CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

The Developmental Puzzle of Irony Understanding: Is Epistemic Vigilance the Missing Piece?

Ana Milosavljevic[1] (iD), Thomas Castelain[1] (iD), Nausicaa Pouscoulous[2] (iD) and Diana Mazzarella[1] (iD)

[1]Cognitive Science Centre, University of Neuchâtel, Switzerland and [2]Linguistics Department, University College London, London, UK
Corresponding author: Diana Mazzarella; Email: diana.mazzarella@unine.ch

Abstract
The prolonged developmental window of irony understanding opens up the question of which socio-cognitive repertoire underlies this pragmatic capacity. In the present study, we investigated the relationship between epistemic vigilance and irony understanding in 5/6- and 6/7-year-old children using a picture selection task. We assessed children's vigilance towards unreliable informants and manipulated the reliability of the irony target. Our findings confirm that irony comprehension is a late-emerging skill and highlight the need to differentiate its full-fledged understanding from mere sensitivity to contextual mismatches. While irony understanding was not affected by our reliability manipulation, our findings revealed that more vigilant children were better at irony understanding than less vigilant ones. This provides the first empirical evidence that epistemic vigilance is a good predictor of irony performance and lays the ground for future research on the intricate relationship between these two capacities.

Résumé
La fenêtre développementale prolongée de la compréhension de l'ironie soulève la question du répertoire socio-cognitif sous-jacent à cette capacité pragmatique. Dans cette étude, nous examinons la relation entre vigilance épistémique et compréhension de l'ironie chez des enfants de 5/6 et 6/7 ans à l'aide d'une tâche de sélection d'images. Nous évaluons la vigilance des enfants envers des informateurs peu fiables et manipulons la fiabilité de la cible ironique. Nos résultats confirment que la compréhension de l'ironie est une compétence qui émerge tardivement et soulignent la nécessité de distinguer une réelle compréhension de l'ironie d'une simple sensibilité aux discordances contextuelles. Nos résultats n'indiquent pas que la compréhension des énoncés ironiques par les enfants ait été affectée par notre manipulation de la fiabilité de l'informateur, ils suggèrent néanmoins que les enfants les plus

---

vigilants comprennent mieux l'ironie que ceux qui sont moins vigilants. Il s'agit là d'une première preuve empirique indiquant que la vigilance épistémique est un bon prédicteur de la capacité à saisir l'ironie chez l'enfant, qui ouvre la voie à de futures recherches sur la relation complexe entre ces deux capacités.

Irony is a form of indirect language where the speaker states something untrue or irrelevant with the intention of expressing a critical or dissociative attitude. Understanding irony thus requires the hearer to go beyond the surface meaning of the utterance, consider available linguistic and contextual cues, and engage in an inferential process to grasp the intended interpretation (Wilson & Sperber, 2012). Such deliberate indirectness, by which the speaker intentionally states some falsehoods to implicitly communicate a critical attitude, may at times be confusing for adults and often poses a challenge for children (Matsui, 2019).

To illustrate the complexity of irony understanding, let us focus on the following example. Suppose that Louis has been playing in his room the whole morning, and the room is now very messy with toys all over the floor. His mother asks Louis to tidy up the room before going outside to play. After some time, when the mother realises that Louis has not done the job properly, she utters, "Well done, your room is really tidy!" The mismatch between the literal meaning of the sentence uttered and the context may offer a cue to understand irony (Colston & O'Brien, 2000; Rivière, Klein, & Champagne-Lavau, 2018), provided that children can infer that the speaker's true belief differs from what is conveyed by her actual words (i.e., that the mother does not genuinely believe that the room is tidy). Crucially, though, understanding irony requires additional insights into the speaker's intention: it requires grasping that the speaker wants the hearer to identify the falsity (or irrelevance) of the proposition literally expressed, thus recognising that the speaker is benevolent and does not intend to deceive the addressee. This allows distinguishing irony from lies: both involve communicating intentional falsehoods, but only irony is driven by honest intentions (Demorest, Meyer, Phelps, Gardner, & Winner, 1984; Winner & Leekam, 1991). Finally, understanding irony involves recognising that the speaker says something false or irrelevant to communicate their critical epistemic stance to the hearer (Wilson & Sperber, 2012). Most theoretical accounts of verbal irony highlight the implicit expression of this dissociative attitude as a pivotal and distinctive characteristic of irony (Clark & Gerrig, 1984; Kumon-Nakamura, Glucksberg, & Brown, 1995; Wilson & Sperber, 2012). The ironical attitude is fundamentally dissociative: in our example, the mother is expressing a dissociative attitude towards the expectation that the room should be tidy, based on the recognition that this expectation is blatantly unsatisfied.

The complexity lying behind irony interpretation opens up the questions of when children acquire a full-fledged understanding of this unique language use and which cognitive capacities may support it. There is still no consensus in the literature on the mechanisms that underpin its development, with successful performance in the irony comprehension tasks often being associated with different socio-cognitive capacities that include language skills, Theory of Mind, executive functions, and emotion recognition (for review, see Pexman, 2023). While many studies have pointed to the crucial role of Theory of Mind (see, e.g., Filippova & Astington, 2008), current pragmatic accounts of irony understanding have highlighted the link between irony understanding and the

cognitive mechanisms that target the risk of misinformation, known under the umbrella term of "epistemic vigilance" (Matsui, 2019; Mazzarella & Pouscoulous, 2021, 2023; Scianna, 2023; Wilson, 2009). Epistemic vigilance involves the assessment of the reliability of the source of information and the evaluation of the plausibility of its content to calibrate epistemic trust (Sperber et al., 2010). To address the puzzle of the development of irony understanding, it is essential to empirically evaluate these pragmatic accounts and thus examine whether epistemic vigilance buttresses irony understanding during childhood.

Therefore, the goal of this study is to examine the hypothesis that epistemic vigilance may play a role in understanding irony. To set the background, we start by discussing the state of the art of research on irony development and highlight some methodological aspects related to measuring irony understanding. We then proceed by outlining the claim that epistemic vigilance should be considered as part and parcel of the socio-cognitive repertoire that makes irony understanding possible. Finally, we present the design and the results of the study, discuss their implications, and put forth some suggestions for future research in this area.

1. Irony development

Irony appears to display a relatively late development compared to other non-literal uses of language, such as metaphor, metonymy, or hyperbole, that children seem to be able to understand already at the age of 3 (see, e.g., Deamer, 2013; Di Paola, Domaneschi, & Pouscoulous, 2020; Falkum, Recasens, & Clark, 2017; Pouscoulous & Tomasello, 2020; for a discussion, see Pouscoulous, 2023). Indeed, most studies on the development of irony indicate that understanding irony does not emerge before the age of 6 years (for a review, see Fuchs, 2023). However, it may even take a few additional years for children to develop a full appreciation of the speaker's attitude and the ironical intent, as well as the ability to differentiate irony from a deceptive act (see, e.g., Dews et al., 1996; Glenwright & Pexman, 2010; Hancock, Dunham, & Purdy, 2000). Unlike adults, who may perceive utterances as ironical when distinct prosodic cues are present (Deliens, Antoniou, Clin, Ostashchenko, & Kissine, 2018), children's reliance on prosodic cues appears to be more fragile, with some studies showing a facilitatory effect of prosody (e.g., Capelli, Nakagawa, & Madden, 1990; Zajączkowska, 2016) and others failing to find any evidence for it (e.g., Filippova & Astington, 2008; Köder & Falkum, 2021). Furthermore, children's difficulties are amplified when irony is used with more complex functions, such as humour and teasing (e.g., Dews et al., 1996), or less familiar ones, such as ironical compliments, in which the speaker combines praise with the expression of the typical dissociative attitude of the irony, thus distancing themself from a negative thought or expectation attributed to someone else (e.g., Filippova & Astington, 2010). These difficulties persist even during middle adolescence (see e.g., Demorest et al., 1984; Massaro, Valle, & Marchetti, 2013). In contrast with this picture, though, a handful of more recent studies employing implicit measures such as eye gaze have made the case for an early sensitivity to irony in preschoolers (see, e.g., Climie & Pexman, 2008; Köder & Falkum, 2021). These mixed findings point to a relatively extended developmental window and make it challenging to establish a clear acquisition pattern. These findings also open up some methodological questions concerning the extent to which the measures employed are targeting irony understanding rather than a mere sensitivity to contextual mismatches.

The development of irony understanding is most often measured by using short narratives presented in the form of vignettes or puppet scenarios that typically involve

interactions between two to three characters (see, e.g., Climie & Pexman, 2008; Filippova & Astington, 2008; Hancock et al., 2000; Köder & Falkum, 2021). Within the story, one character typically behaves in a way that violates a contextual or norm-based expectation and leads to a negative outcome (e.g., Louis fails to tidy up the room, which remains messy), while the other character (the speaker) comments on such a negative outcome with an ironical target utterance that is always opposite in valence (e.g., "Well done, your room is really tidy!"). Finally, children's understanding of the target utterance is measured using different types of verbal or non-verbal responses, such as selecting one out of two or more emoticons that represent the speaker's intent or mental states (happy versus angry face emoticon). Faced with irony, children would thus need to select a negative response (the angry face emoticon) to describe the speaker's feelings despite the positive valence of the phrasing of the target utterance.

While most studies using offline measures suggest that preschoolers tend to fail this type of task and select the positive answer (e.g., the happy face emoticon) that corresponds to the literal interpretation of the ironical statement, online measures based on eye gaze indicate that preschoolers gaze at the correct answer (e.g., the angry face emoticon) more in the presence of irony than of literal praise (Climie & Pexman, 2008; Köder & Falkum, 2021). This has been interpreted as suggesting that, although younger participants demonstrate a strong lexical bias and eventually give preference to the literal meaning of the utterance (Aguert, Laval, Le Bigot, & Bernicot, 2010; Matsui, 2019), offline measures may mask their developing pragmatic skills due to other competing factors. For instance, a growing sensitivity to sense conventions during the preschool and early school years might interfere with children's pragmatic reasoning, especially when faced with less conventional forms of language like irony (Falkum, 2022).

Crucially, to advance our understanding of the emergence of irony understanding, it is important to devise experimental tasks that can clearly distinguish it from mere sensitivity to the presence of a mismatch between some salient (norm-based) expectation and the actual state of affairs (see Schmidt, Rakoczy, Mietzsch, & Tomasello, 2016 for a discussion of children's tendency to follow and enforce social norms). Indeed, while perceiving such a mismatch may lead children to make the right inferences concerning the speaker's feelings (e.g., to infer that Mum is angry when the room is not tidy) and show an appropriate behavioural response (gazing at/selecting the angry emoticon face), irony understanding requires inferring that the speaker has stated some falsehood to express a dissociative attitude towards it, thus integrating the story context and the speaker utterance. In our study, we thus built on the experimental paradigm of Köder and Falkum (2021) and introduced a control condition in which a negative outcome (comparable to the one giving rise to the ironical remark) should lead to a different behavioural answer (the happy face emoticon). This allowed us to isolate the subgroup of children who passed this control, and whose performance in irony understanding could thus not be reduced to mere sensitivity to a contextual mismatch. Before turning to our study, in the next section, we discuss the question of the socio-cognitive repertoire involved in irony understanding.

2. Socio-cognitive correlates of irony understanding

According to most pragmatic accounts of irony, an ironical speaker intentionally says something false or irrelevant to communicate a dissociative attitude towards it (Wilson & Sperber, 2012). Irony understanding would thus require recognising that the speaker is expressing a *thought about a thought*, an ability that is arguably linked to advanced

Theory of Mind (Winner & Leekam, 1991). Brain imaging studies suggest a close connection between irony and Theory of Mind, as they appear to activate common brain areas such as the ventromedial prefrontal cortex (Spotorno, Koun, Prado, Van Der Henst, & Noveck, 2012). This connection is supported by numerous studies with both typical and atypical populations, indicating a positive association between irony performance and higher-order Theory of Mind skills (e.g., Filippova & Astington, 2008; Langdon, Davies, & Coltheart, 2002; Massaro et al., 2013; but see also Angeleri & Airenti, 2014; Panzeri, Giustolisi, & Zampini, 2020).

While Theory of Mind may play an important role in irony understanding, recent pragmatic accounts of irony understanding have highlighted the relevance of other socio-cognitive skills to grasp the dissociative attitude expressed by an ironical speaker (Mazzarella & Pouscoulous, 2023). The ironical speaker expresses a critical attitude towards a thought that they judge ludicrously false or irrelevant, typically attributed to an unreliable source (other than the speaker themselves at the present time). Whether the source is a specific individual, or a group or people in general, this critical, mocking, or scornful attitude is meant to highlight the inappropriateness of a thought (a belief, a desire, a norm-based expectation) and, as a result, the unreliability of its source (Wilson & Sperber, 2012). For this reason, several theorists have suggested that epistemic vigilance may be crucial for navigating the complexities of evaluating both the expressed content and recognising the attitude conveyed by the ironical speaker (Matsui, 2019; Mazzarella & Pouscoulous, 2021; Scianna, 2023; Wilson, 2009). Epistemic vigilance comprises a set of cognitive mechanisms that prevent the risk of being misled by others. These involve the capacity to judge the credibility and acceptability of the incoming information – "epistemic vigilance towards the content" – as well as the capacity to assess the reliability of the one who provides such information – "epistemic vigilance towards the source" (for an overview, see Sperber et al., 2010). When being ironical, the speaker produces a false or irrelevant proposition attributed to another source. Importantly, the speaker neither believes this proposition to be true nor wishes to persuade the hearer into believing it. Evidence from developmental studies on irony comprehension indicates that young children often mistake irony as an error or as a lie (see e.g., Demorest, Silberstein, Gardner, & Winner, 1983; Filippova & Astington, 2008; Winner & Leekam, 1991). For this reason, Mazzarella and Pouscoulous (2021, 2023) suggested that children need to be able to actively assess the competence and honesty of the ironical speaker to rule out, respectively, the possibility of mistake and deception, and understand the motivation to deliberately express a false or irrelevant proposition, even without any intention to induce a false belief in the addressee. Furthermore, to fully grasp irony, children must recognise the speaker's dissociative attitude. By recognising that the ironical speaker judges a thought as untrue or irrelevant and its source as unreliable, children can decipher that the point of the irony is to express the speaker's reaction to it. Mazzarella and Pouscoulous (2021) have therefore suggested that irony understanding may be facilitated when the information about the unreliability of the source (the target of the irony) is salient in context.

Mazzarella and Pouscoulous (2023) have argued that actively assessing the competence and honesty of the ironical speaker and attributing to them the capacity to be critical towards unreliable sources of information require the exercise of increasingly sophisticated forms of epistemic vigilance. In this context, "actively" denotes the capacity to assess the competence and benevolence of the speaker without taking it for granted, thus inferring that the speaker is willing and able to provide accurate information based on an analysis of available contextual cues and the speaker's mental states. Importantly,

children's epistemic vigilance undergoes significant developmental changes in the pre-school years. Specifically, between the ages of 2 and 4 years, children demonstrate the ability to utilise various cues of competence when determining whom to trust and learn from, including past accuracy, confidence, and expertise (for a review, see Harris, Koenig, Corriveau, & Jaswal, 2018; Robinson & Einav, 2014). Around the same age, children develop the ability to differentiate between benevolent and malevolent informants, showing a preference to selectively trust benevolent individuals (e.g., Doebel & Koenig, 2013; Hamlin & Wynn, 2012). However, they encounter difficulty in recognising that a speaker can intentionally communicate falsehood until the age of 5 or 6 (e.g., Jaswal, Croft, Setia, & Cole, 2010; Mascaro & Sperber, 2009; Vanderbilt, Liu, & Heyman, 2011). This capacity appears to be intertwined with social learning and environmental changes. Factors such as frequent peer-to-peer communicative exchanges and increased oppor-tunities for deception in such interactions contribute to the need for a more vigilant approach to communication (Mascaro & Morin, 2014; Matsui, 2019). This enhanced epistemic vigilance capacity may thus become crucial not only for navigating the complexities of deception but also for the development of irony understanding.

Recognising the falsehood of the literal proposition, labelling its source as unreliable, and using this as a basis to recognise the speaker's dissociative stance are all important components of irony understanding (Mazzarella & Pouscoulous, 2023). While a vast amount of research has been conducted on the developmental trajectories of irony comprehension and epistemic vigilance, no attempt has been made to experimentally assess the interplay between these two capacities. The present study aims to fill this gap by investigating whether irony understanding is facilitated when it targets an unreliable source of information and the extent to which children's performance in irony under-standing can be predicted by their epistemic vigilance skills.

3. The current study

The present study addresses two research questions. First, it examines when children begin to understand irony, moving beyond a mere sensitivity to expectation-context mismatches. Second, and more importantly, it explores the role of epistemic vigilance in children's understanding of irony. To do this, we examined the effect of a reliability manipulation of the target of irony on irony understanding and carried out an exploratory analysis to establish whether epistemic vigilance is a good predictor of irony compre-hension. The study has been preregistered on OSF at the following link: https://osf.io/pqkv7/?view_only=e3f86d8b245942bdb110d88ed2bd037e.

To study irony understanding in children from 5/6 to 6/7 years of age, we used a picture selection task adapted from Köder and Falkum (2021), where children assessed the speaker's feelings by choosing a relevant emoticon (happy versus angry emoticon face) after the violation of a social norm. We capitalised on previous findings that indicate that irony understanding is facilitated when the ironical statement echoes a thought previously expressed in the linguistic context (Hancock et al., 2000), making the social norm salient via an explicit request. To experimentally test children's epistemic vigilance and its role in irony understanding, we manipulated the reliability of the target of the irony based on past accuracy as a source of information and tested children's selective trust when seeking new information from one of two sources: a reliable source versus an unreliable source (Vigilance Measure).

With respect to our first research question, we wanted to test the hypothesis that a more comprehensive irony comprehension develops during middle childhood. We

expected to see an improvement in irony understanding with age, with older children being more likely to interpret irony correctly compared to younger children. As for our main research question, we wanted to assess the hypothesis that the unreliability of the ironical source enhances irony understanding by facilitating the recognition of the dissociative attitude communicated by the ironical speaker. Given that the ironical attitude is often directed towards sources deemed unreliable, we predicted that previous information about the source's unreliability should facilitate irony understanding and make the ironical attitude easier to identify (Mazzarella & Pouscoulous, 2023). Therefore, we predicted that children should demonstrate better performance in the irony comprehension task when the target was unreliable than when it was reliable. We also expected that in our exploratory analysis vigilant children (those who selectively trust the reliable informant over the unreliable one) would be better at understanding irony than non-vigilant ones (those who trusted the unreliable informant).

4. Methods

4.1. Participants

Hundred and one Swiss French-speaking children with no history of speech and language difficulties or any known visual, hearing, or cognitive impairment took part in the study. They were recruited and tested in primary schools in Switzerland and distributed across two school levels: 2 HarmoS (5/6-year-olds) and 3 HarmoS (6/7-year-olds)[1]. Eight participants were tested but were not included in the analysis: one was not a native speaker and did not meet the language requirement, two were excluded because of lack of attention or disengaged behaviour, and two were excluded as they met the pre-registered exclusion criteria related to task performance (i.e., they failed to provide first-pass correct answers to at least three out of six reliability questions across the two blocks). Finally, three participants were excluded as they did not meet the age criterion (i.e., a maximum of 12 months age range between participants within a given school level with the critical window set on 1.08 of the relevant year). Therefore, the final sample included 93 participants: 54 5/6-year-olds (girls = 29, M_{age} = 6;43, range:5;11–6;10 years) and 39 6/7-year-olds (girls = 18, M_{age} = 7;26, range:6;10–7;9 years). Written informed parental consent and the participant's verbal assent were obtained before testing. Each child received an age-appropriate book as a reward for taking part in the study. An adult control group of native Swiss French speakers, composed of graduate and undergraduate students (N = 21, female = 10, M_{age} = 23;52, age range:19–30 years) was included to ensure the experimental design was sound and that their performance was at ceiling. The study was approved by the University's Research Ethics Committee.

4.2. Materials

Children were presented with illustrated stories involving a mother and her two children on a computer screen. They saw three types of stimuli administered in two blocks and always in the same order: Induction trials, Familiarization trials, and Test trials.

---

[1]HarmoS stands for a school concordat for the harmonisation of compulsory education (HarmoS) between the cantons throughout Switzerland. "2Harmos" and "3Harmos" refer to specific years within the primary education cycle, which typically correspond to the 2nd and 3rd years of primary school.

4.2.1. Induction trials

Six induction stories with images and prerecorded dialogues were created, each featuring an interaction between a mother and two siblings (either two boys or two girls) (see Figure 1 for an example of an Induction story). In each story, the two children provided contradictory information when answering a question from their mother. Their testimonies concerned a repeated event (e.g., the red team lost every time versus the red team won every time; the policeman caught the robbers every time versus the policeman let the robbers escape every time, etc.) so that a false answer could not simply be construed as an occasional mistake (Picture 1 in Figure 1). Their respective (in)accuracy was then revealed in a second panel, where the mother established who was right and who was wrong based on a piece of evidence visible in the picture (e.g., the page of a newspaper displaying the winning team; a TV screen displaying the outcome of the robberies) (Picture 2 in Figure 1). After hearing the mother's statement and seeing the visual feedback, participants were asked a comprehension question about the key aspect of the story (Q1), after which the pictures of the two siblings would appear on the screen, and participants were asked two control questions about the past accuracy of the two siblings as sources of information (Q2 and Q3), as well as a selective trust question that was administered only after the last induction story of each block (Q4) (Picture 3 in Figure 1):

Q1: Can you tell me if the red team won every time?
Q2: Can you tell me who was right? Louis or Romain?
Q3: Can you tell me who was wrong? Louis or Romain?
Q4: If you got lost, who would you ask your way to? Louis or Romain?

Children were presented with two induction stories per block. An additional story was administered only if children failed to provide correct answers to the control questions in at least one of the first two stories.
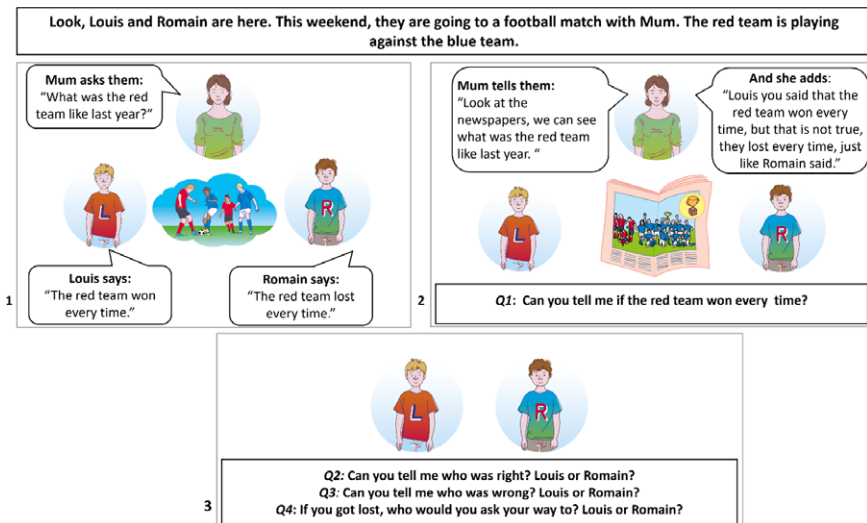

Figure 1. An example of an Induction story (translated into English) followed by a comprehension question (Q1), two control questions (Q2 and Q3), and a selective trust question (Q4). Numbers 1–3 refer to different parts of the experiment in order of their appearance. The text in bold is uttered by the experimenter.

4.2.2. Familiarization trials

The induction trials were followed by two familiarisation trials per block whose aim was to make children familiar with the response measure relevant to the Test trials (happy versus angry face emoticon). Each story involved an interaction between the mother and the two siblings. In each story, the siblings did either something positive (e.g., they nicely decorated a Christmas tree) or something negative (e.g., they destroyed some plants while playing football), thus leading the mother to express a positive remark (e.g., "What a beautiful Christmas tree! You have decorated it really nicely!") or a negative one (e.g., "What a mess! You've destroyed all my plants!"). Each story was accompanied by one illustration depicting the outcome of the siblings' actions while the mother would be depicted from the back to avoid any inference from the facial expression. After hearing the Mum's statement, participants were asked to indicate how the mother was feeling by pointing to one of the two emoticons ("Can you tell me how is Mum feeling? Can you please point to the image?"). The mother's statement was prerecorded in a natural tone of voice for the positive and negative outcomes.

4.2.3. Test trials

The material in the Test trials comprised prerecorded and illustrated stories involving the mother and one of the two siblings (presented in the Induction trials as either reliable or unreliable). All the stories were structured in the following way: (i) the mother would address a request to the child (pre-recorded speech), reflecting some norm-based expectation, (ii) the request of the mother would either be fulfilled or disappointed by the child, (iii) regardless of the outcome, the child would claim that the desired outcome had been achieved (pre-recorded speech), and (iv) the mother would react to the situation by uttering a target statement: Literal praise, Literal criticism, Irony, or Control (pre-recorded speech). Each type of target sentence was recorded by a trained actress with a naturally distinctive tone of voice. After each story, a picture selection task was used to test participants' understanding of the target utterance through the selection of an appropriate emoticon (happy versus angry face) to depict the inner feelings of the mother (adapted from Köder & Falkum, 2021).

Participants saw ten stories across the two blocks, each accompanied by six colourful illustrations depicting all the relevant aspects of the story. Each story required a child to live up to some social norm belonging to one of the following general themes: keeping things or oneself clean, wearing appropriate clothes, and putting something in the right place (see Figure 2 for an example of a Test story). Two images representing either the desired or undesired state of affairs (e.g., a clean room versus a messy room with toys all over the floor) were presented along with a comprehension question to check their understanding of the mother's request. The request of the mother could either be fulfilled or disappointed, but importantly, the negative outcome (e.g., Picture 5a in Figure 2) differed from the undesirable state of affairs depicted in the previous step (comprehension question – Picture 2 in Figure 2). More specifically, we used an image depicting an intermediate negative outcome to make it clear that the child invested some minimal effort into satisfying the request, although the outcome fell short of anything satisfactory. After seeing the positive outcome, the mother would utter an evaluative remark in the form of Literal praise (e.g., "Well done! Your room is really tidy!") while in case of a negative outcome, she could produce either a Literal criticism ("That's bad! Your room is messy!") or an Ironic criticism that had the same form as the Literal praise but was uttered with an ironic intonation. A control condition was included in which the presence of a
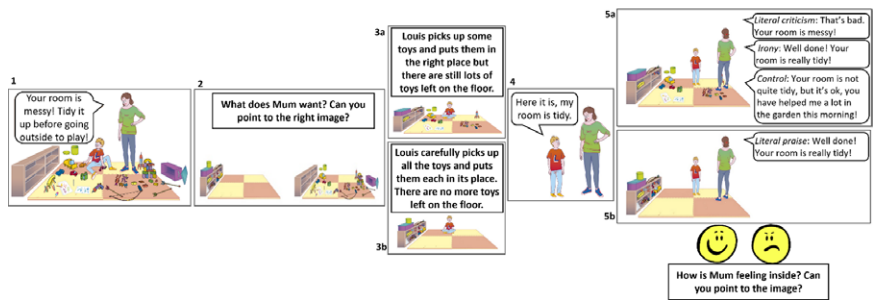
Figure 2. An example of a Test story (translated into English) for all four conditions. Numbers 1–5 correspond to different parts of the story in order of appearance. Parts 3 and 5 have two possible accompanying pictures with each participant being exposed to only one. The text in bold is uttered by the experimenter.

mismatch between the expectation of the mother and the context was followed by a literal, positive statement (e.g., "Your room is not quite tidy, but it's ok, you have helped me a lot in the garden this morning.").

The original scripts for all scenarios in French and their translations in English are available on the Open Science Framework (OSF) at the following link: https://osf.io/y3fet/?view_only=1748761bfc4b4eecadaf30de356e132f.

4.3. Procedure

Each child was tested separately in a quiet room at their school. They were seated next to the experimenter using a child-size chair and table and were presented with the stimuli on a computer screen using PowerPoint. The experimenter first familiarised children with the equipment (computer screen, speakers, cameras) and the overall procedure. The testing of each child was completed within a single uninterrupted session divided into two separate blocks of identical structure, each involving two Induction trials, followed by two Familiarisation trials and five Test trials. The session lasted approximately 15 minutes and was videotaped for data coding (see Table 1 for an overview of the different phases of the experiment in the order of their appearance).

Table 1. An overview of different phases of the experiment in the order of appearance. The order and the number of stories were identical for both blocks. The reliability of the target of the irony was counterbalanced across blocks (e.g., the target of the irony was presented as reliable in Block 1 and unreliable in Block 2). One block involved stories with two girls and their mother, and the other block involved stories with two boys and their mother, and their order was counterbalanced across participants.

	Block 1	Block 2
Induction phase	2 stories with reliability manipulation + vigilance measure	2 stories with reliability manipulation + vigilance measure
Familiarization phase	1 story with a positive outcome 1 story with a negative outcome	1 story with a positive outcome 1 story with a negative outcome
Test phase	1 Irony 1 Literal criticism 1 Literal praise 1 Irony 1 Control	1 Irony 1 Literal criticism 1 Literal praise 1 Irony 1 Control

In the Induction trials, children were first introduced to the main characters and the story setting in which the reliability manipulation of the two siblings was explicitly induced (see Figure 1). After each story, the experimenter asked three questions to check the understanding of the story and the reliability manipulation (Q1, Q2, and Q3). The last induction story ended with an additional selective trust measure to assess children's ability to selectively trust the reliable over the unreliable informant (Q4). The experimenter would allow participants to listen again to the testimonies of the siblings and the mother's feedback if they failed to answer correctly to the comprehension and control questions (Q1–Q3) and correct them at the second failed attempt. No feedback was given to the children following the selective trust question (Q4). An additional story was administered if children failed to provide correct answers to the control questions in at least one of the two stories. In each block, stories with two siblings of the same gender were used to avoid any confounding effect of gender. In each block, one of the siblings was the addressee in the Test trials, and the character's reliability was counterbalanced across blocks, with some children being exposed to the reliable informant in the girls' block and others in the boys' block. Block order was also counterbalanced with some children seeing the girls' block first and others the boys' block first.

Before proceeding to the Test phase, two familiarisation trials were administered to make sure that children understood correctly the meaning of the two emoticons (angry versus happy face) used subsequently to test irony understanding. The experimenter corrected children when they made a mistake and explained why the specific emoticon was used in the given situation to make sure that they were able to differentiate them correctly. The valence of the stories was fixed so that the first story in each block always had a positive outcome (warranting a happy face) while the second one always had a negative outcome (calling for an angry face emoticon).

Finally, participants were exposed to the Test trials involving an interaction between the mother and one of the two siblings (see Figure 2). After being introduced to the story setting and listening to the mother's request (Picture 1 in Figure 2), a comprehension question was administered by the experimenter to ensure that they were following the story (Picture 2 in Figure 2). Then the experimenter would provide a description of the action of the sibling (Picture 3a, b in Figure 2) and their statement about the situation (Picture 4 in Figure 2). Finally, participants heard the prerecorded target utterance while the image depicted the actual state of affairs (Picture 5a, b in Figure 2). The same procedure applied to all the stories. The position of the correct emoticon in each story was counterbalanced across children with each seeing them displayed on the same side on each trial. The position of the correct image in the comprehension question was assigned a fixed pattern across the five stories (Right, Left, Left, Right, Left), and the order of the conditions presented was also fixed (Irony, Literal Criticism, Literal Praise, Irony, Control). The task had a within-subject design with each child seeing two stories in each of the Literal praise, Literal criticism, and Control conditions, as well as four stories in the Irony condition.

4.4. Coding

4.4.1. Induction trials

In the Induction trials, the answers to the first control question (Q2: "Can you tell me who was right?") were deemed correct if the child chose the reliable informant and incorrect if the unreliable informant was selected. The reverse applied to the second control question (Q3: "Can you tell me who was wrong?"). As for the selective trust question (Q4. "If you

got lost, who would you ask your way to?"), the answers were considered correct if the child chose the reliable informant and incorrect when the unreliable informant was selected.

4.4.2. Test trials

In the Test trials, the answers to the comprehension question were considered correct if the child chose the picture representing the desired state of affairs and incorrect if the picture depicting the undesired one was selected. The answers in the Utterance comprehension task were coded differently depending on the condition (see Table 2 for the coding rules for different conditions in the Utterance comprehension task).

5. Results

The dataset and the R scripts for generalised linear mixed-effects models, descriptive statistics, and supplementary materials are available on the Open Science Framework web platform at the link (https://osf.io/y3fet/?view_only=1748761bfc4b4eecadaf30de356e132f). Data processing, analyses, and plotting were conducted in R 4.2.3 (R Core Team, 2023). We decided to fit theoretically driven models for our main hypotheses with further simplification of the random part in case they failed to converge. We checked model assumptions using the DHARMa package for diagnostic inspection (Hartig, 2018). Analysis of Deviance (with type III Wald Chi-square test) was used to check the significance of each term, and only those with significant Wald Chi-square test were considered significant. To further investigate and interpret specific contrasts, we used Tukey Corrected least square means post-hocs, using the emmeans package (Lenth, 2016).

As a first step, we separately checked the possible association between the independent variables Gender and Block and the dependent variable Performance (which measures performance across different conditions). A Mann–Whitney U test was conducted to compare the distribution of Performance scores between the two Gender groups, and it

Table 2. Coding of responses for different conditions in the Utterance comprehension task

Condition	Correct	Incorrect
Literal praise	🙂	🙁
Literal criticism	🙁	🙂
Irony	🙁	🙂
Control	🙂	🙁

showed no significant difference ($W = 1907$, $p = .902$). Therefore, the variable Gender was excluded from further analyses. We then performed a Wilcoxon signed-rank test with continuity correction to assess the differences in the distributions of Performance scores between different blocks. The test revealed a significant difference between the two blocks ($V = 260.5$, $p = .005$), with lower performances in the second block than in the first one. Therefore, the variable Block was included as a factor in the subsequent analyses to account for its impact on the dependent variable Performance.

Before presenting the findings of the main analysis, it is worth mentioning that, across all our analyses, we found a significant order effect. This effect cannot be due to the content of each block, since they were counterbalanced. Moreover, as it went in the direction of a decline in performance from the first to the second block, we attributed this to fatigue and will not discuss this further.[2]

We then performed some descriptive statistical analyses to inspect the dataset and calculate mean performance in percentages (%) for each Utterance type (Literal criticism (LC), Literal praise (LP), Irony, and Control), and age group (5/6-year-olds, 6/7-year--olds, and adults). We ran multiple comparisons t-tests to check whether the performance for each Utterance Type was above chance level and applied the Holm-Bonferroni correction to adjust for multiple comparisons. The results of t-tests show that 6/7-year-olds were performing at ceiling for Literal criticism and Literal praise, while 5/6-year-olds were almost at ceiling for both Literal criticism ($M = 0.98$, $SD = 0.09$, $p < .001$) and Literal praise ($M = 0.99$, $SD =0.07$, $p < .001$). Both groups were performing above chance but not at ceiling for Irony and Control. More precisely, in the Control condition, the performance of both 5/6-year-olds ($M = 0.70$, $SD = 0.41$, $p = .001$) and 6/7-year-olds ($M = 0.85$, $SD = 0.33$, $p < .001$) demonstrated a notable proficiency in the assigned task, as did their performance in the Irony condition: 5/6-year-olds ($M = 0.67$, $SD = 0.43$, $p = .007$) and 6/7-year-olds ($M = 0.74$, $SD = 0.40$, $p = .001$). The control group of adults achieved ceiling performance in Literal criticism and Control conditions, responding correctly to all the questions, and almost at ceiling for Literal praise ($M = 0.98$, $SD = 0.11$, $p < .001$). They performed significantly above chance, albeit not at ceiling, in the Irony condition ($M = 0.90$, $SD = 0.23$, $p < .001$) (See Figure 3).

To investigate the development of irony understanding, we fitted a theoretically driven generalised linear mixed-effects model (GLMM, binomial family, logit link) using the "lme4 package" on a dataset involving only ironic utterances to examine the effect of Age and Block as fixed effects on Irony Performance as the dependent variable. The random part involved random intercepts for subjects and items. The model failed to converge, and Items showed no variance, so we included only by-subject adjustments to the intercept in the analysis. The Analysis of Deviance on the model showed an effect of Block on Irony Performance and no significant effect of Age. To further inspect this result, we used the Tukey Corrected least square test for marginal means, which showed a significant difference in Performance in Block 1 compared to Block 2 ($B = 1.16$, $z = 2.04$, $p = .041$). To assess irony understanding more stringently, we created a data subset that included only those participants who passed the Control condition (i.e., with a Control score = 2). This included a total of 64 participants (33 5/6-year-olds and 31 6/7-year-olds). We fitted a generalised linear mixed effects model with Age and Block as fixed effects,

---

[2]We have also run a separate analysis of children's performance for each block that supports our explanation that this effect was due to tiredness, as it was observed in Block 2 only and particularly in younger children (See Supplementary materials, OSF link: https://osf.io/y3fet/?view_only=1748761bfc4b4ee cadaf30de356e132f).
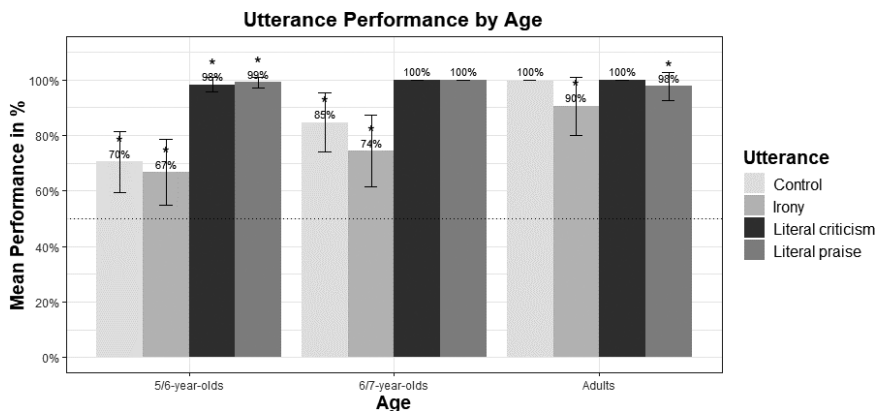
Figure 3. Percentages of correct picture choice for different utterance types (Literal praise, Literal criticism, Irony, Control) and age groups (5/6-year-olds,6/7-year-olds, and Adults). *Note.* The dotted line indicates chance-level performance. Asterisk represents significant values based on the results of multiple comparisons t-tests using Holm Bonferroni correction ($p < .050$). Error bars indicate 95% confidence intervals. No t-tests were conducted for Literal criticism for 6/7-year-olds and Adults, Literal praise for 6/7-year-olds and Control for Adults groups due to the data being essentially constant.

by-subject, and by-item adjustments to the Intercept as random effects, and Irony Performance as the dependent variable. The Analysis of Deviance suggested an effect of Age and an effect of Block. To further inspect these results, we used the Tukey Corrected least square test for marginal means, which confirmed this significant difference between Irony Performance in Block 1 and Block 2 ($B = 1.34$, $z = 2.27$, $p = .023$) and revealed a significant difference in Irony Performance of 5/6-year-olds and 6/7-year-olds who passed the Control condition ($B = -5.19$, $z = -2.03$, $p = .042$). These findings are also in agreement with the descriptive statistical analysis. The results of one-sample t-tests against chance showed that the mean Irony Performance for 5/6-year-olds (M = 0.52, 52%) was not statistically different from chance ($t(32) = 0.29$, $p = .772$), while the mean Irony Performance for 6/7-year-olds (M = 0.70, 70%) was statistically different from chance level ($t(30) = 2.67$, $p = .012$).

To examine the effect of our reliability manipulation on Irony performance, we fitted a theoretically driven generalised linear mixed-effects model on a dataset consisting only of participants who passed the control condition. The model included Reliability (Reliable versus Unreliable) and Block as within-subjects factors, Age as a between-subjects factor, and the interaction between Age and Reliability. The random structure included random intercepts for subjects and items and a by-subject random slope for Reliability, but it failed to converge. We then simplified the random structure of the model by removing the random slope, and the model converged. The Analysis of Deviance revealed an effect of Block and a tendency for Age effect, while no significant effect of Reliability or interaction between Reliability and Age was observed. Tukey Corrected least square means test revealed a significant difference in Irony Performance between Block 1 and Block 2 ($B = 1.26$, $z = 2.09$, $p = .036$).

To investigate the role of epistemic vigilance in irony understanding, we conducted an exploratory analysis to assess whether performance on the selective trust question was a good predictor of Irony Performance. A one-sample t-test was conducted to examine whether children's overall performance in the selective trust task significantly differed

from chance level. The results revealed a significant difference, $t(63) = 15.19$, $p < .001$, indicating that it significantly deviated from the expected chance level. We then created the variable Vigilance, with two levels: vigilant (Trust score = 2) and non-vigilant (Trust score = 0 or 1).

We started by fitting a generalised linear mixed effects model with Irony Performance as the dependent variable and Vigilance, Age, Reliability, and Block as fixed effects, and the resulting interaction between Vigilance, Age, and Reliability in the fixed effect structure. The random structure included random intercepts for subjects and items and a by-subject random slope for Reliability, but it failed to converge. We first tried to simplify the random structure of the model by removing the random slope and also by including by-subject adjustments to the intercept only, but the model still failed to converge. We then simplified the fixed effect structure and fitted a model with Vigilance, Age, Block, and Reliability, and the interaction between Reliability and Vigilance in the fixed effect structure and random intercepts for subjects only.[3] The Analysis of Deviance revealed a main effect of Vigilance and an effect of Block on Irony Performance. More precisely, Tukey Corrected least square means test revealed a significant difference in Irony Performance between non-vigilant and vigilant children ($B = -17$, $z = -4.14$, $p < .001$), with vigilant children performing significantly better than non-vigilant ones. As with previous analyses, Tukey Corrected least square means test also showed a significant difference in Irony Performance between Block 1 and Block 2 ($B = 1.7$, $z = 2.54$, $p = .011$). No effect of Age, Reliability, or interaction between Vigilance and Reliability was found.[4] Finally, to establish that Vigilance correlated with Irony Performance, but not with performance in a comparable task, we carried out the same analysis but used the performance in the Control condition as our dependent variable. We fitted a model with Vigilance, Age, Block, and Reliability, and the interaction between Reliability and Vigilance in the fixed effect structure and random intercepts for subjects only. The Analysis of Deviance revealed only a significant effect of Block on Control Performance, $\chi^2(1) = 11.18$, p < .001. More precisely, Tukey Corrected least square means test showed a significant difference in performance in the control condition between Block 1 and Block 2 ($B = 4.74$, $z = 1.42$, $p < .001$). No effect of Age, Reliability, Vigilance, or interaction between Vigilance and Reliability was found (for the full model selection procedure and model outputs see Supplementary Materials, OSF link:https://osf.io/y3fet/?view_only= 1748761bfc4b4eecadaf30de356e132f).

6. Discussion

The present study investigated irony understanding in children and aimed at disentangling it from a mere sensitivity to the presence of an expectation-context mismatch. To do this, we used an offline picture selection task adapted from Köder and Falkum (2021) and added a control condition that allowed us to identify the subset of children whose responses to the comprehension task fully relied on the integration of the story context

---

[3]An R script with the analyses on 5/6-year-olds only as a group of children who are in a transitional phase in irony acquisition, between a possible literal bias and more robust irony understanding have been included in the Supplementary materials, OSF link:https://osf.io/y3fet/?view_only=1748761bfc4b4eecadaf30de356e132f). The results are consistent with the analysis carried out for the entire dataset.

[4]Full model outputs are presented in Supplementary Materials, OSF link: https://osf.io/y3fet/?view_only= 1748761bfc4b4eecadaf30de356e132f.

and the target utterance. Both 5/6-year-olds and 6/7-year-olds performed well in the irony comprehension task. However, when a more stringent criterion was employed (passing the Control condition) a developmental effect was observed. More specifically, data from children who passed the Control condition indicate that 6/7-year-olds, but not 5/6-year-olds, reliably understand irony above chance.

The main aim of this study was to test experimentally the role of epistemic vigilance in irony comprehension. We did so by assessing whether irony comprehension was facilitated when the ironical target was unreliable and by conducting an exploratory analysis to see if epistemic vigilance was a good predictor of irony understanding. Contrary to our expectations, children were not better at recognising irony when this was directed towards an unreliable target as opposed to a reliable one. Crucially, though, our vigilance measure turned out to be a significant predictor of irony performance, with vigilant children showing a better understanding of irony than non-vigilant children.

6.1. Developmental pattern of irony comprehension

Our findings showed that, for both 5/6- and 6/7-year-olds, irony understanding was not as good as the understanding of its literal counterparts (literal criticism and literal praise). This aligns with prior findings in the developmental literature showing that irony is more challenging than literal remarks (see e.g., Banasik-Jemielniak & Bokus, 2019; Hancock et al., 2000; Köder & Falkum, 2021) as well as with irony processing studies with adults and adolescents, showing increased cognitive costs for processing irony compared to literal statements (for a review see Olkoniemi & Kaakinen, 2021). The generalised linear mixed model used to inspect the developmental trajectory of irony comprehension revealed no age effect with both groups demonstrating a considerable understanding of irony. This pattern of results is consistent with the current findings in the literature suggesting that children start to demonstrate some understanding of irony at around age 6 (e.g., Dews et al., 1996; Glenwright & Pexman, 2010). It is important to stress that in the present study, children were exposed only to counterfactual forms of irony, which are believed to be the easiest to detect, with the echoic nature of the utterance made explicit in the context (e.g., Filippova & Astington, 2008; Hancock et al., 2000). They were asked to answer non-verbally to the test question to avoid imposing additional linguistic or metalinguistic demands (e.g., Köder & Falkum, 2021). However, this may not entirely reflect how ironic communication occurs in real-life situations, where children may encounter less frequent forms that are more difficult to interpret, or when irony may occur in more linguistically complex and less supportive or ambiguous contexts. Nevertheless, the present study stands out as one of the few attempts to explore irony comprehension in younger children; it contrasts with much of the developmental literature, which predominantly focuses on older participants, often after the age of 7 (see also e.g., Angeleri & Airenti, 2014; Banasik-Jemielniak & Bokus, 2019; Köder & Falkum, 2021).

Interestingly, though, when applying a more stringent criterion (excluding participants who failed the Control condition) to assess irony understanding, only 6/7-year-olds performed above chance in our irony comprehension task. Köder and Falkum (2021) showed that from the age of 3, children show an implicit preference for the angry face emoticon in the Irony condition, thus revealing a sensitivity to the mismatch between context and expectations. They also showed that children explicitly manifest this preference by picking the correct emoticon, that is, the angry face, at 5/6 years of age. This aligns

well with the developmental literature showing that the capacity to simultaneously process and consider multiple cues and start interpreting incongruencies between ironic expression and contextual elements does not develop fully before the age of 8 or 9 and that younger children most often tend to demonstrate either lexical or contextual biases (e.g., Aguert et al., 2010; Falkum, 2022; Matsui, 2019). Hence, our findings suggest that a more sophisticated and specific understanding of irony (e.g., demonstrated by passing our control condition) may emerge even later. They underscore the importance of introducing an adequate control condition to distinguish between participants who can consider both the utterance and the context when making their choice and those who are merely guided by the negative valence of the story outcome. Moreover, they suggest that the path to a full-fledged understanding of irony may pass through the recognition that a mismatch between expectations and reality – which younger children can detect – may act as a motive for the speaker to express a dissociative, critical attitude via irony. Future research should incorporate our control condition to capture a more comprehensive understanding of irony, enhancing the validity and depth of the findings.

6.2. Irony and epistemic vigilance

The main aim of the study was to experimentally investigate the role of epistemic vigilance in irony understanding. Our decision to manipulate the reliability level of the target of the ironic remark was based on the idea that the dissociative attitude communicated by the speaker is typically warranted by the epistemic assessment of the literal content, which is often judged as false or irrelevant and attributed to an unreliable source. We anticipated that information about the unreliability of the ironical source (the target of the irony) would make it easier to recognise the speaker's ironical intent, as discussed by Mazzarella and Pouscoulous (2023). We also expected that the performance in the epistemic vigilance task would be a good predictor of children's performance in the irony comprehension task. Indeed, we expected that children who displayed the capacity to actively evaluate the reliability of two informants and calibrate trust accordingly in the selective trust tasks would also be more likely to actively assess the competence and benevolence of the ironical speaker in the irony comprehension task, thus facilitating the recognition of the ironical attitude.

Importantly, our exploratory analysis of children's epistemic vigilance skills (as measured by the selective trust choice) provides compelling evidence supporting our hypothesis that epistemic vigilance may scaffold irony understanding. We found a positive relationship between the performance of children in the selective trust question and their performance in the irony task. More precisely, more vigilant children (those who selectively trusted the reliable informant in the epistemic vigilance stories) demonstrated a better understanding of irony than non-vigilant children regardless of the reliability condition. That the observed link between irony and epistemic vigilance is not due to some other overarching factor (e.g., executive functions, language capacities) is confirmed by our additional analysis on children's performance in the control condition, showing that their epistemic vigilance (measured by their performance in the selective trust task) was not related to their successful performance in the control condition. This confirms that while general factors may contribute to children's performance in these tasks, they are not sufficient to explain their performance in the irony comprehension task, and that epistemic vigilance specifically contributes to detecting irony. Interestingly, although already by the age of 4 children detect different cues of competence and orient their trust towards informants that display them

(Clément, Koenig, & Harris, 2004; Einav & Robinson, 2011), in our selective trust task 5/6-year-olds and 6/7-year-olds performed well but were not at ceiling. The interindividual differences observed may be due to the fact that our task required the exercise of more robust epistemic vigilance skills than most selective trust tasks employed in the literature. Children's assessment of the incoming information could not capitalise on pre-existing background knowledge (as is the case in most paradigms involving object labelling or pointing to a hidden object location; see Harris et al., 2018) and the informants were not explicitly labelled as (in)competent or (dis)honest (thus leaving open the motivation for inaccuracy).

Contrary to our expectations, we found no effect of the reliability manipulation on irony understanding across the two groups. The absence of prior experimental literature on this topic makes it challenging to draw definitive conclusions regarding this null result. However, exploring potential limitations of the study could provide valuable insights for future research.

One possible limitation is that the reliability manipulation used in the present study may have been too weak or not sufficiently reinforced throughout the entire task. Our choice to manipulate the reliability of the two informants based on past (in)accuracy was informed by the findings from the selective trust literature, which indicate that as children grow older (from the age of 4), they tend to perceive epistemic cues (related to an informant's past accuracy or expertise) as stronger indicators than social cues (related to one's social status, relationships, or personality) for obtaining accurate information and avoiding misinformation (see e.g., Tong, Wang, & Danovitch, 2020). However, while in the Induction phase, children were asked to differentiate between the reliable and the unreliable informants and choose which informant to trust, their reliability may have been less salient or relevant during the test phase or less accessible due to cognitive constraints. Hence, future studies may attempt to simplify the structure of the stories and reinforce the reliability manipulation. For instance, explicitly presenting a personality trait associated with one's (in)accuracy might enhance the manipulation's strength and stability across the task. Previous studies on trait reasoning have suggested a developmental shift in children's appreciation and use of personality traits to predict and explain behaviours around the age of 7 to 8 years (i.e., middle childhood). At this stage, children begin to understand and attribute personality traits as stable properties with implications for future behaviour, while younger children appear to associate personality traits with specific situations (see e.g., Kalish, 2002, but see also Liu, Gelman, & Wellman, 2007). Furthermore, to increase the relevance of the reliability manipulation, future studies could consider introducing obvious consequences or rewards for the choice of the accurate informant. Creating a more engaging task with clear incentives may encourage children to pay closer attention to the relevant aspects of the story, enhancing their motivation and involvement throughout the entire task (e.g., Brosseau-Liard, 2014; Ronfard, Nelson, Dunham, & Blake, 2019). It is also worth noting that, in our study, only simple counterfactual forms of irony with explicit antecedents were used, thus possibly mitigating the effect of the reliability manipulation. Such an effect may be more likely to be detected in the presence of more complex or less frequent ironic forms (e.g., ironic praise), particularly in the absence of other contextual and non-contextual cues.

Nevertheless, these represent the first empirical findings linking irony development to that of epistemic vigilance. They corroborate the hypothesis that children need to be able to actively assess the reliability of the speaker – that is, their competence and benevolence – to recognise the ironical intent and avoid misinterpreting irony as an error or a lie (Mazzarella & Pouscoulous, 2021, 2023). This form of epistemic vigilance towards the

source, captured by our selective trust task, is indeed relevant during the test phase to establish that the ironical speaker (here, the mother) is not stating something false because they are incompetent (e.g., that they really believe that the room is tidy, although there are still lots of toys all over the floor) or because they intend to mislead the addressee (e.g., they want the addressee to believe that that is the case, although they know that the room is actually messy). Vigilant children are thus in a better position to recognise that a deliberate falsehood may be motivated by an ironical intent.

7. Conclusions

The present study provides the first empirical evidence supporting the role of epistemic vigilance in the development of irony comprehension and the recognition of the dissociative attitude communicated by the ironical speaker. It also reinforces the claim that irony requires a more sophisticated socio-cognitive repertoire compared to other non-literal uses of language, contributing to its relatively prolonged developmental trajectory. Our findings do not question the role played by factors other than epistemic vigilance (including ToM, executive functions, language skills, and emotion recognition, for review, see Pexman, 2023) in the development of full-fledged irony understanding. However, as a pioneering attempt to explore the development of irony comprehension through the lens of its interplay with the epistemic vigilance capacity, this study opens up a new line of research in developmental pragmatics and provides new insights into how children achieve a mature understanding and use of complex language that characterises adults' communication.

References

Aguert, M., **Laval, V.**, **Le Bigot, L.**, & **Bernicot, J.** (2010). Understanding expressive speech acts: The role of prosody and situational context in French-speaking 5- to 9-year-olds. *Journal of Speech, Language, and Hearing Research*, **53**(6), 1629–1641. https://doi.org/10.1044/1092-4388(2010/08-0078)

Angeleri, R., & **Airenti, G.** (2014). The 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**(2), 133–146. https://doi.org/10.1037/cep0000011

Banasik-Jemielniak, N., & **Bokus, B.** (2019). Children's comprehension of irony: Studies on polish-speaking preschoolers. *Journal of Psycholinguistic Research*, **48**(5), 1217–1240. https://doi.org/10.1007/s10936-019-09654-x

Brosseau-Liard, P. E. (2014). Selective, but only if it is free: Children trust inaccurate individuals more when alternative sources are costly: cost and selectivity. *Infant and Child Development*, **23**(2), 194–209. https://doi.org/10.1002/icd.1828

Capelli, C. A., **Nakagawa, N.**, & **Madden, C. M.** (1990). How children understand sarcasm: The role of context and intonation. *Child Development*, **61**(6), 1824. https://doi.org/10.2307/1130840

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

Clément, F., **Koenig, M.**, & **Harris, P.** (2004). The ontogenesis of trust. *Mind & Language*, **19**(4), 360–379. https://doi.org/10.1111/j.0268-1064.2004.00263.x

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**(3), 257–285. https://doi.org/10.1080/15248370802247939

Colston, H. L., & **O'Brien, J.** (2000). Contrast and pragmatics in figurative language: Anything understatement can do, irony can do better. *Journal of Pragmatics*, **32**(11), 1557–1583. https://doi.org/10.1016/S0378-2166(99)00110-1

Deamer, F. (2013). An investigation into the processes and mechanisms underlying the comprehension of metaphor and hyperbole [Doctoral dissertation, University College London]. https://discovery.ucl.ac.uk/id/eprint/1420534

Deliens, G., **Antoniou, K.**, **Clin, E.**, **Ostashchenko, E.**, & **Kissine, M.** (2018). Context, facial expression and prosody in irony processing. *Journal of Memory and Language*, **99**, 35–48. https://doi.org/10.1016/j.jml.2017.10.001

Demorest, A., **Meyer, C.**, **Phelps, E.**, **Gardner, H.**, & **Winner, E.** (1984). Words speak louder than actions: understanding deliberately false remarks. *Child Development*, **55**(4), 1527. https://doi.org/10.2307/1130022

Demorest, A., **Silberstein, L.**, **Gardner, H.**, & **Winner, E.** (1983). Telling it as it isn't: Children's understanding of figurative language. *British Journal of Developmental Psychology*, **1**(2), 121–134. https://doi.org/10.1111/j.2044-835X.1983.tb00550.x

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**(6), 3071. https://doi.org/10.2307/1131767

Di Paola, S., **Domaneschi, F.**, & **Pouscoulous, N.** (2020). Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. *Journal of Pragmatics*, **156**, 235–251. https://doi.org/10.1016/j.pragma.2019.08.008

Doebel, S., & **Koenig, M. A.** (2013). Children's use of moral behavior in selective trust: Discrimination versus learning. *Developmental Psychology*, **49**(3), 462–469. https://doi.org/10.1037/a0031595

Einav, S., & **Robinson, E. J.** (2011). When being right is not enough: Four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological Science*, **22**(10), 1250–1253. https://doi.org/10.1177/0956797611416998

Falkum, I. L. (2022). The development of non-literal uses of language: Sense conventions and pragmatic competence. *Journal of Pragmatics*, **188**, 97–107. https://doi.org/10.1016/j.pragma.2021.12.002

Falkum, I. L., **Recasens, M.**, & **Clark, E. V.** (2017). "*The moustache* sits down first": On the acquisition of metonymy. *Journal of Child Language*, **44**(1), 87–119. https://doi.org/10.1017/S0305000915000720

Filippova, E., & **Astington, J. W.** (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development*, **79**(1), 126–138. https://doi.org/10.1111/j.1467-8624.2007.01115.x

Filippova, E., & **Astington, J. W.** (2010). Children's understanding of social-cognitive and social-communicative aspects of discourse irony. *Child Development*, **81**(3), 913–928. https://doi.org/10.1111/j.1467-8624.2010.01442.x

Fuchs, J. (2023). 40 years of research into children's irony comprehension: A review. *Pragmatics & Cognition*, **30**(1), 1–30. https://doi.org/10.1075/pc.22015.fuc

Glenwright, M., & **Pexman, P. M.** (2010). Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language*, **37**(2), 429–451. https://doi.org/10.1017/S0305000909009520

Hamlin, J. K., & **Wynn, K.** (2012). Who knows what's good to eat? Infants fail to match the food preferences of antisocial others. *Cognitive Development*, **27**(3), 227–239. https://doi.org/10.1016/j.cogdev.2012.05.005

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**(2), 227–248. https://doi.org/10.1207/S15327647JCD010204

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, **69**(1), 251–273. https://doi.org/10.1146/annurev-psych-122216-011710

Hartig, F. (2018). DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models (Version 0.4.6) [Software]. Retrieved from https://CRAN.R-project.org/package=DHARMa

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, **21**(10), 1541–1547. https://doi.org/10.1177/09567976 10383438

Kalish, C. (2002). Children's predictions of consistency in people's actions. *Cognition*, **84**(3), 237–265. https://doi.org/10.1016/S0010-0277(02)00052-5

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

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

Langdon, R., Davies, M., & Coltheart, M. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind & Language*, **17**(1–2), 68–104. https://doi.org/10.1111/1468-0017.00189

Lenth, R. V. (2016). Least-squares means: The R package LS means. *Journal of Statistical Software*, **69**(1), 1–33. doi:10.18637/jss.v069.i01.

Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of young children's trait understanding: Behavior-to-trait Inferences and trait-to-behavior predictions. *Child Development*, **78**(5), 1543–1558. https://doi.org/10.1111/j.1467-8624.2007.01082.x

Mascaro, O., & Morin, O. (2014). Gullible's travel: How honest and trustful children become vigilant communicators. In *Trust and skepticism* (pp. 69–82). Psychology Press.

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition*, **112**(3), 367–380. https://doi.org/10.1016/j.cognition.2009.05.012

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**(3), 301–317. https://doi.org/10.1080/17405629.2012.672272

Matsui, T. (2019). Component processes of irony comprehension in children: Epistemic vigilance, mind-reading and the search for relevance. In K. Scott, B. Clark, & R. Carston (Eds.), *Relevance, pragmatics and interpretation* (1st ed., pp. 231–239). Cambridge University Press. https://doi.org/10.1017/978110829 0593.020

Mazzarella, D., & Pouscoulous, N. (2021). Pragmatics and epistemic vigilance: A developmental perspective. *Mind & Language*, **36**(3), 355–376. https://doi.org/10.1111/mila.12287

Mazzarella, D., & Pouscoulous, N. (2023). Ironic speakers, vigilant hearers. *Intercultural Pragmatics*, **20**(2), 111–132. https://doi.org/10.1515/ip-2023-2001

Olkoniemi, 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**(2), 99–106. https://doi.org/10.1037/cep0000216

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

Pexman, P. M. (2023). Irony and thought: Developmental insights. In R. W. Gibbs, Jr & H. L. Colston (Eds.), *The Cambridge handbook of irony and thought* (1st ed., pp. 181–196). Cambridge University Press. https://doi.org/10.1017/9781108974004.015

Pouscoulous, N. (2023). More than one path to pragmatics? Insights from children's grasp of implicit, figurative and ironical meaning. *Cognition*, **240**, 105531. https://doi.org/10.1016/j.cognition.2023.105531

Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, **156**, 160–167. https://doi.org/10.1016/j.pragma.2019.05.021

R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rivière, E., Klein, M., & Champagne-Lavau, M. (2018). Using context and prosody in irony understanding: Variability amongst individuals. *Journal of Pragmatics*, **138**, 165–172. https://doi.org/10.1016/j.pragma.2018.10.006

Robinson, E. J., & **Einav, S.** (2014). *Trust and skepticism: Children's selective learning from testimony.* Psychology Press.

Ronfard, S., **Nelson, L.**, **Dunham, Y.**, & **Blake, P. R.** (2019). How children use accuracy information to infer informant intentions and to make reward decisions. *Journal of Experimental Child Psychology*, **177**, 100–118. https://doi.org/10.1016/j.jecp.2018.07.017

Schmidt, M. F. H., **Rakoczy, H.**, **Mietzsch, T.**, & **Tomasello, M.** (2016). Young children understand the role of agreement in establishing arbitrary norms—but unanimity is key. *Child Development*, **87**(2), 612–626. https://doi.org/10.1111/cdev.12510

Scianna, C. (2023). When children acquire irony: The role of epistemic vigilance. *Intercultural Pragmatics*, **20**(3), 323–343. https://doi.org/10.1515/ip-2023-3006

Sperber, D., **Clément, F.**, **Heintz, C.**, **Mascaro, O.**, **Mercier, H.**, **Origgi, G.**, & **Wilson, D.** (2010). Epistemic vigilance. *Mind & Language*, **25**(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

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

Tong, Y., **Wang, F.**, & **Danovitch, J.** (2020). The role of epistemic and social characteristics in children's selective trust: Three meta-analyses. *Developmental Science*, **23**(2), e12895. https://doi.org/10.1111/desc.12895

Vanderbilt, K. E., **Liu, D.**, & **Heyman, G. D.** (2011). The development of distrust. *Child Development*, **82**(5), 1372–1380. https://doi.org/10.1111/j.1467-8624.2011.01629.x

Wilson, D. (2009). Irony and metarepresentation. *UCL Working Papers in Linguistics*, **21**, 183–226.

Wilson, D., & **Sperber, D.** (2012). *Meaning and relevance* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139028370

Winner, E., & **Leekam, S.** (1991). Distinguishing irony from deception: Understanding the speaker's second-order intention. *British Journal of Developmental Psychology*, **9**(2), 257–270. https://doi.org/10.1111/j.2044-835X.1991.tb00875.x

Zajączkowska, M. K. (2016). Influence of voice intonation on understanding irony by polish-speaking preschool children. *Psychology of Language and Communication*, **20**(3), 278–291. https://doi.org/10.1515/plc-2016-0017