committee for Human Sciences at the University of Turku approved the study.

### 2.3. Apparatus

Eye movements were recorded using EyeLink Portable Duo, and EyeLink 1000 Plus eye-trackers (SR Research Ltd. Ontario, Canada) at 500 Hz sampling frequency. EyeLink Portable Duo was used with children, and EyeLink 1000 Plus with adults. With Portable Duo, the stimuli were presented on a 17.3” Asus ROG G752V laptop monitor, and participants were seated 60 cm from the screen. With EyeLink 1000 the stimuli were presented on a 24” Asus VG248QE monitor, and due to larger monitor size, they were seated 90 cm from the screen. With both monitors, refresh rate of 120 Hz and a resolution of 1920 × 1080 pixels were used. Chin-and-forehead rest was used to stabilize the head of the participant.

### 2.4. Materials

In the study, each participant read a total of 26 experimental stories on a computer screen (font: Courier New, font size: 27, line height: 3) while their eye movements were recorded. Half of the stories contained literal, and the other half ironic, target phrases. There were literal and ironic versions of each story (26 stories × 2 text types, resulting in 52 experimental stories). Each story consisted of 4–5 sentences, and their length varied between 20 and 41 words ( $M_{Words} = 29.33$ ,  $SD_{Words} = 4.53$ ) (see Table 1 for an example). The stories started with context sentences: First, there were 1–2 context sentences that gave general description in which the events of the story happened, and these were the same in both literal and ironic versions ( $M_{Words} = 9.04$ ,  $SD_{Words} = 3.63$ ). Second, there was a context sentence that contained critical information (i.e., critical context) that made the following phrase either ironic or literal (for ironic versions  $M_{Words} = 9.31$ ,  $SD_{Words} = 3.08$ ; and literal versions ( $M_{Words} = 9.15$ ,  $SD_{Words} = 2.88$ ). Context sentences were followed by a target phrase that was the same in both literal and ironic versions of the stories ( $M_{Words} = 4.24$ ,  $SD_{Words} = 0.86$ ), but could be interpreted as literal or ironic depending on critical context. The target phrase was followed by a spillover region (see e.g., Rayner & Duffy, 1986) describing who had uttered the target phrase ( $M_{Words} = 2.27$ ,  $SD_{Words} = 0.60$ ). Last, every story ended with a neutral sentence describing how the story events ended ( $M_{Words} = 5.46$ ,  $SD_{Words} = 1.29$ ). The spillover region and the end of the story sentences were the same between literal and ironic versions.

**Table 1**

An example of an experimental story and text memory and inference questions translated from Finnish.

Story type	Region	Story content
Literal	Beginning	Onni and Alekski are in the school canteen.
	Critical Context	Onni gathers his plate full and eats up fast all the food.
	Target Phrase	“Well, you were hungry,”
	Spillover Region	Alekski says to him.
	End	The boys need to hurry to the next lesson.
Ironic	Beginning	Onni and Alekski are in the school canteen.
	Critical Context	Onni gathers a little food on the plate but eats only a small portion of it.
	Target Phrase	“Well, you were hungry,”
	Spillover Region	Alekski says to him.
	End	The boys need to hurry to the to the next lesson.
Questions	Inference Question	Did Alekski think that Onni was hungry?
	Text Memory Question	Were the boys at home?

Another function for the first and last sentences of the story was to prevent the target phrase and spillover region being the first or last sentence of the story, as these are typically read differently. Only one version of each story was shown to the participants. The presentation order of the stories was randomized. Participants' memory for the story context and comprehension of the intended meaning of the target statement were also assessed after reading of each story (examples of the questions are presented in Table 1). For both types of questions, correct answers were counted, and the proportion of correct answers was computed.

In preparation for the main study, two separate rating studies were conducted to test: (1) how familiar the target statements were as ironic in comparison with literal meaning and how natural the experimental stories were, and (2) how accurately ironic intent was comprehended in the experimental items by children. An Internet survey tool was used to collect the data (Webropol, [www.webropol.com](http://www.webropol.com)).

A group of 46 adults ( $M_{Age} = 34.36$ ,  $SD_{Age} = 11.55$ ; 40 females, 5 males) participated in the first rating study. All were native Finnish speakers and did not take part in the actual experiment. Raters evaluated (from 0 = not familiar at all to 10 = very familiar) ironic target phrases ( $M = 2.43$ ,  $SD = 1.50$ ) as less familiar compared to literal target phrases ( $M = 5.76$ ,  $SD = 1.37$ ),  $t(25) = -5.50$ ,  $p < .001$ ,  $d = 0.93$ . As for naturalness, raters evaluated (from 0 = not natural at all to 10 = very natural) ironic versions of the stories ( $M = 6.63$ ,  $SD = 1.16$ ) being equally natural compared to literal versions ( $M = 7.12$ ,  $SD = 1.17$ ),  $t(25) = -1.02$ ,  $p = .063$ ,  $d = 0.42$ .

A group of 18 children aged 7–14 ( $M_{Age} = 10.28$ ,  $SD_{Age} = 1.81$ , 5 males, 12 females, and 1 who chose not to answer) participated in the second rating study tapping into the comprehensibility of the experimental items. All the children were native Finnish speakers, and none of them participated in the actual experiment. Proportion of correct answers to comprehension questions for literal target phrases was higher ( $M = 0.92$ ,  $SD = 0.27$ ) than for ironic target phrases ( $M = 0.63$ ,  $SD = 0.48$ ). However, the accuracy scores were in expected range based on previous studies with children (e.g., Nicholson et al., 2013).

## 2.5. Procedure

Participants were tested individually. Upon arrival, participants were informed that the experiment assessed reading. The specific nature of the experiment was explained to participants when the experiment was over. Before the reading task, the eye-tracking system was introduced to each participant, and the experimental procedure was explained. The eye-tracker then was set up and calibrated using a 9-point calibration screen. Participants were instructed to read each story for comprehension at their own pace. Each story was presented on one screen. Participants were told to press the space bar on the keyboard when they finished reading the paragraph. After each story, two questions were presented one at a time. Participants answered the questions by pressing designated "Yes" and "No" buttons on the keyboard. After the participant answered the second question, the next story was presented. For children, the reading task was followed by the Digit Span, word fluency test, and Index of Empathy for Children and Adolescents. The experimental sessions lasted about 30–50 min. The reading comprehension task (i.e., the Maze task) was completed by children in their class group after the eye movement registrations.

## 3. Results

### 3.1. Preprocessing of data

Fixations shorter than 50 ms were either merged with a nearby fixation (if the distance between the fixations was  $<1^\circ$ ) or removed from the data. Sentence-level measures were calculated for the target phrase, the critical context, and the spillover region from the eye movement data (Hyönä, Lorch, & Rinck, 2003). *First-pass reading time* is the summed

duration of fixations made within the sentence during first reading of the sentence. First-pass reading time was further divided into *forward-fixation time* (i.e., summed duration of fixations landing on unread parts of the sentence during first-pass reading) and *number of first-pass rereading fixations* (i.e., summed number of fixations made re-inspecting a sentence before moving on in the text). Number of *look-back fixations* is the summed number of fixations returning to the sentence from other parts of text made after the first-pass reading, and *number of look-from fixations* is the summed number of look-back fixations that were initiated from the sentence. All the measures were analyzed for the target phrase, and first-pass reading time was analyzed for the spillover region. As for the critical context, the content of the critical context sentences was not tightly matched between Story Types (literal vs. ironic), which made reliably comparing reading times not possible, consequently, probability to look-back (binomial measure) was analyzed.

The exclusion criterion for participants' comprehension accuracy was that they should have more than one correctly comprehended ironic item. This criterion was chosen instead of using data-driven cut-off point (e.g.,  $\pm 2SD$ ), as this would have left participants with only a single correctly comprehended item in the data. Seven participants (2 children and 5 adults) had *below acceptable level* irony comprehension accuracy, and their data were excluded from the analyses, resulting in a total of 63 participants (33 children and 30 adults). Observed means and standard deviations of the measures are presented in Table 2.

### 3.2. Analyses

The data were analyzed with linear or generalized linear mixed-effects models (Baayen, Davidson, & Bates, 2008) using R statistical software (Version 4.1.2; R Core Team, 2021), except for the dependent variables with the problem of zero-inflation (i.e., number of first-pass rereading fixations, number of look-back fixations, number of look-from fixations, and probability to look-back to critical context) which were analyzed using glmmTMB package (Brooks et al., 2017).

Separate models were built for each eye movement measure for the different text regions (target sentence, critical context, and spillover region), and for comprehension and text memory question accuracies. *Story Type* (Literal vs. Irony) and *Age Group* (Adults vs. Children) were fitted in the models as successive difference contrast coded fixed effects. *Trial Order* (i.e., order in which the items were presented) was fitted in models as a centered continuous variable. Last, effect of *Item-Comprehension* (i.e., whether intended meaning of the target phrase was comprehended Correctly vs. Incorrectly) was added to the models as a treatment coded fixed effect variable, in which correct answer was set as baseline. Thus, the model intercepts, in which *Item-Comprehension* is fitted, assess reading in which intended meaning of the target phrase was correctly comprehended (see Schad, Vasishth, Hohenstein, & Kliegl, 2020 for review of contrasts), and not reading time over correctly and incorrectly comprehended items (see Ferreira & Yang, 2019 for discussion about comprehension). The *Item-Comprehension* measure was fitted only in models concerning reading. The text memory question accuracy values were close to ceiling across the conditions, consequently, trial order was not fitted into the respective model to keep them as simple as possible. In the rest of the models, we explored three-way interactions that included *Story Type*. Participants and items were fitted into the models as random intercepts, and *Story Type* and *Age Group* were fitted as random slopes. If a model failed to converge, it was trimmed top-down starting with removing covariance between random effects (Barr, Levy, Scheepers, & Tily, 2013; Brauer & Curtin, 2018). The reading time measures were skewed and consequently logarithmically transformed (best fitting transformation was selected for the data using Box-Cox Power transform). As for the zero-inflated models, the predictors to be fitted in the zero-inflation part of the models were determined by comparing models with each other. We started with a model with only an intercept, followed by models in which we added Age

**Table 2**  
Descriptive statistics of the reading and comprehension measures for both story types and age groups.

Region	Measure	Adults				Children			
		Literal		Irony		Literal		Irony	
		M	SD	M	SD	M	SD	M	SD
Target Phrase									
	First-pass reading time	844	440	862	421	1997	1199	1949	1288
	Forward fixation time	709	294	731	304	1489	672	1422	652
	Number of first-pass rereading fixations	1.98	1.58	1.75	1.44	3.05	3.12	3.18	3.13
	Probability to first-pass reread	0.38	0.49	0.40	0.49	0.74	0.44	0.74	0.44
	Number of look-back fixations	3.59	3.38	3.71	3.06	4.66	3.64	4.90	3.97
	Probability to look-back	0.52	0.50	0.53	0.50	0.18	0.39	0.16	0.37
	Number of look-from fixations	4.08	4.28	3.73	3.15	3.75	2.98	4.00	3.93
	Probability to look-from	0.42	0.49	0.44	0.50	0.13	0.34	0.13	0.33
Spillover Region									
	First-pass reading time	449	313	462	288	901	643	882	667
Critical Context									
	Probability to look-back	0.63	0.48	0.59	0.49	0.21	0.41	0.16	0.37
Questions									
	Correct Inference	0.97	0.17	0.86	0.34	0.89	0.31	0.50	0.50
	Text memory	0.96	0.21	0.98	0.15	0.96	0.20	0.94	0.23

Note. Reading time measures are reported in ms. Inference and text memory question accuracies reported are proportion of correct answers.

Group (as it was the main cause of zero inflation), Age Group  $\times$  Story Type, and finally all the predictors and their three-way interactions. The model providing the best fit to the data (lowest AIC value) was selected.

The exact degrees of freedom were difficult to determine for the  $t$ - and  $z$ -statistics estimated by mixed-effects models, leading to a problem determining exact  $p$ -values (Baayen et al., 2008). Consequently, degrees of freedom or  $p$ -values are not reported; statistical significance at the 0.05 level is indicated by values of  $t$  and  $z > |1.96|$ . For the sake of brevity, only significant main effects of Story Type or Age Group, and interactions involving Story Type are reported. Significant main effects are reported in text and for interactions the model estimates, and their 95% CIs, are illustrated in figures. All the model summaries are reported in Appendix A Tables A1 – A9.

### 3.3. Changes to the statistical analyses in comparison to pre-registration

There were some changes that we needed to make for the pre-registered analyses. First, in the rereading measures (i.e., first-pass rereading time, look-back time, look-from time, and probability to look-back to critical context) there were more zero-values than expected. This made it problematic to analyze the data using linear or traditional generalized linear mixed-effect models (see e.g., Baayen et al., 2008). To be able to take this into account, we analyzed the number of fixations instead of rereading times using *glmmTMB* package (Brooks et al., 2017) in order to be able to analyze count data (e.g., number of fixations) that have higher number of zero values. In sentence-level processing fixation times and number of fixations are tightly related (in our sample, correlations between reading times and number of fixations ranged between  $r = 0.96$ – $0.98$ ). Second, while pre-registering the experiment we failed to mention the analysis of trial effects, although it was likely that they would occur with adults (see Olkonieni & Kaakinen, 2021 for review). Consequently, in order to reliably compare reading of irony between adults and children trial effects were taken into account. Last, as comprehension accuracies were lower than expected (see Table 2), we decided to take the Item-Comprehension into account in the models. This solution has two advantages: First, we are able to explore how successful vs. non-successful comprehension of the intended meaning of the target phrase is reflected in the reading data. This has not been previously possible, as all the previous experiments on the processing of written irony have been conducted with adults. Second, this allowed us to use all the items, thus, strengthening the models. Finally, as adults showed variance in their irony comprehension accuracies (see Table 2), we were also able to compare adults to children in the models concerning relationships

between irony comprehension accuracy and reading.

### 3.4. Comprehension and text memory questions

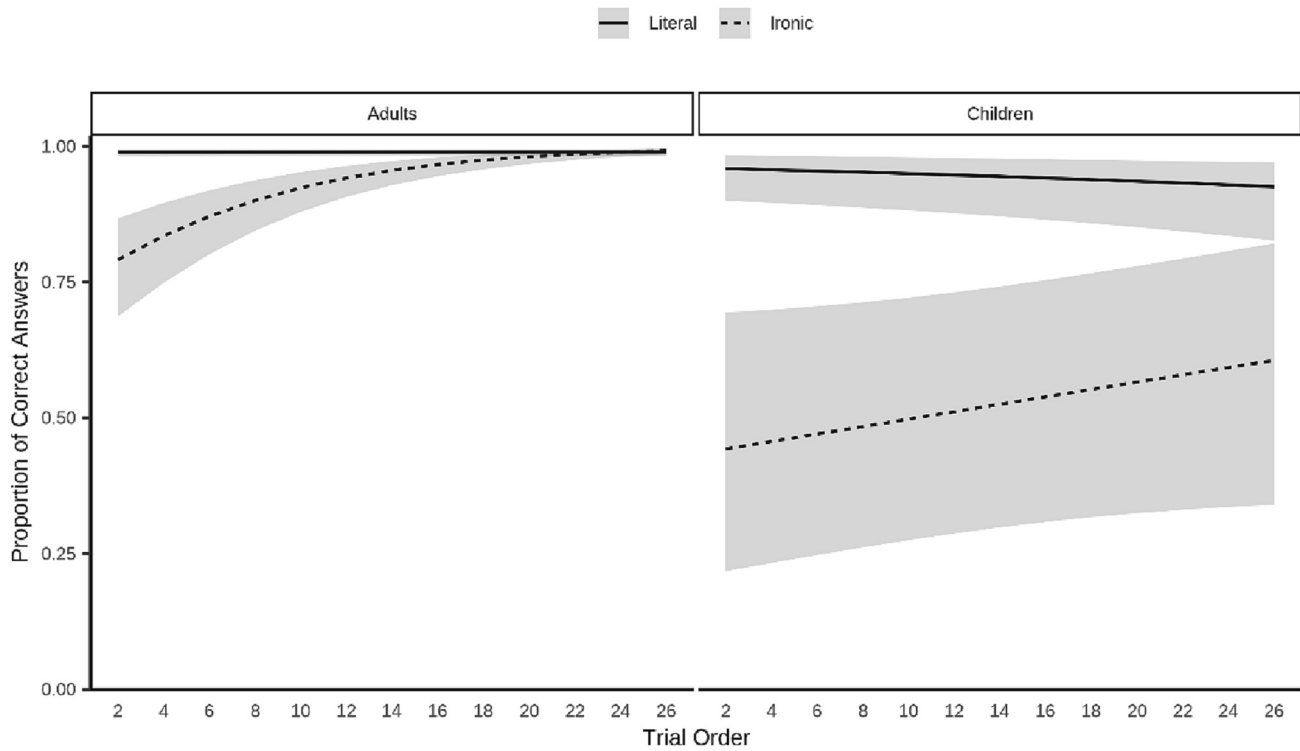
The descriptive statistics for proportions of correct answers on text memory and inference questions are presented in Table 2. Overall, we observed a ceiling effect in text memory question performance, suggesting that all participants were attentive to the task. Consequently, the model for accuracy in answering text memory questions did not show effects of Story Type (literal vs. irony), Age Group (adults vs. children), or their interaction (see Table A1).

The model for comprehension question accuracy showed two main effects (see Table A2). First, there was an effect of Story Type (literal vs irony), indicating that the ironic phrases were harder to comprehend than their literal counterparts,  $\beta = -2.16$ , 95% CI  $[-2.34, -1.97]$ ,  $z = -23.09$ . Second, there was an effect of Age Group, indicating that children showed lower comprehension accuracy than adults,  $\beta = -2.30$ , 95% CI  $[-3.14, -1.46]$ ,  $z = -5.38$ . These effects were qualified by a three-way interaction between Story Type, Age Group, and Trial Order,  $\beta = -0.67$ , 95% CI  $[-1.03, -0.32]$ ,  $z = -3.70$ . This interaction indicates that adults had lower proportion of correct answers to comprehension questions for ironic than literal target statements in the beginning of the experiment, but that the effect wore off toward the end of the experiment (see Fig. 1). Similarly, children showed lower accuracy in ironic than literal target phrases in the beginning of the experiment but showed less improvement toward the end than adults. Overall, children showed lower comprehension accuracy scores than adults, and the difference was larger for ironic than literal phrases.

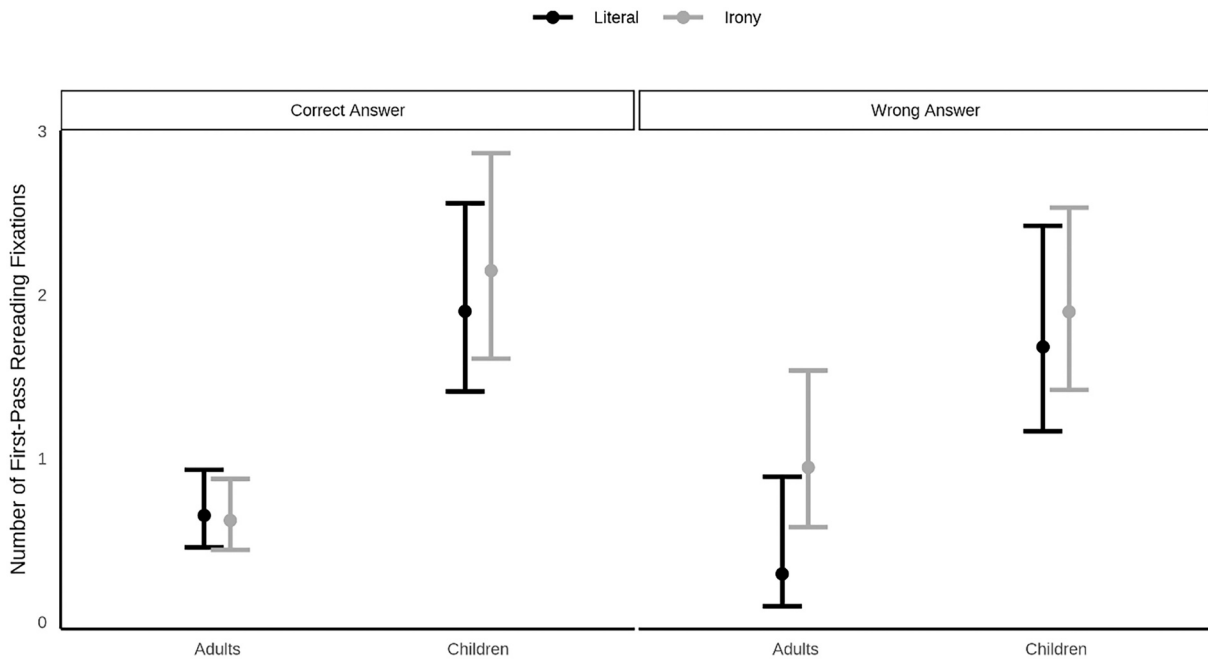
### 3.5. Reading of the ironic and literal stories

The model for first-pass reading time on the target phrase revealed an effect of Age Group (see Table A3). This effect indicates that children showed longer first-pass reading times than adults, when the intended meaning was correctly comprehended,  $\beta = 0.81$ , 95% CI  $[0.66, 0.95]$ ,  $t = 11.14$ . The model did not show effects of Story Type (literal vs. irony). Similarly, the model for forward-fixation time on the target phrase revealed an effect of the Age Group (see Table A4), indicating that children showed longer forward-fixation times than adults,  $\beta = 0.69$ , 95% CI  $[0.56, 0.83]$ ,  $t = 10.31$ . This model did not show effects related to Story Type either.

The model for number of first-pass rereading fixations on the target phrase showed an effect of Age Group (see Table A5). This indicates that children did more first-pass rereading than adults, when the



**Fig. 1.** Model estimates for the comprehension question accuracy. Note. Model values are back-transformed from log-values. Shaded areas represent 95% confidence intervals.



**Fig. 2.** Model estimates for the number of first-pass rereading fixations. Note. Y-axis values are model estimates that are back-transformed from log-values, and error bars represent 95% Confidence Intervals.

intended meaning was correctly comprehended,  $\beta = 1.15$ , 95% CI [0.81, 1.50],  $z = 6.56$ . The model also revealed a three-way interaction between Story Type, Age Group, and Item-Comprehension,  $\beta = -1.21$ , 95% CI [-2.39, -0.03],  $z = -2.02$ . This interaction indicates that when the comprehension question was answered correctly, both children and adults showed no difference in the number of first-pass rereading

fixations between literal and ironic target phrases (see Fig. 2). However, when the comprehension question was answered incorrectly, adults did more first-pass rereading for ironic than literal target phrases, whereas children showed no difference between Story Types. In addition, zero-inflation model for first-pass rereading fixations showed an effect of Age Group, indicating that children had a lower

probability of making exactly zero first-pass rereading fixations than adults,  $\beta = -1.03$ , 95% CI [-1.67, -0.39],  $z = -3.17$ .

The model for *number of look-back fixations on the target phrase* revealed a three-way interaction between Story Type, Age Group, and Trial Order,  $\beta = 0.50$ , 95% CI [0.24, 0.76],  $z = 3.30$  (see Table A6). This interaction indicates that, in the beginning of the experiment, adults made higher numbers of look-backs to ironic than literal target phrases, but the effect turned the other way around toward the end of the experiment (see Fig. 3). Children showed an opposite trend making higher numbers of look-backs to literal than ironic target phrases in the beginning of the experiment, but the effect turned the other way around toward the end of the experiment. In addition, the zero-inflation model for *look-back fixations on the target phrase* showed a main effect of Age Group, indicating that children had a higher probability of making exactly zero look-back fixations than adults,  $\beta = 2.09$ , 95% CI [1.80, 2.39],  $z = 13.82$

The model for *number of look-from fixations made from the target phrase* showed an interaction between Story Type and Trial Order,  $\beta = -0.27$ , 95% CI [-0.45, -0.08],  $z = -2.81$ . This indicates that in the beginning of the experimental session readers made higher numbers of look-foms from the ironic than literal target phrase, but the effect turned the other way around toward the end of the experiment (see Fig. 4A). In addition, the zero-inflation model for *look-from fixations from the target phrase* showed a main effect of Age Group, indicating that children had a higher probability for making exactly zero look-back fixations than adults,  $\beta = 1.74$ , 95% CI [1.43, 2.05],  $z = 11.08$ .

The model for *first-pass reading time on the spillover region* showed an effect of Age Group (see Table A8). This effect indicates that children showed longer reading times in comparison to adults, when the intended meaning of the target phrase was correctly comprehended,  $\beta = 0.68$ , 95% CI [0.54, 0.83],  $z = 9.32$ . The model showed no main effect of Story Type, nor interactions related to Story Type.

Last, the model for *probability to look-back to critical context* revealed two main effects (see Table A9). First, there was an effect of Story Type, indicating that readers had a lower probability of look-backs to critical context of ironic than literal stories when the intended meaning of the target phrase was correctly comprehended,  $\beta = -0.34$ , 95% CI [-0.68, -0.003],  $z = -1.98$ . Second, there was an effect of Age Group, indicating that children were less likely to look-back to critical context from

subsequent text parts than adults (when the target phrase was comprehended correctly),  $\beta = -2.58$ , 95% CI [-3.37, -1.79],  $z = -6.39$ . Moreover, the model revealed an interaction between Story Type and Trial Order,  $\beta = -0.68$ , 95% CI [-1.03, -0.32],  $z = -3.76$ . This indicates that in the beginning of the experiment readers showed slightly higher probability of looking back to critical context of ironic than literal stories, but this effect turned the other way around toward the end (see Fig. 4B).

### 3.6. Relationships between reading of ironic texts and irony comprehension

The results showed that there was large variance in irony comprehension accuracies between participants, especially in children. In addition, successful vs. non-successful comprehension of ironic and literal phrases was related to processing of irony. However, we do not know how individuals' irony comprehension ability is related to processing of ironic meaning. Thus, we explored the question of whether better irony comprehenders exhibit different processing patterns for ironic stories. Relationships between reading of ironic stories and irony comprehension were analyzed by building separate models in which only reading data of correctly interpreted ironic items were considered. As a new variable, *Comprehension Accuracy* of the participants was fitted into the models as a continuous variable. Before fitting Comprehension Accuracy, the measure was scaled per Age Group as good vs. poor comprehension accuracy was different for children and adults (see Table 2). For the sake of brevity, only significant effects involving Irony Comprehension Accuracy and interactions between Age Group and Irony Comprehension Accuracy are reported. All the final models are reported in the Appendix B Tables B1 – B7.

The model for *first-pass reading time on ironic target phrases* did not show an effect of Irony Comprehension Accuracy (see Table B1). However, the model revealed an interaction between Irony Comprehension Accuracy and Age Group,  $\beta = -0.24$ , 95% CI [-0.38, -0.10],  $z = -3.41$ . This interaction indicates that when adults' irony comprehension accuracy increased their first-pass reading time also increased, but for children the opposite was true (see Fig. 5A). Also, the model for *forward-fixation time on the ironic target phrases* showed no effect of Irony Comprehension Accuracy (see Table B2), but revealed an interaction between Irony Comprehension Accuracy and Age Group,  $\beta = -0.22$ ,

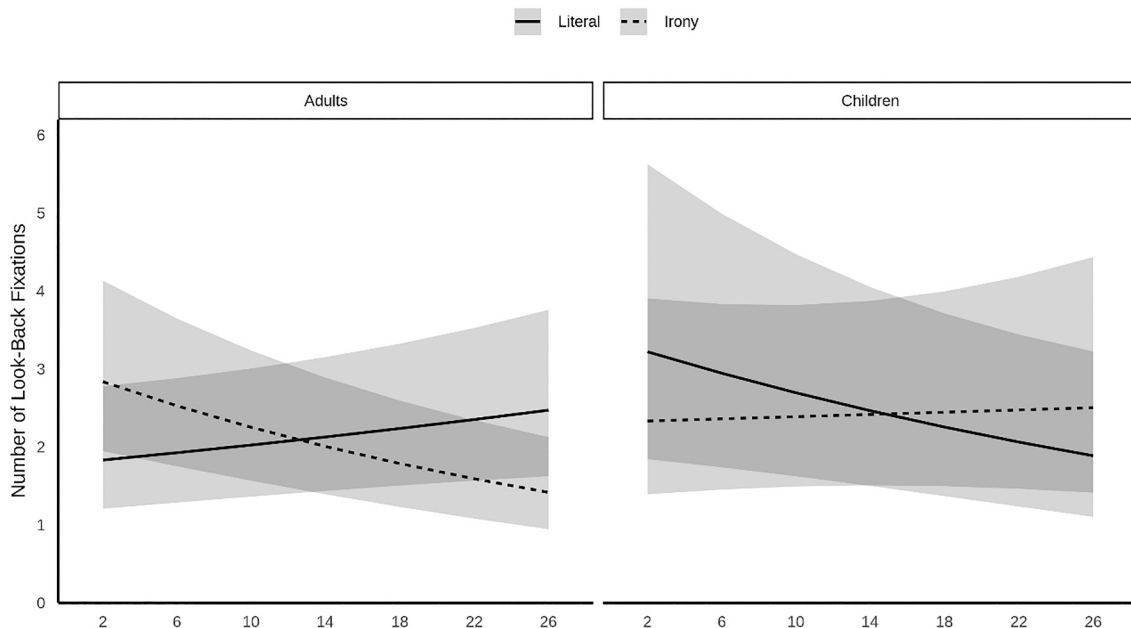
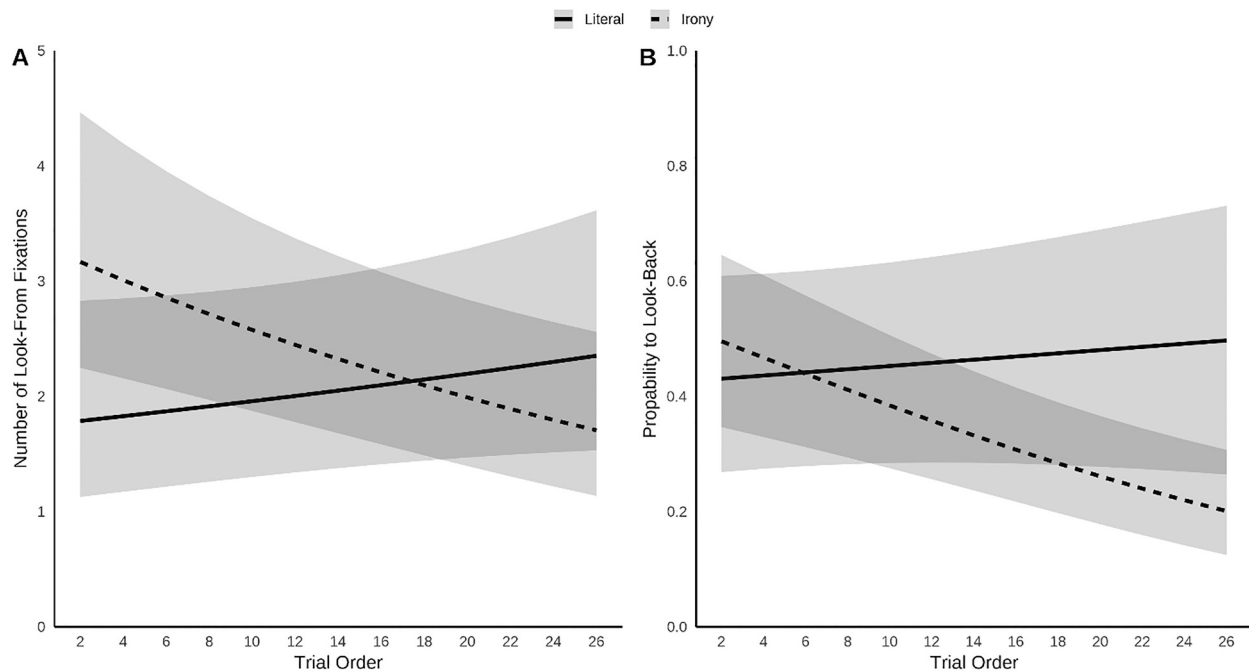


Fig. 3. Model estimates for the number of look-back fixations to target phrase. Note. Y-axis values are model estimates that are back-transformed from log-values, and error bars represent 95% Confidence Intervals.



**Fig. 4.** Model estimates for the number of look-from fixations from target phrase and probability to look-back to critical context.

Note. Panel A: The interaction between Story Type and Trial Order on the number of look-from fixations made from the target phrase. Panel B: The interaction between Story Type and Trial Order on the probability to look-back to critical context. In both panels, Y-axis values are model estimates that are back-transformed from log-values, and shaded areas represent 95% Confidence Intervals.

95% CI  $[-0.35, -0.08]$ ,  $z = -3.20$ . This interaction was similar to the interaction revealed by the model on first-pass reading time (see Fig. 5B). Similar to previous models, the model for *number of first-pass rereading fixations on ironic target phrase* did not show an effect of Irony Comprehension Accuracy (see Table B3), but revealed an interaction between Comprehension Accuracy and Age Group,  $\beta = -0.42$ , 95% CI  $[-0.79, -0.04]$ ,  $z = -2.17$  (see Fig. 5C).

The model for *number of look-back fixations to ironic target phrase* did not show an effect of Irony Comprehension Accuracy or an interaction between Comprehension Accuracy and Age Group (see Table B4). The model for *number of look-from fixations made from the ironic target phrase* did not show an effect of Irony Comprehension Accuracy or an interaction between Comprehension Accuracy and Age Group (see Table B5). The model for *first-pass reading time on the spillover region following ironic target phrase* did not show a main effect of Irony Comprehension Accuracy (see Table B6). However, the model revealed an interaction between Comprehension Accuracy and Age Group,  $\beta = -0.29$ , 95% CI  $[-0.44, -0.14]$ ,  $z = -3.85$ . This indicates that when adults' irony comprehension accuracy increased their first-pass reading time on spillover region also increased, but for children the opposite was true (see Fig. 5D). Last, the model on the *probability to look-back to critical context of the ironic stories* did not show a main effect of Irony Comprehension Accuracy nor interaction between Comprehension Accuracy and Age Group (see Table B7).

#### 4. Discussion

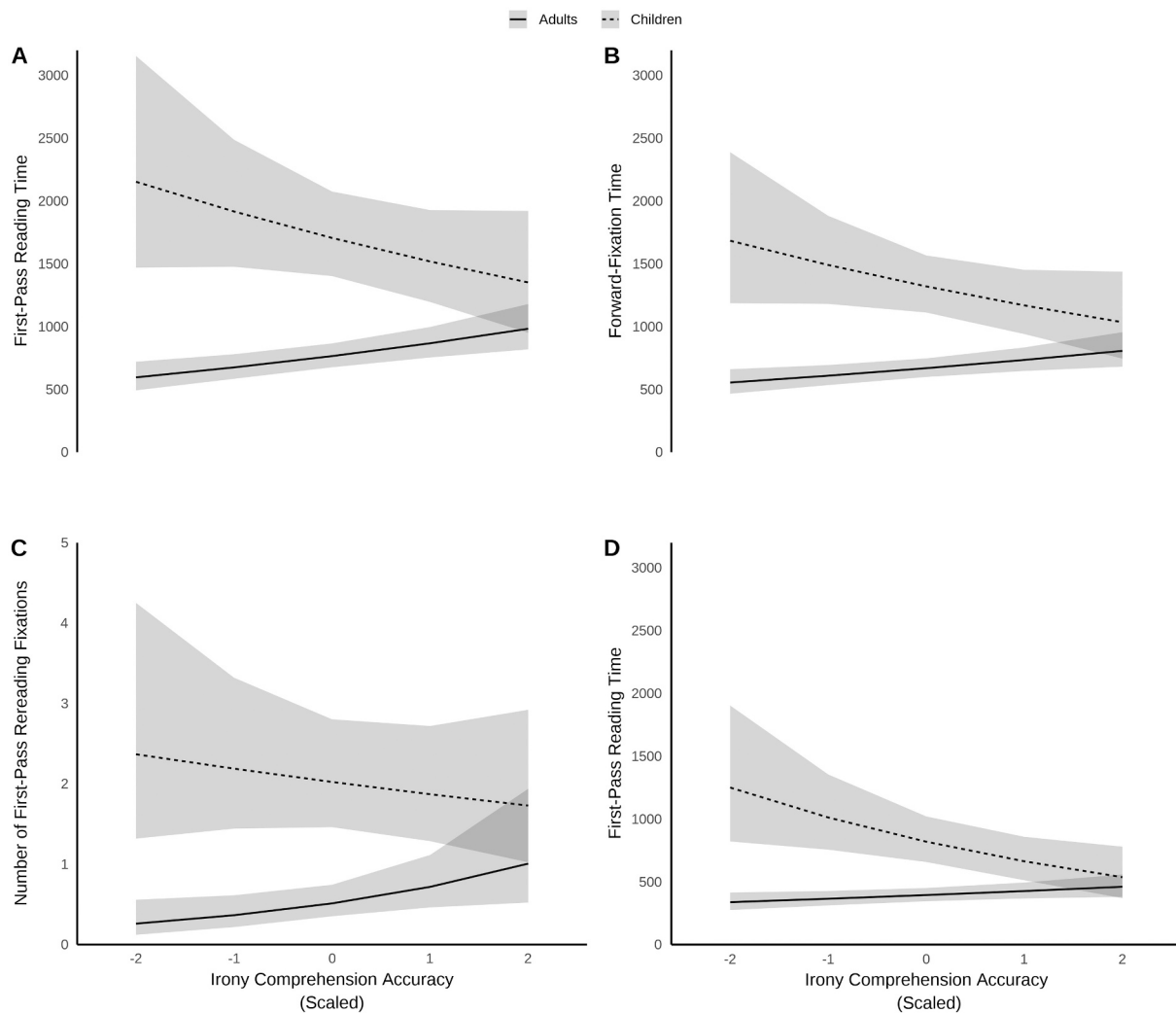
For the first time, by employing eye-tracking, we explored how children process and comprehend written irony in comparison to adults. We expected 10-year-old children to show lower comprehension of intended meaning of ironic but not literal statements when compared to adults. Moreover, we expected adults to show more re-reading of ironic than literal phrases, and that this irony effect would be most apparent in the beginning of the experiment and would wear off toward the end of

the experiment. We expected children to struggle in their irony comprehension more than adults and assumed that this struggle would be associated with more pronounced rereading of the ironic stories compared to adults.

##### 4.1. Processing of irony

As expected, adults showed increased later rereading of ironic phrases in comparison to literal, and they also made more returns to already read text parts (i.e., look-firms) from the ironic than literal phrase. It seems that these returns were predominantly made to critical context of ironic stories, as the probability to look-back to the critical context of the ironic stories was higher when compared to literal. Moreover, these effects were mediated by trial effects. In other words, increased rereading of ironic stories was seen in the beginning of the experiment, but this effect turned other way around toward the end. These results for adults replicate previous findings of eye-tracking studies on processing written irony (see Olkonieni & Kaakinen, 2021 for review).

Like adults, children also made higher numbers of returns from the ironic than literal phrase and showed higher probability to look-back to the critical context of the ironic stories. Surprisingly, these effects did not differ between adults and children, and children also showed trial effects for irony. This suggest that both children and adults needed extra processing in order to integrate ironic target phrases with context, consistent with traditional theories on irony comprehension (e.g., Grice, 1975). However, repeated exposure to irony seemed to reduce this need. This trial effect result is similar to that of previous studies with adults (Olkonieni & Kaakinen, 2021; Spotorno & Noveck, 2014), and suggests that children also become sensitized by the repeated exposure of irony and are able to some extent to improve their performance. However, this sensitization was not similar in all the measures, as trial effect for irony was only observed for adults in look-back fixations to ironic phrase. It should be also noted that the improvement in reading times was not a



**Fig. 5.** Model estimates for the interactions between irony comprehension accuracy and age group.

Note. Panel A: The interaction between Irony Comprehension Accuracy and Age Group on the first-pass reading time on the correctly interpreted ironic target phrase. Panel B: The interaction between Irony Comprehension Accuracy and Age Group on the forward-fixation time on the correctly interpreted ironic target phrase. Panel C: The interaction between Irony Comprehension Accuracy and Age Group on the number of first-pass rereading fixations on the correctly interpreted ironic target phrase. Panel D: The interaction between Irony Comprehension Accuracy and Age Group on the first-pass reading time on the spillover region following the correctly interpreted ironic target phrase. In all the panels y-axis values are model estimates that are back-transformed from log-values, and shaded areas represent 95% Confidence Intervals.

clear indication of better comprehension, as children showed a relatively minor increase in comprehension accuracy during the course of the experiment. This might be due to the fact that, overall, children's probability to do later rereading was much smaller than for adults.

For adult readers, but not for children, there was a relationship between comprehending the intended meaning of the target phrase and early and later rereading of that phrase. This suggests that adults may have higher sensitivity for difficulties in categorizing the target phrase as being ironic or not. This could be interpreted, as we had hypothesized, as reflecting children's insensitivity for intended ironic meaning. However, it is also possible that reading for comprehension is overall more effortful for children (e.g., Kaakinen et al., 2015). Because of this, it might be that the irony effect is difficult to detect in children's reading, as it may be masked by the higher general effort that children need to invest for the reading task. This possibility is at least partly supported by our data. First of all, children who had higher irony comprehension accuracy showed faster first-pass reading times than those with poorer comprehension accuracy. In other words, those children who are better at categorizing phrases as being ironic need not invest as much processing effort as others. The opposite was true for adults who showed

higher comprehension accuracy, suggesting that for adults reading in general is very efficient and irony-related processing is clearly evident in additional processing. This is in line with previous studies on adults' written irony processing (see Olkonieni & Kaakinen, 2021, for a review).

For both children and adults, better irony comprehension was not associated with the number of look-fors made from the ironic target phrase. This contrasts with previous findings by Olkonieni, Johander, and Kaakinen (2019), who found in their eye-tracking study on processing written irony that the probability of making a look-from from the target phrase had a negative correlation with comprehension accuracy. However, differences in the tasks used (moving window paradigm was used by Olkonieni, Johander, & Kaakinen, 2019, which prevented natural reading) and the age groups (Olkonieni, Johander, & Kaakinen, 2019 studied adults only) makes comparing the results of these studies difficult. It is evident that more research is needed to explore the role of later rereading on irony comprehension.

#### 4.2. Irony comprehension

As predicted, 10-year-old children showed lower irony comprehension accuracy than adults, replicating previous findings (e.g., Ackerman, 1983; Capelli et al., 1990; Climie & Pexman, 2008; Dews et al., 1996; Filippova & Astington, 2008; Köder & Falkum, 2021). However, children's irony comprehension accuracy was lower than what was reported in several previous studies using non-written materials containing irony (e.g., Banasik-Jemielniak et al., 2020; Dews et al., 1996; Filippova & Astington, 2008; Köder & Falkum, 2021; Loukusa & Leinonen, 2008; Pexman et al., 2005). This raises the question of whether comprehending written irony is harder for children compared to, for example, when irony is spoken and contains intonation cues as well as picture, video, or puppet show illustration, as in the previous studies with children (e.g., Climie & Pexman, 2008; Filippova & Astington, 2008; Harris & Pexman, 2003).

The answer to this question is not that clear: First, it is important to note that this issue was not directly explored in the present study, so strong conclusions cannot be drawn from the results of this experiment. Second, some of the previous studies on children of the same age have shown lower, or about the same, levels of irony comprehension than what was observed in the present study, despite tone of voice information and illustration (e.g., Capelli et al., 1990; Zajaczkowska & Abbot-Smith, 2020). For example, in their study, Zajaczkowska and Abbot-Smith (2020, Exp 2) presented short video-recorded dialogues containing irony to 10–12-year-old children. Their results showed that average comprehension accuracy for ironic materials that needed social reasoning (they called this complex irony) was on average 55% with accuracies ranging between 0 and 100%. This suggests that task demands in previous studies might have varied, so it is hard to tell how much differences are due to, for example, demands of spoken vs. written modalities and how much to other task demands. Moreover, in our study, like in the study by Zajaczkowska and Abbot-Smith (2020), there was large variance in children's irony comprehension accuracies. This shows that it should not be expected that children's irony comprehension ability would develop in an invariant manner. Rather these results suggest that there are large individual differences (see also Loukusa & Leinonen, 2008) and that the ability to comprehend irony develops for each child at a somewhat individual pace. As we know that there are also individual differences among adults in processing and comprehending irony (see e.g., Kaakinen et al., 2014), this is not surprising.

As mentioned, current theories of irony comprehension do not explicitly take developmental changes into account. It is possible, however, to accommodate developmental differences in theories that take individual differences into account, such as the *parallel constraint-satisfaction framework* (Pexman, 2008) and the *predictive coding theory*

of irony (Fabry, 2021). These theories could be extended to make the prediction that as irony comprehension ability is still developing in children, they should show more difficulty in processing and comprehending irony than adults. In future studies, those theories could be further refined by examining which developing social, cognitive, and linguistic skills are related to children's irony processing. The costs of irony processing may be higher for children, but there is still a great deal to learn about the factors that mitigate that cost.

#### 5. Conclusion

The present study explored, for the first time, children's processing and comprehension of written irony and compared children's performance to that of adults. The results showed that for both children and adults comprehending irony poses a greater cost than processing of its literal counterpart. Moreover, it seems that for 10-year-old children reading for comprehension is more demanding than it is for adults. Nonetheless, children are able to adapt to task context and improve their irony processing based on previous exposure to irony. The results also show that development of irony comprehension does not happen on a similar timeline for all children; 10-year-olds show large variance between individuals in their ability to comprehend irony. In addition, the results suggest that children who are better in comprehending irony are also faster to resolve the ironic meaning while reading. As the present study is the first study to explore children's processing and comprehension of written irony, more studies are needed to confirm and extend the results.

#### Declaration of Competing Interest

None

#### Data availability

The data and analysis code are made available via Open Science Framework, and the link is provided in the manuscript.

#### Acknowledgements

This study was supported by the Academy of Finland grant #338712 awarded to Henri Olkonieni. Portions of the data were presented at North-West Visual Cognition seminar in Preston, UK, the 21st European Conference on Eye Movements in Leicester, UK, and 9th Experimental Pragmatics Conference in Pavia, Italy. We would like to thank Dr. Seppo Vainio for his help in creating experimental stories and Ms. Stella Puurunen for her help in the data collection.

#### Appendix A. Final models for each measure on reading and comprehending of the ironic and literal stories

**Table A1**  
Final model for text memory question accuracy.

Random effects		<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)		63	4.74	2.18
Participant (Story Type)			6.63	2.57
Item (Intercept)		26	2.13	1.46
Fixed effects		<i>Estimate</i>	<i>95% CI</i>	<i>z</i>
Intercept		5.66	4.65 – 6.67	<b>10.94</b>
Story Type (Literal vs. Ironic)		0.82	–0.57 – 2.20	1.16
Age Group (Adult vs. Child)		–0.28	–1.53 – 0.96	–0.45
Story Type × Age Group		–1.66	–3.35 – 0.04	–1.91

Note. *z* values > |1.96| are bolded.

**Table A2**  
Final model for inference question accuracy.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	2.56	1.60
Item (Intercept)	26	0.40	0.63
Fixed effects	Estimate	95% CI	<i>z</i>
Intercept	2.62	2.13 – 3.10	<b>10.52</b>
Story Type (Literal vs. Ironic)	-2.16	-2.34 to -1.97	<b>-23.09</b>
Age Group (Adult vs. Child)	-2.30	-3.14 to -1.46	<b>-5.38</b>
Trial Order	0.28	0.19 – 0.37	<b>6.10</b>
Story Type × Age Group	-1.21	-1.58 to -0.85	<b>-6.56</b>
Story Type × Trial Order	0.74	0.56 – 0.92	<b>8.12</b>
Age Group × Trial Order	-0.54	-0.72 to -0.37	<b>-6.03</b>
Story Type × Age Group × Trial Order	-0.67	-1.03 to -0.32	<b>-3.70</b>

Note. *z* values > |1.96| are bolded.

**Table A3**  
Final model for first-pass reading time on target phrase.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.07	0.27
Participant (Story Type)		0.001	0.03
Item (Intercept)		0.06	0.25
Item (Age Group)	26	0.01	0.08
Residual		0.10	0.31
Fixed effects	Estimate	95% CI	<i>t</i>
Intercept	7.04	6.92 – 7.16	<b>116.01</b>
Story Type (Literal vs. Ironic)	-0.01	-0.04 – 0.03	-0.28
Age Group (Adult vs. Child)	0.81	0.66 – 0.95	<b>11.14</b>
Trial Order	-0.03	-0.04 to -0.01	<b>-2.87</b>
Comprehension	-0.01	-0.07 – 0.05	-0.28
Age Group × Trial Order	-0.002	-0.04 – 0.03	-0.11
Age Group × Comprehension	-0.03	-0.16 – 0.09	-0.54
Trial Order × Comprehension	-0.02	-0.07 – 0.03	-0.70
Story Type × Age Group	-0.04	-0.12 – 0.03	-1.16
Story Type × Trial Order	-0.03	-0.07 – 0.01	-1.54
Story Type × Comprehension	0.11	-0.02 – 0.24	1.67
Story Type × Age Group × Trial Order	0.01	-0.06 – 0.08	0.29
Story Type × Age Group × Comprehension	-0.03	-0.29 – 0.22	-0.24
Story Type × Trial Order × Comprehension	0.03	-0.07 – 0.13	0.60

Note. *t* values > |1.96| are bolded.

**Table A4**  
Final model for forward-fixation time on target phrase.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.07	0.26
Participant (Story Type)		0.002	0.04
Item (Intercept)		0.05	0.22
Item (Age Group)	26	0.002	0.04
Residual		0.07	0.26
Fixed effects	Estimate	95% CI	<i>t</i>
Intercept	6.84	6.74 – 6.95	<b>126.82</b>
Story Type (Literal vs. Ironic)	0.001	-0.03 – 0.03	0.09
Age Group (Adult vs. Child)	0.69	0.56 – 0.83	<b>10.31</b>
Trial Order	-0.01	-0.03 – 0.002	-1.74
Comprehension	0.01	-0.04 – 0.06	0.37
Age Group × Trial Order	0.01	-0.01 – 0.04	0.92
Age Group × Comprehension	-0.05	-0.15 – 0.05	-0.95
Trial Order × Comprehension	-0.02	-0.06 – 0.02	-0.86
Story Type × Age Group	-0.06	-0.12 – 0.005	-1.81
Story Type × Trial Order	-0.02	-0.05 – 0.01	-1.01
Story Type × Comprehension	-0.01	-0.12 – 0.09	-0.21
Story Type × Age Group × Trial Order	0.02	-0.03 – 0.08	0.77
Story Type × Age Group × Comprehension	0.11	-0.10 – 0.32	0.99
Story Type × Trial Order × Comprehension	0.07	-0.02 – 0.15	1.56

Note. *t* values > |1.96| are bolded.

**Table A5**  
Final model for number of first-pass rereading fixations on target phrase.

Random Effects	n	Variance	SD
Participant (Intercept)	63	0.33	0.57
Participant (Story Type)		0.15	0.39
Item (Intercept)	26	0.22	0.47
Item (Age Group)		0.06	0.24
Count Model Predictors	Estimate	95% CI	z
Intercept	0.13	-0.12 – 0.37	1.01
Story Type (Literal vs. Ironic)	0.04	-0.13 – 0.20	0.45
Age Group (Adult vs. Child)	1.15	0.81 – 1.50	<b>6.56</b>
Trial Order	-0.06	-0.12 – 0.001	-1.94
Comprehension	-0.15	-0.45 – 0.14	-1.03
Age Group × Trial Order	-0.09	-0.21 – 0.02	-1.62
Age Group × Comprehension	0.06	-0.52 – 0.65	0.21
Trial Order × Comprehension	-0.01	-0.14 – 0.12	-0.17
Story Type × Age Group	0.17	-0.15 – 0.49	1.05
Story Type × Trial Order	-0.04	-0.16 – 0.08	-0.67
Story Type × Comprehension	0.60	0.01 – 1.20	<b>1.99</b>
Story Type × Age Group × Trial Order	-0.10	-0.33 – 0.13	-0.85
Story Type × Age Group × Comprehension	-1.21	-2.39 to -0.03	<b>-2.02</b>
Story Type × Trial Order × Comprehension	-0.02	-0.29 – 0.25	-0.15
Zero-Inflated Model Predictors	Estimate	95% CI	z
Intercept	-1.81	-2.13 to -1.48	<b>-10.96</b>
Age Group (Adult vs. Child)	-1.03	-1.67 – -0.39	<b>-3.17</b>

Note. z values > |1.96| are bolded.

**Table A6**  
Final model for look-back fixations on the target phrase.

Random Effects	n	Variance	SD
Participant (Intercept)	63	0.88	0.94
Participant (Story Type)		0.22	0.47
Item (Intercept)	26	0.05	0.22
Item (Age Group)		0.03	0.17
Count Model Predictors	Estimate	95% CI	z
Intercept	0.74	0.44 – 1.03	<b>4.87</b>
Story Type (Literal vs. Ironic)	0.02	-0.24 – 0.29	0.19
Age Group (Adult vs. Child)	0.17	-0.38 – 0.72	0.61
Trial Order	-0.14	-0.21 – -0.07	<b>-3.94</b>
Comprehension	0.39	0.23 – 0.55	<b>4.81</b>
Age Group × Trial Order	-0.01	-0.14 – 0.12	-0.17
Age Group × Comprehension	-0.02	-0.33 – 0.30	-0.12
Trial Order × Comprehension	0.36	0.20 – 0.53	<b>4.33</b>
Story Type × Age Group	-0.10	-0.53 – 0.33	-0.45
Story Type × Trial Order	-0.09	-0.23 – 0.05	-1.27
Story Type × Comprehension	-0.30	-0.62 – 0.02	-1.82
Story Type × Age Group × Trial Order	0.50	0.24 – 0.76	<b>3.80</b>
Story Type × Age Group × Comprehension	0.51	-0.12 – 1.15	1.59
Story Type × Trial Order × Comprehension	0.15	-0.17 – 0.48	0.91
Zero-Inflated Model Predictors	Estimate	95% CI	z
Intercept	0.29	0.14 – 0.44	<b>3.68</b>
Age Group (Adult vs. Child)	2.09	1.80 – 2.39	<b>13.82</b>

Note. z values > |1.96| are bolded.

**Table A7**  
Final model for the number of look-from fixations made from the target phrase.

Random Effects	n	Variance	SD
Participant (Intercept)	63	0.50	0.70
Participant (Story Type)		0.27	0.52
Item (Intercept)	26	0.08	0.28
Item (Age Group)		0.30	0.55
Count Model Predictors	Estimate	95% CI	z
Intercept	0.74	0.41 – 1.06	<b>4.43</b>
Story Type (Literal vs. Ironic)	0.18	-0.14 – 0.51	1.13
Age Group (Adult vs. Child)	-0.16	-0.73 – 0.41	-0.54
Trial Order	-0.12	-0.21 to -0.04	<b>-2.73</b>

(continued on next page)

**Table A7** (continued)

Comprehension	0.26	0.05 – 0.47	<b>2.46</b>
Age Group × Trial Order	-0.16	-0.32 to -0.005	<b>-2.02</b>
Age Group × Comprehension	-0.16	-0.58 – 0.25	-0.78
Trial Order × Comprehension	0.37	0.17 – 0.57	<b>3.64</b>
Story Type × Age Group	0.16	-0.40 – 0.71	0.55
Story Type × Trial Order	-0.27	-0.45 to -0.08	<b>-2.81</b>
Story Type × Comprehension	-0.21	-0.64 – 0.22	-0.95
Story Type × Age Group × Trial Order	-0.20	-0.54 – 0.14	-1.15
Story Type × Age Group × Comprehension	0.10	-0.77 – 0.96	0.23
Story Type × Trial Order × Comprehension	-0.05	-0.45 – 0.35	-0.25
<b>Zero-Inflated Model Predictors</b>			
	<i>Estimate</i>	<i>95% CI</i>	<i>z</i>
Intercept	0.76	0.58 – 0.93	<b>8.31</b>
Age Group (Adult vs. Child)	1.74	1.43 – 2.05	<b>11.08</b>

Note. *z* values > |1.96| are bolded.

**Table A8**

Final model for first-pass reading time on the spillover region.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.07	0.26
Participant (Story Type)		0.01	0.08
Item (Intercept)	26	0.08	0.29
Item (Age Group)		0.01	0.11
Residual		0.16	0.40
Fixed effects	<i>Estimate</i>	<i>95% CI</i>	<i>t</i>
Intercept	6.29	6.16 – 6.42	<b>94.57</b>
Story Type (Literal vs. Ironic)	0.03	-0.02 – 0.08	1.32
Age Group (Adult vs. Child)	0.68	0.54 – 0.83	<b>9.32</b>
Trial Order	0.003	-0.02 – 0.03	0.27
Comprehension	-0.002	-0.08 – 0.08	-0.05
Age Group × Trial Order	0.02	-0.02 – 0.07	1.04
Age Group × Comprehension	-0.02	-0.18 – 0.14	-0.26
Trial Order × Comprehension	-0.01	-0.07 – 0.06	-0.26
Story Type × Age Group	-0.05	-0.15 – 0.06	-0.89
Story Type × Trial Order	-0.04	-0.09 – 0.01	-1.62
Story Type × Comprehension	0.08	-0.08 – 0.25	1.02
Story Type × Age Group × Trial Order	-0.03	-0.12 – 0.05	-0.75
Story Type × Age Group × Comprehension	-0.21	-0.53 – 0.11	-1.28
Story Type × Trial Order × Comprehension	0.03	-0.10 – 0.16	0.49

Note. *t* values > |1.96| are bolded.

**Table A9**

Final model for probability to look-back to the critical context.

Random Effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	1.26	1.12
Item (Intercept)	26	0.15	0.39
Item (Age Group)		0.23	0.48
Count Model Predictors	<i>Estimate</i>	<i>95% CI</i>	<i>z</i>
Intercept	-0.49	-0.97 – -0.01	<b>-1.98</b>
Story Type (Literal vs. Ironic)	-0.34	-0.68 – -0.003	<b>-1.98</b>
Age Group (Adult vs. Child)	-2.58	-3.37 – -1.79	<b>-6.39</b>
Trial Order	-0.19	-0.35 – -0.03	<b>-2.30</b>
Comprehension	0.68	-0.49 – 1.85	1.13
Age Group × Trial Order	-0.14	-0.44 – 0.16	-0.91
Age Group × Comprehension	-1.39	-3.75 – 0.97	-1.15
Trial Order × Comprehension	0.09	-0.45 – 0.64	0.34
Story Type × Age Group	-0.01	-0.66 – 0.64	-0.02
Story Type × Trial Order	-0.68	-1.03 to -0.32	<b>-3.76</b>
Story Type × Comprehension	-0.90	-3.21 – 1.40	-0.77
Story Type × Age Group × Trial Order	0.13	-0.50 – 0.75	0.39
Story Type × Age Group × Comprehension	1.11	-3.54 – 5.76	0.47
Story Type × Trial Order × Comprehension	0.85	-0.24 – 1.93	1.53
<b>Zero-Inflated Model Predictors</b>			
	<i>Estimate</i>	<i>95% CI</i>	<i>z</i>
Intercept	-2.60	-4.36 to -0.84	<b>-2.89</b>

Note. *z* values > |1.96| are bolded.

**Appendix B. Final models on relationships between reading of ironic texts and irony comprehension**

**Table B1**

Final model for first-pass reading time on ironic target phrase.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.06	0.25
Item (Intercept)		0.07	0.27
Item (Age Group)	26	0.01	0.11
Residual		0.08	0.29
Fixed effects	<i>Estimate</i>	95% <i>CI</i>	<i>t</i>
Intercept	7.04	6.92 – 7.16	<b>111.89</b>
Comprehension Accuracy	0.004	–0.07 – 0.07	0.12
Age Group (Adult vs. Child)	0.80	0.66 – 0.95	<b>10.93</b>
Age Group × Comprehension Accuracy	–0.24	–0.38 to –0.10	<b>–3.41</b>

Note. *t* values > |1.96| are bolded.

**Table B2**

Final model for forward-fixation time on ironic target phrase.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.06	0.24
Item (Intercept)		0.05	0.23
Item (Age Group)	26	0.000004	0.002
Residual		0.06	0.24
Fixed effects	<i>Estimate</i>	95% <i>CI</i>	<i>t</i>
Intercept	6.85	6.74 – 6.96	<b>121.76</b>
Comprehension Accuracy	–0.01	–0.08 – 0.05	–0.42
Age Group (Adult vs. Child)	0.68	0.55 – 0.81	<b>10.27</b>
Age Group × Comprehension Accuracy	–0.22	–0.35 to –0.08	<b>–3.20</b>

Note. *t* values > |1.96| are bolded.

**Table B3**

Final model for number of first-pass rereading fixations on ironic target phrase.

Random Effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.30	0.55
Item (Intercept)	26	0.26	0.51
Item (Age Group)		0.35	0.59
Count Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	0.02	–0.26 – 0.30	0.13
Comprehension Accuracy	0.13	–0.06 – 0.32	1.35
Age Group (Adult vs. Child)	1.37	0.95 – 1.80	<b>6.31</b>
Age Group × Comprehension Accuracy	–0.42	–0.79 to –0.04	<b>–2.17</b>
Zero-Inflated Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	–1.98	–2.56 to –1.40	<b>–6.70</b>

Note. *z* values > |1.96| are bolded.

**Table B4**

Final model for number of look-back fixations on ironic target phrase.

Random Effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.60	0.77
Item (Intercept)	26	0.21	0.46
Item (Age Group)		0.25	0.50
Count Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	0.79	0.42 – 1.15	<b>4.19</b>
Comprehension Accuracy	0.10	–0.19 – 0.40	0.69
Age Group (Adult vs. Child)	0.08	–0.58 – 0.74	0.24
Age Group × Comprehension Accuracy	–0.03	–0.62 – 0.57	–0.10
Zero-Inflated Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	0.31	0.06 – 0.55	<b>2.47</b>
Age Group (Adult vs. Child)	1.82	1.35 – 2.30	<b>7.48</b>

Note. *z* values > |1.96| are bolded.

**Table B5**  
Final model for number of look-from fixations made from the ironic target phrase.

Random Effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.17	0.41
Item (Intercept)	26	0.11	0.33
Item (Age Group)		1.36	1.16
Count Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	0.87	0.54 – 1.19	<b>5.21</b>
Comprehension Accuracy	0.23	-0.02 – 0.49	1.79
Age Group (Adult vs. Child)	-0.09	-0.82 – 0.65	-0.23
Age Group × Comprehension Accuracy	0.02	-0.49 – 0.53	0.08
Zero-Inflated Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	0.83	0.57 – 1.09	<b>6.31</b>
Age Group (Adult vs. Child)	1.63	1.12 – 2.14	<b>6.26</b>

Note. *z* values > |1.96| are bolded.

**Table B6**  
Final model for first-pass reading time on the spillover region following ironic target phrase.

Random effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	0.06	0.25
Item (Intercept)		0.08	0.29
Item (Age Group)	26	0.02	0.15
Residual		0.15	0.38
Fixed effects	<i>Estimate</i>	95% <i>CI</i>	<i>t</i>
Intercept	6.34	6.21 – 6.48	<b>93.62</b>
Comprehension Accuracy	-0.07	-0.14 – 0.01	-1.79
Age Group (Adult vs. Child)	0.73	0.57 – 0.89	<b>9.17</b>
Age Group × Comprehension Accuracy	-0.29	-0.44 to -0.14	<b>-3.85</b>

Note. *t* values > |1.96| are bolded.

**Table B7**  
Final model for probability to look-back to critical context of ironic story.

Random Effects	<i>n</i>	Variance	<i>SD</i>
Participant (Intercept)	63	1.00	1.00
Item (Intercept)	26	0.40	0.63
Count Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	-0.81	-1.26 to -0.35	<b>-3.49</b>
Comprehension Accuracy	0.34	-0.03 – 0.71	1.79
Age Group (Adult vs. Child)	-2.23	-3.00 to -1.47	<b>-5.73</b>
Age Group × Comprehension Accuracy	-0.07	-0.81 – 0.67	-0.18
Zero-Inflated Model Predictors	<i>Estimate</i>	95% <i>CI</i>	<i>z</i>
Intercept	-19.24	-11493.98 – 11455.50	-0.003

Note. *z* values > |1.96| are bolded.

## References

- Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, 35, 487–508. [https://doi.org/10.1016/0022-0965\(83\)90023-1](https://doi.org/10.1016/0022-0965(83)90023-1)
- Aguert, M., Laval, V., Gauducheau, N., Atifi, H., & Marcocchia, M. (2016). Producing irony in adolescence: A comparison between face-to-face and computer-mediated communication. *Psychology of Language and Communication*, 20, 199–218. <https://doi.org/10.1515/plc-2016-0013>
- Angeleri, R., & Airenti, G. (2014). Development of joke and irony understanding: A study with 3- to 6-year-old children. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 68, 133–146. <https://doi.org/10.1037/cep0000011>
- Au-Yeung, S. K., Kaakinen, J. K., Liversedge, S. P., & Benson, V. (2015). Processing of written irony in autism spectrum disorder: An eye-movement study. *Autism Research*, 8, 749–760. <https://doi.org/10.1002/aur.1490>
- 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. <https://doi.org/10.1016/j.jml.2007.12.005>
- Banasik-Jemielniak, N., & Bokus, B. (2019). Children's comprehension of irony: Studies on polish-speaking preschoolers. *Journal of Psycholinguistic Research*, 48, 1217–1240. <https://doi.org/10.1007/s10936-019-09654-x>
- Banasik-Jemielniak, N., Bosacki, S., Mitrowska, A., Wyrębek Walters, D., Wisiecka, K., Copeland, N. E., ... Siemieniuk, A. (2020). "Wonderful! We've just missed the bus." – Parental use of irony and children's irony comprehension. *PLoS One*, 15, Article e0228538. <https://doi.org/10.1371/journal.pone.0228538>
- 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, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Blythe, H. I., & Joseph, H. S. S. L. (2011). Children's eye movements during reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 643–662). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199539789.013.0036>
- Bosco, F. M., & Gabbatore, I. (2017). Sincere, deceitful, and ironic communicative acts and the role of the theory of mind in childhood. *Frontiers in Psychology*, 8, 21. <https://doi.org/10.3389/fpsyg.2017.00021>
- Bosco, F. M., Tirasa, M., & Gabbatore, I. (2018). Why pragmatics and theory of mind do not (completely) overlap. *Frontiers in Psychology*, 9, 1453. <https://doi.org/10.3389/fpsyg.2018.01453>
- Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23, 389–411. <https://doi.org/10.1037/met0000159>
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages

- for zero-inflated generalized linear mixed modeling. *The R Journal*, 9, 378–400. <https://doi.org/10.32614/RJ-2017-066>
- Bryant, B. (1982). An index of empathy for children and adolescents. *Child Development*, 53, 413–425. <https://doi.org/10.2307/1128984>
- Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, 48, 257–277. <https://doi.org/10.1177/00238309050480030101>
- 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. <https://doi.org/10.1016/j.ridd.2014.08.002>
- Capelli, C. A., Nakagawa, N., & Madden, C. M. (1990). How children understand sarcasm: The role of context and intonation. *Child Development*, 61, 1824–1841. <https://doi.org/10.2307/1130840>
- 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. <https://doi.org/10.1080/15248370802247939>
- Connor, C. M., Radach, R., Vorstius, C., Day, S. L., McLean, L., & Morrison, F. J. (2015). Individual differences in fifth graders' literacy and academic language predict comprehension monitoring development: An eye-movement study. *Scientific Studies of Reading*, 19, 114–134. <https://doi.org/10.1080/10888438.2014.943905>
- Del Sette, P., Bambini, V., Bischetti, L., & Lecce, S. (2020). Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development*, 56, Article 100958. <https://doi.org/10.1016/j.cogdev.2020.100958>
- Demorest, A., Christine, M., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development*, 55, 1527–1534. <https://doi.org/10.2307/1130022>
- Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., McGovern, A., Qualter, A., & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child Development*, 67, 3071–3085. <https://doi.org/10.1111/j.1467-8624.1996.tb01903.x>
- Ehri, L. C. (1995). Phases of development in learning to read by sight. *Journal of Research in Reading*, 18, 116–125. <https://doi.org/10.1111/j.1467-9817.1995.tb00077.x>
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9, 167–188. <https://doi.org/10.1207/s1532799xssr0902.4>
- Fabry, R. E. (2021). Getting it: A predictive processing approach to irony comprehension. *Synthese*, 198, 6455–6489. <https://doi.org/10.1007/s11229-019-02470-9>
- Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, 56, 485–495. <https://doi.org/10.1080/0163853X.2019.1591885>
- Filik, R., & Moxey, L. M. (2010). The on-line processing of written irony. *Cognition*, 116, 421–436. <https://doi.org/10.1016/j.cognition.2010.06.005>
- Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, 79, 126–138. <https://doi.org/10.1111/j.1467-8624.2007.01115.x>
- Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.
- Giora, R. (2003). *On our mind: Saliency, context, and figurative language*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195136166.001.0001>
- Glenwright, M., & Pexman, P. M. (2010). Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language*, 37, 429–451. <https://doi.org/10.1017/S0305000909009520>
- Glenwright, M., Tapley, B., Rano, J. K. S., & Pexman, P. M. (2017). Developing appreciation for sarcasm and sarcastic gossip: It depends on perspective. *Journal of Speech, Language, and Hearing Research*, 9, 3295–3309. <https://doi.org/10.1044/2017.jshlr-17-0058>
- Godbee, K., & Porter, M. (2013). Comprehension of sarcasm, metaphor and simile in Williams syndrome. *International Journal of Language & Communication Disorders*, 48, 651–665. <https://doi.org/10.1111/1460-6984.12037>
- Grice, H. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). Academic Press.
- Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, 102, 167–181. <https://doi.org/10.1016/j.jecp.2008.04.002>
- Häikiö, T., Heikkilä, T. T., & Kaakinen, J. K. (2018). The effect of syllable-level hyphenation on reading comprehension: Evidence from eye movements. *Journal of Educational Psychology*, 110, 1149–1159. <https://doi.org/10.1037/edu0000261>
- Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development*, 1, 227–248. <https://doi.org/10.1207/s15327647jcd010204>
- Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, 48, 101–109. [https://doi.org/10.1016/0010-0277\(93\)90026-r](https://doi.org/10.1016/0010-0277(93)90026-r)
- Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes*, 36, 147–165. <https://doi.org/10.1207/s15326950dp3603.1>
- Hayashi, H., & Ban, Y. (2021). Children's understanding of unintended irony and unsuccessful irony. *European Journal of Developmental Psychology*, 18, 230–256. <https://doi.org/10.1080/17405629.2020.1783528>
- Häyrynen, T., Serenius-Sirve, S., & Korkman, M. (2013). *Lukilasse 2: Lukemisen, kirjoittamisen ja laskemisen seulontatesti 1.–6. vuosiluokille. [Lukilasse 2: Screening test for reading, writing and calculus for 1st to 6th grades.]*. Hogrefe Psychologien Kustannus Oy.
- Hyönä, J., Lorch, R. F., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: Elsevier Science. <https://doi.org/10.1016/B978-0-44451020-4/50018-9>
- Kaakinen, J. K., Lehtola, A., & Paattilampi, S. (2015). The influence of a reading task on children's eye movements during reading. *Journal of Cognitive Psychology*, 27, 640–656. <https://doi.org/10.1080/20445911.2015.1005623>
- Kaakinen, J. K., Olkonieni, H., Kinnari, T., & Hyönä, J. (2014). Processing of written irony: An eye movement study. *Discourse Processes*, 51, 287–311. <https://doi.org/10.1080/0163853X.2013.870024>
- Kim, J., & Lantolf, J. P. (2018). Developing conceptual understanding of sarcasm in L2 English through explicit instruction. *Language Teaching Research*, 22, 208–229. <https://doi.org/10.1177/1362168816675521>
- Köder, F., & Falkum, I. L. (2021). Irony and perspective-taking in children: The roles of norm violations and tone of voice. *Frontiers in Psychology*, 12, 624604. <https://doi.org/10.3389/fpsyg.2021.624604>
- Loukusa, S., & Leinonen, E. (2008). Development of comprehension of ironic utterances in 3- to 9-year-old Finnish-speaking children. *Psychology of Language and Communication*, 12, 55–69. <https://doi.org/10.2478/v10057-008-0003-0>
- Massaro, D., Valle, A., & Marchetti, A. (2013). Irony and second-order false-belief in children: What changes when mothers rather than siblings speak? *European Journal of Developmental Psychology*, 10, 301–317. <https://doi.org/10.1080/17405629.2012.672272>
- Massaro, D., Valle, A., & Marchetti, A. (2014). Do social norms, false belief understanding, and metacognitive vocabulary influence irony comprehension? A study of five- and seven-year-old children. *European Journal of Developmental Psychology*, 11, 292–304. <https://doi.org/10.1080/17405629.2013.821407>
- Mazzarella, D., & Pouscoulous, N. (2020). Pragmatics and epistemic vigilance: A developmental perspective. *Mind & Language*, 36, 355–376. <https://doi.org/10.1111/mila.12287>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 international results in reading. Retrieved from: <http://timssandpirls.bc.edu/isc/publications.htm>
- Nicholson, A., Whalen, J. M., & Pexman, P. M. (2013). Children's processing of emotion in ironic language. *Frontiers in Psychology*, 4, 691. <https://doi.org/10.3389/fpsyg.2013.00691>
- Nilsen, E. S., Glenwright, M., & Huyder, V. (2011). Children and adults understand that verbal irony interpretation depends on listener knowledge. *Journal of Cognition and Development*, 12, 374–409. <https://doi.org/10.1080/15248372.2010.544693>
- Noveck, I. A., Bianco, M., & Castry, A. (2001). The costs and benefits of metaphor. *Metaphor and Symbol*, 16, 109–121. <https://doi.org/10.1080/10926488.2001.9678889>
- Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and Writing*, 18, 657–686. <https://doi.org/10.1007/s11145-005-3355-z>
- Olkonieni, H., Johander, E., & Kaakinen, J. K. (2019). The role of look-backs in the processing of written sarcasm. *Memory & Cognition*, 47, 87–105. <https://doi.org/10.3758/s13421-018-0852-2>
- Olkonieni, H., & Kaakinen, J. K. (2021). Processing of irony in text: A systematic review of eye-tracking studies. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 75, 99–106. <https://doi.org/10.1037/cep0000216>
- Olkonieni, 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, 433–450. <https://doi.org/10.1037/xlm0000176>
- Olkonieni, H., Strömberg, V., & Kaakinen, J. K. (2019). The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, 72, 1212–1223. <https://doi.org/10.1177/1747021818807864>
- Panzeri, F., Giustolisi, B., & Zampini, L. (2020). The comprehension of ironic criticisms and ironic compliments in individuals with down syndrome: Adding another piece to the puzzle. *Journal of Pragmatics*, 156, 223–234. <https://doi.org/10.1016/j.pragma.2019.08.009>
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383. <https://doi.org/10.1080/10888430701530730>
- Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory of mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development*, 83, 469–485. <https://doi.org/10.1111/j.1467-8624.2011.01728.x>
- Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, 17, 286–290. <https://doi.org/10.1111/j.1467-8721.2008.00591.x>
- Pexman, P. M., Glenwright, M., Krol, A., & James, T. (2005). An acquired taste: Children's perceptions of humor and teasing in verbal irony. *Discourse Processes*, 40, 259–288. <https://doi.org/10.1207/s15326950dp4003.5>
- Pexman, P. M., Zdrzilova, L., McConnachie, D., Deater-Deckard, K., & Petrill, S. A. (2009). "That was smooth, mom!": Children's production of verbal and gestural irony. *Metaphor and Symbol*, 24, 237–248. <https://doi.org/10.1080/10926480903310286>
- R Core Team. (2021). *R: A language and environment for statistical computing*. Retrieved from: <http://www.R-project.org>. Retrieved from:
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201. <https://doi.org/10.3758/bf03197692>
- Recchia, H. E., Howe, N., Ross, H. S., & Alexander, S. (2010). Children's understanding and production of verbal irony in family conversations. *British Journal of*

- Developmental Psychology*, 28, 255–274. <https://doi.org/10.1348/026151008X401903>
- Rivière, E., Klein, M., & Champagne-Lavau, M. (2018). Using context and prosody in understanding irony: Variability amongst individuals. *Journal of Pragmatics*, 138, 165–172. <https://doi.org/10.1016/j.pragma.2018.10.006>
- Ronimus, M., Tolvanen, A., & Hautala, J. (2022). The roles of motivation and engagement in computer-based assessment of children's reading comprehension. *Learning and Individual Differences*, 98, Article 102197. <https://doi.org/10.1016/j.lindif.2022.102197>
- Rowling, J. K. (2005). *Harry Potter and the half-blood prince*. Bloomsbury Publishing plc.
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, 110, Article 104038. <https://doi.org/10.1016/j.jml.2019.104038>
- van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. D. M. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*, 25, 1665–1690. <https://doi.org/10.1007/s11145-011-9337-4>
- Schroeder, S., Häikiö, T., Pagán, A., Dickins, J. H., Hyönä, J., & Liversedge, S. P. (2022). Eye movements of children and adults reading in three different orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48, 1518–1541. <https://doi.org/10.1037/xlm0001099>
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174. <https://doi.org/10.1348/000712603321661859>
- Spotorno, N., & Noveck, I. (2019). Ironic utterances. In C. Cummings, & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics* (pp. 283–297). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198791768.013.16>
- Spotorno, N., & Noveck, I. A. (2014). When is irony effortful? *Journal of Experimental Psychology: General*, 143, 1649–1665. <https://doi.org/10.1037/a0036630>
- Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology*, 13, 191–204. <https://doi.org/10.1111/j.2044-835X.1995.tb00673.x>
- Vorstius, C., Radach, R., Mayer, M. B., & Lonigan, C. J. (2013). Monitoring local comprehension monitoring in sentence reading. *School Psychology Review*, 42, 191–206. <https://doi.org/10.1080/02796015.2013.12087484>
- Wechsler, D. (2010). *WISC IV - Wechsler intelligence scale for children* (4th ed.). NCS Pearson Ltd.
- Whalen, J. M., Doyle, A., & Pexman, P. M. (2020). Sarcasm between siblings: Children's use of relationship information in processing ironic remarks. *Journal of Pragmatics*, 156, 149–159. <https://doi.org/10.1016/j.pragma.2019.05.005>
- Whalen, J. M., & Pexman, P. M. (2010). How do children respond to verbal irony in face-to-face communication? The development of mode adoption across middle childhood. *Discourse Processes*, 47, 363–387. <https://doi.org/10.1080/01638530903347635>
- Wilson, D. (2013). Irony comprehension: A developmental perspective. *Journal of Pragmatics*, 59, 40–56. <https://doi.org/10.1016/j.pragma.2012.09.016>
- Winner, E., & Leekam, S. (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, 9, 257–270. <https://doi.org/10.1111/j.2044-835X.1991.tb00875.x>
- Zajackowska, M., & Abbot-Smith, K. (2020). “Sure I'll help—I've just been sitting around doing nothing at school all day”: Cognitive flexibility and child irony interpretation. *Journal of Experimental Child Psychology*, 199, Article 104942. <https://doi.org/10.1016/j.jecp.2020.104942>
- Zargar, E., Adams, A. M., & Connor, C. M. (2020). The relations between children's comprehension monitoring and their reading comprehension and vocabulary knowledge: An eye-movement study. *Reading and Writing*, 33, 511–545. <https://doi.org/10.1007/s11145-019-09966-3>