Journal of Experimental Child Psychology 199 (2020) 104942

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

"Sure I'll help—I've just been sitting around doing nothing at school all day": Cognitive flexibility and child irony interpretation

Maria Zajączkowska, Kirsten Abbot-Smith *

School of Psychology, University of Kent, Canterbury CT2 7NP, UK

ARTICLE INFO

Article history:
Received 7 May 2020
Revised 26 June 2020

Keywords:
Irony
Sarcasm
Executive functioning
Theory of mind
Mentalising
Cognitive flexibility

ABSTRACT

Successful peer relations in older children depend on proficiency with banter, which in turn frequently involves verbal irony. Individual differences in successful irony interpretation have traditionally been attributed to theory of mind. Our premise was that the key factor might in fact be cognitive flexibility, that is, the ability to switch between different perspectives (here, on the same utterance). We also wished to extend the focus of previous irony studies, which have almost exclusively examined *simple irony,* where the literal meaning conflicts with observable physical evidence (e.g., "Great day for a picnic" when viewing a downpour). Therefore, we also examined how children interpreted more *complex irony,* where listeners must consider at a deeper level the common ground shared with the speakers (e.g., general knowledge/cultural common ground or information about the particular speaker). In Study 1, we found that for 6- to 8-year-olds, both cognitive flexibility and theory of mind contributed unique variance to simple irony interpretation while statistically controlling for nonverbal reasoning and structural language standardized scores. Neither inhibitory control, nor working memory, nor general knowledge correlated with irony interpretation. The 6- to 8-year-olds were at floor for complex irony. In Study 2, we found that cognitive flexibility contributed unique variance to how 10- to 12-year-olds interpreted complex irony while controlling for nonverbal reasoning, structural language, and specific knowledge required. We are the first to examine the relationship with cognitive flexibility and conclude that it must be taken into account

* Corresponding author. Fax: +44 1227 82 7030.
E-mail address: k.abbot-smith@kent.ac.uk (K. Abbot-Smith).

https://doi.org/10.1016/j.jecp.2020.104942

when investigating the relationship between theory of mind and irony interpretation.

Introduction

As children get older, one of the most obvious individual differences in their social communication and interaction is the ability to quickly interpret the communicative intent associated with banter, teasing, and joking around with their peers. A vital aspect of such banter and joking often involves proficiency with verbal irony interpretation (e.g., Leggitt & Gibbs, 2000). Verbal irony prototypically involves a contrast between an utterance's literal meaning and its intended meaning (e.g., Wilson, 2013).

In the developmental literature, children's ability to interpret verbal irony has almost ubiquitously been tested using vignettes in which the addressee can observe that the literal meaning cannot be true (see Filippova, 2014, for an overview). For example, if a boy says "Well done!" to a friend when she spills juice down her shirt, the visual contrast with the ruined shirt makes the boy's intended meaning evident. We refer to this type of irony usage as *simple irony*. However, as children become teenagers, they increasingly also use more *complex irony* (e.g., Aijmer, 2019). In Example (1) below, the listener must engage in a more in-depth analysis to determine the speaker's communicative intent.

(1) *Tom:* Can you help me cook the dinner? I'm tired.
Sally: Oh yes, because I have just been sitting around doing nothing at school today.

To date, only Bosco and Bucciarelli (2008) have examined how children interpret simple irony versus complex irony. Their characterization of this distinction focused on the fact that complex irony requires the listener to compute a chain of interferences. Our characterization differs a little in focus. In the simplest forms of irony—and the type most frequently tested in the child literature—the literal meaning conflicts with physical evidence from the real-world context. In contrast, the correct interpretation of complex ironic statements, such as that of Sally's in Example (1) above, must require cultural/general knowledge (here, that people generally do not do "nothing" during the school day). Our aim in the current two studies was to explore which cognitive abilities relate to individual differences in children's proficiency with inferring the communicative intent of verbal irony and whether these relationships with other cognitive abilities differ for simple irony versus complex irony.

Cognitive correlates of irony interpretation

The ability of a child or an adolescent to successfully interpret the communicative intent of an ironic utterance depends on a number of factors. Some influential factors relate to the child's prior language experience. For example, the frequency with which irony is used for certain functions in a particular language community affects ease of acquisition (e.g., Filippova, 2014; see also Banasik-Jemielniak & Bokus, 2019, for a discussion of cross-cultural variation). Furthermore, children build up expectations regarding which types of speaker relationships (e.g., between siblings vs. between strangers) result in more frequent use of irony (e.g., Whalen, Doyle, & Pexman, 2020). Learning the role played by prosody is also arguably dependent on language experience.

Other factors driving proficiency in irony interpretation concern child-internal cognitive abilities. These are the focus of the current article. One cognitive skill that is an obvious prerequisite for irony interpretation is structural language, that is, the words and morphosyntax used to convey conventionalized meaning (e.g., Abbot-Smith, 2020). A large number of studies have found that structural language—or at least vocabulary—relates to children's ability to successfully interpret irony (Angeleri & Airenti, 2014; Caillies, Bertot, Motte, Raynaud, & Abely, 2014; Filippova & Astington, 2008; Godbee

& Porter, 2013; Massaro, Valle, & Marchetti, 2014; Mewhort-Buist & Nilsen, 2019; Nilsen, Glenwright, & Huyder, 2011).

Another cognitive skill that seems likely to relate to irony interpretation is theory of mind. This is the ability to understand that others may have perspectives, knowledge, or beliefs that differ from one's own (e.g., Wellman, 2014). There is a fairly long history to claims that the ability to interpret verbal irony may depend on theory of mind—particularly more advanced theory of mind such as the ability to represent one individual's knowledge of a second individual's knowledge or belief (Happé, 1993). The logic behind this claim is that to understand ironic intent, a child must understand that the speaker has a particular mental state attitude toward the content of her or his utterance. Advanced theory of mind ability is often assessed in tests such as Happé's (1994) "Strange Stories," in which children are asked to explain white lies, double bluffs, and miscommunication. There are indeed a considerable number of studies that found relationships between theory of mind and irony interpretation in children (e.g., Angeleri & Airenti, 2014; Banasik, 2013; Caillies et al., 2014; Filippova & Astington, 2008; Mewhort-Buist & Nilsen, 2013; Nilsen et al., 2011), although not all of them controlled for structural language in their analyses (see Matthews, Biney, & Abbot-Smith, 2018, for a review).

However, other less frequently investigated cognitive abilities may potentially be as important—or more important—for successful irony interpretation. Some likely contenders can be found under the umbrella term of executive functions (EFs), which encompasses higher-level cognitive skills required for cognitive control (Diamond, 2013). One EF component is working memory. This is the ability to maintain and update information held in short-term memory. Although a certain working memory capacity is likely to be a prerequisite for any aspect of language processing and interpretation (e.g., Kidd, 2013), it may come under particular pressure in the types of tasks typically used to assess children's interpretation of verbal irony. These tasks involve a participant listening to a background story vignette about two characters, where the vignette ends with a character making an ironic statement. Thus, the task itself clearly requires a child to retain and build a mental model of a fair amount of information. Nonetheless, a study by Filippova and Astington (2008), which examined relationships between irony interpretation and both advanced theory of mind and working memory, established that the latter—although important—was not the most crucial factor in accounting for how 5-, 7-, and 9-year-olds interpreted irony. That is, whereas the relationship between irony and working memory showed a medium effect size, advanced theory of mind (assessed by a combination of second-order false belief, Strange Stories, and faux pas understanding) still contributed a significant amount of variance to children's irony interpretation scores over and above working memory and structural language.[1]

The second key component of EF is inhibitory control (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), which is the ability to suppress a habitual response or distracting information in order to focus on the task at hand. Inhibitory control is plausibly implicated in irony interpretation because to access the ironic interpretation of a particular utterance, a listener may need to suppress the same utterance's literal interpretation. To date, only Caillies and colleagues have examined whether inhibitory control correlated with child irony interpretation (e.g., Caillies et al., 2014). However, their findings are difficult to interpret. Although they found a significant correlation between irony interpretation and inhibitory control for their sample of typically developing children (aged 6–10 years), this was not the case for their sample of children with attention-deficit/hyperactivity disorder, which is widely thought to involve an impairment in inhibitory control. Most problematic was that Caillies et al. (2014) included only 15 participants per group and did not statistically control for structural language. Therefore, the specificity of the relationship between inhibitory control and irony interpretation in children is as yet unclear.

The importance of cognitive flexibility for irony interpretation

The third important component of EF is cognitive flexibility (Pennington & Ozonoff, 1996), which is the ability to switch between "multiple representations, strategies or responses" (Chevallier et al.,

[1] Nilsen and Bacso (2017) also found a relationship with working memory in their sample of adolescents, but unfortunately their outcome variable conflated irony interpretation with referential communication.

2012, p. 99). The most commonly used experimental means of assessing cognitive flexibility involves Card Sort tasks, in which a participant must first sort cards according to one rule (e.g., same color) and then switch to sorting the same cards using a different strategy (e.g., same shape). Here, the measure of inflexibility is the number of perseverative errors made, that is, where a participant continues to use the previous strategy after the rule switch has occurred. Cognitive flexibility is clearly a multifaceted ability requiring both working memory to retain the current rule and inhibitory control to suppress the previous rule (Diamond, 2013), and thus it is often considered to be the epitome of EF (Zelazo & Müller, 2007). Nonetheless, cognitive flexibility is not merely a composite skill given that it crucially also involves the ability to switch between multiple perspectives simultaneously (Jacques & Zelazo, 2005).

This ability to switch between multiple perspectives is logically essential for irony interpretation for two reasons. The first reason is that to interpret an utterance as ironic, one needs to be aware of not only its nonliteral meaning (to correctly interpret and respond) but also its literal meaning (to understand that usage of this literal meaning is being "mocked" in some sense) (e.g., Wilson, 2013). The second reason why cognitive flexibility is likely to be pivotal is that if irony interpretation is at least partially dependent on some form of theory of mind, this in itself requires a child to simultaneously hold in mind his or her own perspective as well as the perspective of the interlocutor. Indeed, this may be why training cognitive flexibility in preschoolers leads to better performance in theory of mind tasks (Kloo & Perner, 2003). Strangely, despite this logical fundamental relationship between irony interpretation and cognitive flexibility, no study to date has investigated whether cognitive flexibility correlates with irony understanding (cf. Zajączkowska, Abbot-Smith, & Kim, 2020). Thus, the relationship between cognitive flexibility and irony interpretation was the primary research question of the current two studies.

The current studies

We carried out two studies to determine the degree to which cognitive flexibility accounts for unique variance in children's ability to interpret simple and complex irony. To this end, we also assessed children on theory of mind, requisite knowledge, structural language, and nonverbal reasoning. The latter is usually assessed via a form of the Matrices task—a test of visual analogical reasoning that in some cases comprises the only ability assessed by certain IQ tests (e.g., Raven, 1986). Nonverbal reasoning is likely to be important for strategy use in any type of development test. It is also demonstrably highly intercorrelated with EF in typically developing children (e.g., Ardila, Pineda, & Rosselli, 2000; Arffa, 2007). Problematically, nonverbal reasoning to date has been given no consideration in child irony interpretation literature (see Matthews et al., 2018).

For our irony measure, we wished to develop a new method. This was because previous methods either required children to formulate in their own words the speaker's intent (e.g., Angeleri & Airenti, 2014; Caillies et al., 2014; Filippova & Astington, 2008) or required them to choose between two options, namely a literal meaning versus a nonliteral meaning (e.g., Happé, 1993; Massaro et al., 2014; Mewhort-Buist & Nilsen, 2013, 2019; Nilsen et al., 2011). The latter—binary forced choice—is problematic given that children show an awareness that simple ironic utterances are not intended literally long before they are able to determine the actual communicative intent of the utterances (e.g., Hancock, Dunham, & Purdy, 2010). Therefore, they may succeed in a binary forced choice without truly comprehending the ironic meaning.

A binary choice is particularly problematic when complex irony is included; a listener may understand in Example (1) above that Sally has not literally been sitting around doing nothing all day without necessarily understanding Sally's communicative intent. That is, Sally is not necessarily refusing to help cook dinner. Rather, she is primarily expressing her annoyance at the assumption that only Tom is tired. For this reason, for our two studies, the key irony outcome variable required children to choose among three options. Nonetheless, for consistency with the previous literature, in Study 1 we also asked children (prior to the forced choice) to explain in their own words what the speaker meant.

In Study 1, we tested typically developing, monolingual English-speaking children on their ability to interpret simple and complex irony. For this study, we selected children aged 6, 7, and 8 years. This

age group overlaps with, but is somewhat toward the younger end of, the age range typically assessed in child irony interpretation studies (e.g., Caillies et al., 2014; Filippova & Astington, 2008). We also assessed their structural language, nonverbal reasoning, general knowledge, theory of mind, inhibitory control, working memory, and cognitive flexibility. Because the children in Study 1 were essentially at floor in their interpretation of complex irony, in our follow-up study we assessed the interpretation of complex irony by 10-, 11-, and 12-year-olds. The majority of these children were in their first year of secondary school in England, which children usually start on the September following their 11th birthday. In secondary school, proficiency with irony starts to play a more important role in banter between peers. Our focus here was the relationship between complex irony and children's structural language, nonverbal reasoning, knowledge, advanced theory of mind, and cognitive flexibility. Our overarching hypothesis was that whereas knowledge, nonverbal reasoning, structural language, and theory of mind all would significantly correlate with irony interpretation, cognitive flexibility would contribute unique variance to the ability to interpret both simple and complex irony.

Study 1

Method

Participants

We tested 57 typically developing, monolingual English-speaking 6-, 7-, and 8-year-olds. All were recruited via and tested in the Kent Child Development Unit in southern England. Children whose parents said they suspected language or reading difficulties or any developmental difficulties were not invited to attend testing. Of the sample, 6 children were tested but excluded from all analyses because they failed the "attention check," which is described below. Thus, the final sample consisted of 51 children (22 boys) with a mean age of 90.44 months (SD = 9.68, range = 74–109). All were within the typically developing range on our measure of structure language, namely the Formulated Sentences subtest of the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5; Wiig, Semel, & Secord, 2013): M = scaled score of 12.33, range = scaled score of 8 to 19. Ethical approval was obtained from the School of Psychology ethics committee at the University of Kent. Children attended appointments in the Kent Child Development Unit with their parents. The vast majority of parents remained in the lab waiting room during testing, and the rest sat behind their children and were requested not to speak. Child assent was also always obtained, and a number of testing breaks were offered.

Irony task

Stimuli development. We first developed 42 vignettes ending in ironic criticism based on items from the previous literature as well as examples from children's books and films. We categorized these items as *simple irony* if the ironic intent was obvious from real-world evidence in the immediate context (i.e., what the addressee had just seen and/or heard). Conversely, we categorized these items as *complex irony* if the ironic intent could not be deduced in this manner but instead depended on accessing cultural common ground and/or information about the protagonists' relationship.

These items were then presented individually in a written format to 38 neurotypical, monolingual English-speaking adults who were naïve to the research question. The majority completed this for coursework credit as part of an undergraduate psychology program. For each item, participants were first asked to select the correct response (among three possible responses) and were then asked to rank, on a 9-point Likert scale, the degree to which they thought an 8-year-old would have difficulty in understanding the item, where 9 was the highest level of difficulty. All items for which the mean accuracy was 2.5 standard deviations below the mean for all irony items were removed from the level of difficulty ratings. Across all participants, only items for which a particular participant gave a correct response were included in our analysis of the level of difficulty ratings.

When selecting items for the child studies, there were a few items that, on consideration, were deemed too difficult to film. From the remaining items, we selected for the complex irony condition the five items that received the highest difficulty ratings within this category. For simple irony, we selected the five items that received the lowest difficulty ratings out of the simple irony items.

However, it can be seen from Appendices B and C that there was a great deal of overlap between the simple irony and complex irony categories regarding the difficulty ratings and some overlap in terms of mean accuracy.

Finally, we also created five *literal* interpretation items for our attention check that were equated with the irony items in complexity. An example of one literal item is given in Example (2) below. The complex irony, simple irony, and literal conditions were then matched in terms of the mean number of words and the number of mental state verbs in their vignettes. To achieve this, some of the original stories were slightly modified for the purpose of matching.

(2) Tom and Sally are wrapping presents.
Sally: What picture is on your wrapping paper?
Tom: It's got Father Christmas on it!
Q: What does Tom mean?
A. I don't really like wrapping presents at Christmas time.
B. A birthday present shouldn't be wrapped in Christmas wrapping paper.
C. I'm just telling you what picture is on the paper.

Development of video-recorded vignettes. The task itself consisted of five simple irony, five complex irony, and five literal (attention check) items. For each item, participants saw and heard a short video of a two- or three-line dialogue between a male speaker and a female speaker, ending with the target utterance. To ensure that the literal and ironic target utterances did not differ in terms of intonation, the target utterances were video-recorded before the actors knew about the prior context of the vignettes. Thus, the actors were blind to whether the target utterances were intended ironically or literally and were also instructed to maintain neutral intonation and emphasis. The simple irony vignettes all involved a conflict between the visual evidence and the literal meaning of the ironist's final statement (see Appendix B). The literal meaning of the ironist's statement in the complex irony vignettes was also untrue [see Example (3) below], but to determine this the listener needed the requisite cultural knowledge (here, an invitation to a party given by the Queen would be an extremely rare and thus highly unlikely occurrence, and so Sally is insinuating that this also holds true of Tom's claim).

(3) *Tom:* I have been invited to a party by the most beautiful girl in my class.
Sally: Yeah, and I have been invited to the Queen's party.
Q: What does she mean?
A. I think that the Queen's party would be more interesting.
B. I don't believe that you were invited to that girl's party.
C. I don't want to talk about that beautiful girl's party.

Task administration and coding

All participants were presented with the irony task via PsychoPy (Peirce, 2007). First, the experimenter explained the task to participants and administered two practice trials, for which feedback was given if required. These practice trials were always the two literal language trials shown in Appendix A and were similar in structure and sentence length to the test trials. The order of the simple irony, complex irony, and literal (attention check) items was fully randomized. For each item, participants first saw the video, which they could replay as many times as they wished.

For this study, we had two outcome measures for irony interpretation. The first involved open-ended questions following a number of studies in this field (e.g., Angeleri & Airenti, 2014; Caillies et al., 2014; Filippova & Astington 2008). Thus, the next screen for each item showed a visual display of the final part of the video interaction, and the experimenter asked an open-ended question ("Why did [the speaker] say [target statement]?"). Responses to the irony items all were audio-recorded, transcribed, and coded by the first author and another independent rater into one of the three following categories. Cohen's kappa was .81, reflecting a good level of agreement. Items for which raters differed were recoded by the first author following discussion.

0. Answers referring just to the literal meaning of the ironic utterance, repetitions of the context of the story, or saying "I don't know."
1. Responses reflecting simple surface-level justifications or a reference to (accurately described) states or feelings of the speaker ("Because she was mad") or reflecting learned conventional answers (e.g., "That is not a nice thing to do").
2. Responses with a reference made to the speaker's attitude toward the situation and to the pragmatic function of irony (e.g., "She did not want to be too harsh at her since she was just a little girl and she didn't mean to spill it all over the picture," "She wanted to make the accident a little less serious").

Following this, for each item participants were assessed on the second irony outcome measure (forced choice). Here, participants simultaneously heard (and saw in writing) the question "What does (s)he mean?" while seeing a "still" visual display of the interaction. Below this still clip, participants saw a picture of the speaker's face accompanied by three initially blank speech bubbles. Participants could click on each speech bubble in any order, upon which they would simultaneously hear (and read if at reading age) that particular response option. The simple irony vignettes can be seen in Appendix B together with their response options and the mean accuracy scores of the adult participants.

After listening to all possible answers, participants selected one by clicking on a "pick me" button, which was located on each of the speech bubbles. Children could also play each of the possible answers as many times as they needed by clicking on the speaker icon in each speech bubble. There were three different script orders so that for each vignette across the sample of participants the target answer could appear either first, second, or third.

Additional measures

The following tasks were always presented after the above task and were carried out in the same order as listed below. The testing was divided into three 30-min sessions carried out on the same day but with lengthy breaks for free play and snacks between each session. Additional breaks were added if children requested.

Inhibitory control. This was assessed via a computer-based version of the Stroop task. Importantly, children in England start school at 4 years of age, and thus all children in our sample could read the color words used. Computer keys were overlaid with colors, and participants were asked to respond by pressing the color that matched the color of the font (and ignore the meaning of the word, which described a different color). For each participant, we excluded data from error trials and outlier trials (i.e., all reaction times < 200 ms or 2.5 standard deviations above the mean). The difference between mean reaction times in the incongruent condition and the congruent condition was calculated for each participant. A higher score is taken to indicate more inference and thus poorer inhibitory control.

Working memory. This was assessed using a semi-computerized version of the Backward Digit Span subscale of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 2003). That is, the experimenter explained the instructions "live" and provided feedback on the practice items, but the actual stimuli for the test items were prerecorded and administered via button press. For this measure, higher scores indicate greater working memory capacity. The total possible *raw* score was 16, but because the ceiling (stopping rule) is one score of zero for an item, the actual highest raw score achieved in our sample was 12. (Details regarding standardized scores are given in Table 1.)

Structural language. This was assessed using the Formulated Sentences subtest of the CELF-5. This requires children to formulate a series of increasingly complex sentences on the basis of a picture and some keywords (e.g., *car, before*). Each item is scored on a 3-point scale in relation to semantics and morphosyntax. The total possible raw score was 48, but because the ceiling (stopping rule) was four consecutive scores of zero, the actual highest raw score achieved in our sample was 44. (Details regarding standardized scores are given in Table 1.)

Cognitive flexibility. This was assessed using a computerized version of the Wisconsin Card Sort Task, whereby participants were informed every time the rule changed (but not what the new rule was). The key measure was the raw number of perseverative errors recorded in PsychoPy. Thus, for this measure, higher scores indicate *poorer* cognitive flexibility.

General knowledge. This was assessed using the Information subtest of the WISC (Wechsler, 2003), which consists of 33 questions such as "What must you do to make water boil?".

Nonverbal reasoning. This was assessed using the Matrices subtest of the British Ability Scale–Third Edition (BAS-3; Elliot & Smith, 2011). For each item, children were shown an incomplete matrix or pattern of pictures or else abstract geometric shapes at the top of the page. For each item, children were asked to select the option (from a row at the bottom of the page) which best completed the matrix. This subtest is extremely similar to Matrices subtests on other intelligence tests standardized for children such as the Wechsler Preschool and Primary Sale of Intelligence–Fourth Edition (Wechsler, 2012). The total possible raw score was 51, but due to the application of ceiling (stopping rules), the actual highest raw score achieved in our sample was 26. (Details regarding standardized scores are given in Table 1.)

Theory of mind direct measure. This was assessed using four items from Happé's (1994) Strange Stories. The particular stories selected assessed understanding of lying to persuade ("Kittens"), white lies ("Christmas Present"), miscommunication ("Burglar's Glove"), and double bluff (Malkin, Abbot-Smith, Williams, & Ayling's [2018] "Biscuits" version). Vocabulary was simplified, and many more illustrations were used than in the original versions. All stories were presented on a laptop, and the narration was prerecorded and played via loudspeakers. The experimenter administered the test questions "live." All responses were audio-recorded, transcribed, and scored by the first author. The scoring system was derived from White, Hill, Happé, and Frith (2009). Two points were given if children referred to the character's inference about the mental state of the other person (e.g., "persuade," "things that [X] doesn't know," "trick," "thought that [X] knew," "sparing [X's] feelings"). One point was given if children referred to the outcome (e.g., "getting rid of the kittens," "to stop [X] eating biscuits") or trait (e.g., "she's nice," "she's lying") with no reference to protagonist's understanding of thoughts. Zero points were given if children referred to irrelevant or incorrect facts. An additional one half (0.5) point was given for passing the control questions for the "" "Kittens"' (white lie) and "Biscuits" (double bluff) vignettes. This meant that the total possible maximum score that children could attain for Strange Stories was 9. An experienced independent rater scored the performance of 11 participants blind to the ratings of the first author. Reliability was excellent; the average intraclass correlation coefficient (ICC) was .99.

Theory of mind indirect measure. Parents completed the Theory of Mind Inventory (ToMI; Hutchins, Prelock, & Bonazinga, 2012), which is a standardized questionnaire assessing the precursor, first-order, second-order, and advanced theory of mind. The original version of the ToMI did contain a few items specifically examining understanding of figurative language or hypotheticals (Items 2, 13, 29, and 36) rather than theory of mind. However, these four items were removed prior to analysis. For this reason, we used raw scores rather than standardized scores for this measure in our analyses. In the version we used, there were 38 items, each of which was scored on a scale from 0 to 20 (where 0 = *definitely not* and 20 = *definitely*). The overall mean was calculated. Thus, the maximum possible score was 20.

Results

The full anonymized datasets are available on the Open Sciences Framework web pages (https://osf.io/25k9m/).

Descriptive statistics for irony measures

For the forced-choice measure, children performed significantly better for simple irony (M = .76/1.00 correct, SD = .25) than for complex irony (M = .26/1.00 correct, SD = .23), $t(50) = 12.89$, $p < .001$. The same pattern was found for the open-ended question measures, $t(45) = 9.12$, $p < .001$, with participants performing better for simple irony (M = .48/1.00 correct, SD = .29), whereas for complex irony they were at floor (M = .10/1.00 correct, SD = .12). Because participants found the open-ended response measure difficult, only forced-choice responses were included in the main analyses.

Predictors

The nonverbal reasoning data were missing for 4 children. Parents of 2 additional children did not complete the ToMI questionnaire. Therefore, 45 children (19 boys) with a mean age of 89.8 months (range = 74–109) were included in the following analyses. For predictors based on standardized measures (nonverbal reasoning, structural language (CELF-5), general knowledge (WISC Information subtest), and Backward Digit Span (WISC), we used the standardized scores (i.e., T-scores or scaled scores). The descriptive statistics for the predictors of theoretical interest are shown in Table 1.

Table 1
Descriptive statistics for Study 1.

	M	*SD*	Min score	Max score	*N*
Nonverbal reasoning (BAS-3) *T* score	49.24	11.11	21	72	45
Structural language (Formulated Sentences scaled score)	12.42	2.50	8	19	45
Theory of mind direct (Strange Stories)	4.22	2.29	0	9	45
Theory of mind indirect (ToMI questionnaire)	16.80	1.62	13.20	19.70	45
General knowledge (WISC Information Test scaled score)	11.42	2.45	5	18	45
Working memory (Backward Digit Span scaled score)	11.69	2.54	8	18	45
Inhibitory control (Stroop RT incongruent − RT congruent)	190.04	141.22	−71.31	533.70	45
Cognitive flexibility (Card Sort) raw perseverative errors	10.29	7.06	1	33	45
Simple irony (forced choice)	3.71/5	1.27	0	5	45
Complex irony (forced choice)	1.27/5	1.16	0	5	45

Note. BAS-3, British Ability Scale–Third Edition; ToMI, Theory of Mind Inventory; WISC, Wechsler Intelligence Scale for Children; RT, reaction time.

Correlational analyses

Table 2 outlines which factors correlated with simple irony versus complex irony for Study 1. As can be seen, only parent-assessed theory of mind (ToMI), $r(45) = .40$, $p = .007$, and cognitive flexibility, $r(45) = -.44$, $p = .003$, correlated significantly with simple irony. The relationship with nonverbal reasoning was of marginal significance, $r(45) = .29$, $p = .058$.

Predictors of simple irony interpretation in 6- to 8-year-olds

We first converted all variables to z scores and then entered them into a hierarchical linear regression. Scaled scores for nonverbal reasoning (Matrices subtest of BAS-3) and structural language (CELF Formulated Sentences subtest) were entered into the first step. This led to a model that approached significance, $F(2, 42) = 2.59$, $p = .087$, and accounted for 11% of the variance. Then, theory of mind, executive functioning, and general knowledge predictors all were entered into the second step. This led to a significant model, $F(8, 36) = 3.35$, $p = .006$, accounting for 43% of variance overall and showing an R^2 change of .32. As can be seen from Table 3, in this second step only Card Sort (cognitive flexibility) and the ToMI (indirect measure of theory of mind) were significant predictors, accounting for 12% and 9% of variance, respectively. The ToMI showed a positive relationship with simple irony interpretation (as expected given that positive scores on both measures indicate better performance). The Card Sort task showed a negative relationship with simple irony interpretation (again as expected given that higher scores on the Card Sort task indicate poorer cognitive flexibility). Multicollinearity was not a concern (tolerance for all > .60).

To verify our results, we reran the same analysis, this time including only those predictors that were significant or of marginal significance in the first analysis, namely nonverbal reasoning, cognitive

Table 2
Correlations with irony for Study 1.

	Simple Irony	Complex Irony
NV Reasoning (BAS Matrices T-score)	.29 (p = .058)	.23 (p = .14)
Structural Language (Formulated Sentences scaled score)	.21 (p = .17)	.05 (p = .77)
Theory of Mind Direct (Strange Stories)	.23 (p = .12)	.27 (p = .07)
ToM Indirect (ToMI questionnaire with items on figurative language removed))	.40 (p = .007)	.22 (p = .15)
General Knowledge (WISC Information Test scaled score)	.03 (p = .87)	-.03 (p = .83)
Working Memory (WISC Backwards Digit Span scaled score)	-.099 (p = .52)	-.18 (p = .24)
Inhibitory Control (Stroop Incongruent RT – Congruent RT)	.27 (p = .07)	-.19 (p = .21)
Cognitive Flexibility (Perseverative Errors)	-.44 (p = .003)	-10 (p = .51)

Note. BAS-3, British Ability Scale–Third Edition; ToMI, Theory of Mind Inventory; WISC, Wechsler Intelligence Scale for Children; RT, reaction time.

flexibility (Card Sort), and the indirect theory of mind measure (ToMI). The same pattern of results was found, whereby Card Sort (β = −.37, t = −2.90, p = .006) and the ToMI (β = .35, t = 2.78, p = .008) still accounted for unique variance (13% and 12%, respectively).[2] Regarding potential multicollinearity, here the values for tolerance were excellent (all > .94).

Predictors of complex irony interpretation in 6- to 8-year-olds

When the same hierarchical linear regression was carried out for complex irony, the model at the first step was not significant, $F(2, 44) = 1.12$, $p = .34$. This did not improve ($p = .20$) at the second step once the predictor variables were entered. Indeed, the only predictor to even marginally correlate with complex irony was direct theory of mind (Strange Stories), $r(47) = .28$, $p = .06$.

[2] When only cognitive flexibility (Card Sort) and the indirect theory of mind measure (ToMI) were included, these accounted for 17% ($p = .002$) and 14% ($p = .006$) of the unique variance, respectively.

Discussion

For the interpretation of simple irony by children of this age, cognitive flexibility accounted for unique variance when nonverbal reasoning, structural language, and theory of mind were entered into the same analysis, thereby supporting our main hypothesis. However, no factors were significantly related to complex irony interpretation by children in this age group. The most likely interpretation for this latter finding is that this age group did not comprehend complex irony at all; among a choice of three items per trial, children selected the correct option only 26% of the time on average.

Study 2

To examine the correlates of complex irony, we carried out a follow-up study (Study 2) with 10- to 12-year-olds. We assessed only the interpretation of complex irony in this second study because we assumed that this age group would be at ceiling on simple irony (see, e.g., Demorest, Meyer, Phelps, Gardner, & Winner, 1984; Pexman & Glenwright, 2007). We again controlled for nonverbal reasoning and structural language, selecting the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2011) Matrices and Vocabulary measures because they are standardized for this age group. Because it is much easier to test the target age group in schools, we needed to reduce the total amount of testing time per child and avoid asking parents to complete a lengthy questionnaire. Therefore, we only directly assessed theory of mind, and regarding executive functioning we assessed only cognitive flexibility.

Method

Participants

We tested 97 typically developing, monolingual English-speaking 10-, 11-, and 12-year-olds. Of these children, 20% were recruited via and tested in the Kent Child Development Unit, as for Study 1. The remainder were tested in a quiet area of a school in the same region. Of the sample, 9 children were tested but excluded from all analyses because they failed the "attention check." An additional 2 children were removed from the sample because they had a diagnosed neurodevelopmental disorder, and 1 child was removed because scores were anomalous (4 standard deviations away from the mean). When rescoring offline our measure of structural language—namely the Vocabulary subtest of the WASI (Wechsler, 2011)—it was ascertained that 9 children in the remaining sample did not reach ceiling (prespecified stopping point) on this measure due to experimenter error. There were missing data for the Card Sort for 1 child, for the knowledge test for 1 child, and for theory of mind for 2 children. Thus, the final sample consisted of 72 children (35 boys) with a mean age of 143 months (SD = 6.02, range = 126–156). Ethical approval was obtained from the School of Psychology ethics committee at the University of Kent. In addition to parental and teacher consent, child assent was always obtained and children were told that they could take a break whenever they wished or desist testing entirely if they preferred.

Irony and literal interpretation task

We used only the forced-choice method for complex irony interpretation because children in Study 1 found the open-ended response measure to be more difficult than forced choice. (In addition, although interrater reliability for the open-ended responses was good, it was not excellent.) For the forced-choice measure in Study 2, we extended the scale to include seven items (shown in Appendix C).

Additional measures

WASI Vocabulary subtest. In this test, participants were asked to define word meanings. For each item, participants' response could receive a score of either two points, one point, or zero points. For example, for the item "tradition," participants would receive zero points if they referred to either something one celebrates or a legend. Participants would receive one point if they referred to a family pattern,

Table 3
Predictors of simple irony.

Step	Variable	*B*	*SE*	*t*	sr^2
1	Nonverbal reasoning (BAS-3) *T* score	.26	.15	1.78$^{\forall}$	.07
	Structural language (Formulated Sentences scaled score)	.17	.15	1.16	.03
2	Nonverbal reasoning (BAS-3) *T* score	.27	.15	1.88$^{\forall}$	.06
	Structural language (Formulated Sentences scaled score)	.08	.17	0.50	.00
	Theory of mind direct (Strange Stories)	−.03	.16	0.20	.00
	Theory of mind indirect (ToMI)	.33	.14	2.38**	.09
	General knowledge (WISC Information Test)	−.24	.16	1.49	.03
	Working memory (Backward Digit Span)	−.18	.14	1.34	.03
	Inhibitory control (Stroop RT incongruent − RT congruent)	.07	.15	0.49	.00
	Cognitive flexibility (Card Sort)	−.41	.15	2.70**	.12

Note. sr^2, squared part correlation; BAS-3, British Ability Scale–Third Edition; ToMI, Theory of Mind Inventory; WISC, Wechsler Intelligence Scale for Children; RT, reaction time.
$^{\forall}$ $p < .10$.
** $p < .01$.

ritual, or something that has been passed on for generations. To achieve a score of two points, participants needed to be more specific (e.g., something one does/a ritual/a theme that is passed on/repeated from generation to generation or else a set of values/behaviors rooted in the history/past). Participant responses to this test were audio-recorded and transcribed offline. The ceiling (stopping level) was three consecutive scores of zero. There were 31 items, so the maximum possible raw score was 62. However, the highest score obtained in our sample was a raw score of 37.

Cognitive flexibility. We used the same computer-based Wisconsin Card Sort Task as in Study 1 following the same procedure.

WASI Matrices (nonverbal reasoning). Similarly to the BAS Matrices used in Study 1, in the WASI Matrices used in Study 2, participants saw for each item a matrix of four geometric shapes, where the fourth item was missing. Participants needed to choose one of five possibilities to complete the matrix. The ceiling (stopping level) was three consecutive scores of zero. The maximum possible raw score was 30, but the highest score obtained in our sample was a raw score of 23.

Theory of mind. We were concerned that typically developing young adolescents might score toward ceiling on Strange Stories. Therefore, we selected the widely used Animations task (Abell, Happé, & Frith, 2000) as an age-appropriate theory of mind measure. In this task, participants were asked to narrate the actions of two inanimate shapes, specifically four animations, which tend to elicit use of mental state terms. Participants were first shown a practice clip, which they saw three times. On the first viewing, participants watched the clip passively. On the second viewing, participants narrated the clip. On the third viewing, the experimenter said "This is what I think is happening" and narrated the clip. For each of the four test items, participants viewed the clips twice, the first time without attempting narration. No feedback (third viewing) was given for the test items.

The first author scored narrations of each animation on a 3-point scale in accordance with the criteria outlined in Abell et al. (2000). For example, one clip showed a large triangle coaxing a small triangle out of a box/house. If children merely described "pushing" or "fighting," then this item would receive a score of zero points. If children described reluctance to exit on the part of the small triangle, then this item would receive a score of one point. If children described this item in terms of the big triangle trying to urge (or something similar) the small triangle out, then this item would receive the full score of two points. An independent rater scored the narrations produced by 10 participants (13% of the data) with excellent reliability (ICC = .92).

Assessment of knowledge

At the end of the session, participants were asked seven binary-choice questions that tested their background knowledge relevant for the ironic videos presented at the beginning of the session (e.g., "Do most people get invited to the Queen's parties?"). For two items, a score of one half point was also given for partially correct responses in which children provided an explanation as to why they thought it was not a problem to be late for the cinema or why brothers might like helping their sisters to wash the dishes (e.g., "There are 30-minute [advertisements] at the beginning anyway").

Procedure

The tasks were always administered in the following order: irony interpretation, WASI Vocabulary, Card Sort task, WASI Matrices (nonverbal reasoning), theory of mind, and the knowledge test. The entire testing session lasted 45 min. Participants were offered a break halfway through, but most opted not to take it.

Results

The full anonymized datasets are available on the Open Sciences Framework web pages (https://osf.io/25k9m/). The overall performance for complex irony in Study 2 was significantly above chance, $t(71) = 5.85$, $p < .001$. The descriptives for each variable are reported in Table 4.

Correlational analyses

Complex irony interpretation correlated with all the predictor variables except age: nonverbal reasoning *T* score, $r(72) = .39$, $p = .001$; Vocabulary *T* score, $r(72) = .39$, $p = .001$; total specific knowledge score, $r(72) = .24$, $p = .043$; and raw number of perseverative errors in the Card Sort, $r(72) = -.37$, $p = .001$. It correlated marginally with the theory of mind Animations task, $r(72) = .23$, $p = .054$.[3]

Main analysis

We converted all variables to *z* scores and then entered them into a hierarchical linear regression. In our regression analysis, *T* scores for nonverbal reasoning (Matrices) and structural language (Vocabulary) were entered into the first step. This led to a model that was significant ($p < .001$) and accounted for 23% of the variance. Then, theory of mind, cognitive flexibility, and specific knowledge predictors all were entered into the second step. This led to a significant model, $F(5, 66) = 4.700$, $p < .001$, accounting for 33% of variance overall and showing an R^2 change of .10. As can be seen from Table 5, in this second step only cognitive flexibility and nonverbal reasoning were significant. Multicollinearity was not a concern (tolerance for all > .80).

To verify our results, we reran the same analysis, this time including only those predictors that were significant in the first analysis, namely nonverbal reasoning, structural language, and cognitive flexibility (Card Sort). The same pattern of results was found in this as in the previous analysis; here, Card Sort ($\beta = -.27$, $t = 2.60$, $p = .011$) accounted for a slightly greater proportion of unique variance (7%), whereas both nonverbal reasoning ($\beta = .24$, $t = 2.20$, $p = .031$) and Vocabulary ($\beta = .24$, $t = 2.27$, $p = .026$) accounted for 5% of unique variance. The values for tolerance were similar to the above values (all > .83).

Knowledge interpretation

For those knowledge items that were answered incorrectly, 62.5% of the equivalent irony interpretation items were also answered incorrectly. Conversely, for those knowledge items that were answered correctly, 57.2% of the equivalent irony interpretation items were also answered correctly. Thus, although the relationship between knowledge and irony analyzed in this manner was significant ($\chi^2 = 6.215$, $p = .045$)—and also approached significance in the regression analysis (see Table 5)—requisite background knowledge is clearly insufficient for the correct interpretation of complex irony.

[3] The theory of mind Animations task showed a negative correlation with cognitive flexibility, $r(72) = -.37$, $p = .002$.

General discussion

In two studies, we investigated whether the previously reported relationship between theory of mind and verbal irony interpretation in children might in fact be better accounted for in terms of executive functioning, specifically cognitive flexibility. Unlike the vast majority of previous studies, we did not solely examine children's interpretation of simple irony, where the concurrent physical evidence from the real world reveals that the speaker cannot intend the utterance literally. Rather, we also assessed children's understanding of more complex irony for which the listener must access cultural or general knowledge to determine that the literal meaning is not intended. Both studies statistically controlled for nonverbal reasoning, structural language, and knowledge understanding. Study 1 found that both cognitive flexibility and (parent-assessed) theory of mind accounted for unique variance in how 6- to 8-year-olds interpreted simple irony (whereas inhibitory control and working memory did not). In Study 2, theory of mind showed a marginally significant correlation (with a small to medium effect size) with complex irony interpretation by 10- to 12-year-olds. However, theory of mind did not contribute unique variance to the regression analysis, whereas cognitive flexibility did.

The role of knowledge and nonverbal reasoning

General knowledge was not found to correlate with simple irony interpretation in Study 1, but there was a correlation (with a small to medium effect size) in Study 2 between complex irony interpretation and the specific knowledge required for the vignettes we used. In the regression analysis, this relationship was only of marginal significance. This—and our chi-square analysis—suggests that although the requisite knowledge is accessed in order to interpret irony, this does not play as important a role as cognitive flexibility or even as important a role as nonverbal reasoning ability.

Indeed, the current studies are the first in this field to statistically control for the role played by nonverbal reasoning. Our findings are an important contribution on this point because in Study 2 we found that nonverbal reasoning was a significant contributor to unique variance in irony interpretation (and showed a marginally significant relationship in Study 1). It is difficult to know exactly how to interpret the role of nonverbal reasoning in irony interpretation. It could merely play a relatively superficial role; children with better reasoning ability might be better able to determine the correct answer from the selection of three choices simply by induction. A non-mutually exclusive possibility is that, on the whole, neurotypical children who attain higher nonverbal reasoning scores tend to also perform better on both executive functioning and structural language measures and indeed on all developmental measures. Thus, nonverbal reasoning might have mediated the reported relationships between irony interpretation and theory of mind in previous studies.

The role of theory of mind

Regarding theory of mind, we found in Study 2 that the relationship between the Animations theory of mind task (Abell et al., 2000) and complex irony had a small to medium effect size (and was marginally significant). The relationship between theory of mind, as assessed by Strange Stories (e.g., White et al., 2009) and irony, had a similar effect size in Study 1. However, these relationships between irony interpretation and directly assessed theory of mind were not retained in regression analyses in either study. In contrast, in Study 1 we found a medium to strong effect size for the relationships between the parental questionnaire theory of mind measure (ToMI) and simple irony interpretation, which was retained in our regression analyses.

These somewhat inconclusive results regarding the relationship between theory of mind and child irony interpretation led us to reexamine the previous literature in a more fine-grained manner. We noted that whereas two studies had found large effect sizes for the relationship between theory of mind and child irony interpretation (Caillies et al., 2014; Filippova & Astington, 2008), others had found only medium effect sizes (Angeleri & Airenti, 2014; Banasik, 2013) or small effect sizes (Mewhort-Buist & Nilsen, 2013; Nilsen et al., 2011). Moreover, of those studies that had found medium or large effect sizes, all except one study were difficult to interpret because they either had not

Table 4
Descriptive statistics for Study 2.

	M	*SD*	Min score	Max score	*N*
Nonverbal reasoning (WASI Matrices) *T* score	46.28	7.29	25	62	72
Structural language (WASI Vocabulary) *T* score	45.51	6.62	29	61	72
Theory of mind Animations (maximum possible = 8)	5.10	1.94	0	8	72
Specific knowledge raw total (maximum possible = 7)	6.35	0.60	4.5	7	72
Cognitive flexibility (Card Sort) raw perseverative errors	6.40	3.76	1	16	72
Complex irony (proportion correct)	.55	.31	0	1	72

Note. WASI, Wechsler Abbreviated Scale of Intelligence.

Table 5
Study 2 predictors of complex irony.

Step	Variable	*B*	*SE*	*t*	sr^2
1	Nonverbal reasoning (WASI Matrices) *T* score	.30	.12	2.72**	.08
	Structural language (WASI Vocabulary) *T* score	.30	.11	2.67**	.08
2	Nonverbal reasoning (WASI Matrices) *T* score	.24	.11	2.19*	.05
	Structural language (WASI Vocabulary) *T* score	.21	.11	1.92∀	.04
	Theory of mind direct (Animations)	.11	.12	1.03	.01
	Specific knowledge	.18	.11	1.68∀	.03
	Cognitive flexibility (Card Sort)	−.23	.18	2.10*	.04

Note. sr^2, squared part correlation; WASI, Wechsler Abbreviated Scale of Intelligence.
∀ $p < .10$.
* $p < .05$.
** $p < .01$.

statistically controlled for structural language (Banasik, 2013; Caillies et al., 2014) or had conflated irony with nonironic joke understanding (Angeleri & Airenti, 2014).

The results of the remaining studies were also less clear than at first sight. Filippova and Astington (2008) reported that the relationship with theory of mind was retained in their regression analyses, but in fact this relationship was only marginally significant. Nilsen et al. (2011) only found a significant relationship between theory of mind and whether children interpreted irony as mean versus nice but not whether children correctly interpreted the communicative intent of the ironist. In addition, the particular component of theory of mind that was found to relate to irony varied between studies. Worse, in some studies theory of mind showed larger effect sizes for relationships with literal language interpretation than for relationships with irony interpretation (e.g., Mewhort-Buist & Nilsen, 2013).

Despite the aforementioned lack of clarity regarding the empirical findings, we argue that listeners logically must access some form of shared mental representations to interpret irony—particularly more complex irony. For example, if Tom says "I was invited to a party by the most beautiful girl in my class" and Sally replies "Yeah, and I was invited to the Queen's party," it does not suffice to know that the Queen is unlikely to have invited Sally to a party. Listeners must also take into account that Sally knows that Tom knows the Queen has probably not invited her, otherwise Sally's utterance would merely be interpreted as a lie. Other examples of complex irony may require listeners to access the common ground understanding that they share with a *particular* speaker. For example, if a speaker asks someone to wash the speaker's plate and the addressee responds by asking whether he or she should tidy the speaker's room as well, the first speaker must consider his or her personal relationship with the addressee (e.g., a sibling vs. an overly submissive romantic partner) and other circumstances (e.g., perhaps the addressee knows that the speaker's arms are injured). Therefore, we conclude that the relationship between theory of mind and child irony interpretation warrants much closer examination in future research.

The role of cognitive flexibility

We found that cognitive flexibility emerged as a consistent predictor of verbal irony interpretation across our two studies. Moreover, cognitive flexibility accounted for unique variance when controlling for structural language, nonverbal reasoning, theory of mind, and knowledge. One possible explanation for why cognitive flexibility correlates with irony interpretation could simply be that cognitive flexibility requires both working memory and inhibitory control, and both of these are logically implicated in irony interpretation. This particular account does not receive empirical support from the current study; rather, Study 1 found no relationships between irony interpretation and either working memory (as assessed by a standardized test, the WISC Backward Digit Span) or inhibitory control (as assessed by the commonly used Stroop task).

An alternative explanation for the current finding that cognitive flexibility correlates with irony interpretation is that this is due to the "switching" element of cognitive flexibility. Switching is required in two aspects of irony interpretation. First, dynamic theory of mind itself requires the ability to switch between one's own perspective and the interlocutor's perspective. In addition, when processing an ironic remark, the listener needs to consider more than one possible interpretation of the speaker's communicative intent. We propose that this (rapid and often unconscious) consideration of multiple intended meanings is highly likely to involve an element of switching.

This suggestion that irony interpretation always involves the activation of multiple potential meanings does not, of course, fit with all theories of irony, particularly not with Gibbs's (2002) direct access account, in which it is proposed that the ironic meaning can in some contexts be activated directly without first activating the literal meaning. Importantly, recent studies of adults using timecourse-sensitive techniques, such as eye tracking and event-related potentials, throw clearer light on this issue; they show that although highly familiar irony, for example, may initially be processed more rapidly, it loses this advantage in later processing stages and becomes more difficult to process than literal items (Turcan & Filik, 2016). Indeed, overall this body of literature indicates that although irony familiarity and context play an important role, irony or sarcasm incurs a higher processing load overall than literal language (e.g., Filik, Leuthold, Wallington, & Page, 2014; Olkoniemi, Strömberg, & Kaakinen, 2019).

Moreover, we emphasize that for more complex irony, the requisite switching is not only between a literal meaning and an ironic meaning but also between various possible interpretations of the ironic content. For example, when processing "Sure, I've just been sitting round all day at school" in response to "Can you help me cook because I am tired," the listener very likely needs to switch between a consideration of whether this response indicates a refusal to help and one of whether the response is merely an expression of annoyance.

Limitations

The current research should be considered with three key limitations in mind. First, our results (and indeed those in the previous literature) regarding the relationship between theory of mind are not conclusive in either direction. Future developmental studies should include a number of theory of mind—and emotion recognition—measures and should control for both nonverbal reasoning and structural language. In an ideal world, researchers would have access to scalable advanced measures of theory of mind, which are not so language heavy. Future studies should also include a sample size suitable for medium effect sizes in multiple regression.

Second, although we did not find any relationships in Study 1 between simple irony interpretation and either inhibitory control or working memory, this null result is based on one study—and using only one measure for each construct. Therefore, it would be well worth replicating using multiple working memory and inhibitory control measures and also examining any potential relationships with complex irony.

Finally, we only investigated relationships between these variables—and only at one time point during development at that. A more robust test would be to train children's cognitive flexibility (see, e.g., Kenworthy et al., 2014, for a child EF training program that appears to affect social communication). If training cognitive flexibility led to improved irony interpretation, this would allow us to

infer that irony interpretation proficiency depends—at least in part—on cognitive flexibility. This could be an important step forward for children with atypical development who struggle with irony interpretation—particularly because theory of mind training does not seem to generalize to social communication (e.g., Begeer et al., 2015; Hadwin, Baron-Cohen, Howlin, & Hill, 1996; see also Pexman, Reggin & Lee, 2019, for a general discussion of irony training in children).

Conclusions

The cognitive correlates of complex irony interpretation in young adolescents (structural language, nonverbal reasoning, and cognitive flexibility) appear to be very similar to those of simple irony interpretation in primary school-aged children. Importantly, cognitive flexibility retains significance as a predictor of irony interpretation ability even when controlling for requisite knowledge, structural language, and nonverbal reasoning. Given our weaker and less consistent results for relationships between theory of mind and irony interpretation, we suggest that the role of executive functioning—particularly cognitive flexibility—may be a more fruitful line of research for those interested in how children learn to interpret irony.

CRediT authorship contribution statement

Maria Zajączkowska: Conceptualization, Methodology, Formal analysis, Writing - original draft, Investigation, Project administration, Data curation. **Kirsten Abbot-Smith:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Formal analysis, Supervision.

Acknowledgments

This study was funded by a Ph.D. scholarship to the first author. Many thanks go to Frank Gasking for technical assistance. Many thanks also go to Toby Nicholson for training on the Animations task. A special thank you goes to the children, the parents who brought their children into the Kent Child Development Unit, and the staff and caregivers of local secondary schools in Kent, United Kingdom. Many thanks also go to Ellis Cross, Megan Day, Adela Jurkova, Ishka Trace-Kleeberg, Jessica Rigden, and Gemma Wilby for assistance with testing and to Ellis Cross and Dr Louise Malkin for coding reliabilities.

Appendix A

Practice vignettes

1. *Tom:* Where are you going?
   Sally: I am going to the toilet.
   A. I want to tell you that I'm going to the shop.
   B. I want to tell that I'm going home.
   C. I want to tell you that I'm going to the toilet.
2. *Sally:* Where did you buy your shoes?
   Tom: In the shoe shop around the corner.
   A. My sister bought me this new pair of shoes.
   B. This is not my new pair of shoes.
   C. I bought them in the shop around the corner.

Appendix B

Simple irony vignettes with adult response accuracy and difficulty ratings

Vignette	Response options	Mean adult accuracy (%)	Mean adult difficulty rating (1 = easy; 9 = difficult)
1. *Boy:* Can I have a sweet (candy)?[a] *Girl:* [hands him an empty wrapper] *Boy:* Thanks a lot!	A. It's kind of her to give me the wrapper. B. It's selfish of her to give me the wrapper. C. She has helped me not to get fat.	97	4.46
2. Tom knocked over a glass and spilled the juice over the clean tablecloth. *Sally:* Well done!	A. I am glad that you didn't break the glass. B. I am cross that you knocked over the juice. C. The tablecloth looks better with juice stains.	100	4.71
3. Sally wants to help Tom with math, but he does not follow her instructions. *Sally:* You are a great listener.	A. You didn't interrupt me when I was giving the instructions. B. You listened to me, and you are great at following the instructions. C. You weren't listening to me when I was giving the instructions.	95	4.97
4. Sally and Tom are eating crisps (chips) in the park.[b] One crisp falls on the ground. Sally picks it up and eats it. *Tom:* Yummy!	A. I wish I could eat a crisp, too. B. I am disgusted that you ate a dirty snack. C. I am very worried about your health.	97	4.97
5. Sally and Tom want to go for a picnic.[c] It has just started to rain. *Tom:* It's a perfect day for picnic.	A. I like it when it's sunny when we go for a picnic. B. I don't like when it's hot and sunny in the park. C. I am upset that we can't go out because it's raining.	87	5.03

[a]This vignette is adapted from an item in Bosco and Bucciarelli (2008).
[b]This vignette is loosely based on an item from Bucciarelli, Colle, and Bara (2003).
[c]This picnic vignette is adapted from Happé (1994).

Appendix C

Complex irony vignettes with adult response accuracy and difficulty ratings

Vignette	Response options	Mean adult accuracy (%)	Mean adult difficulty rating (1 = easy; 9 = difficult)
1. *Tom:* I have been invited to a party by the most beautiful girl in my class. *Sally:* Yeah, and I have been invited to the queen's party.	A. I think that the queen's party would be more interesting. B. I don't believe that you were invited to that girl's party. C. I don't want to talk about that beautiful girl's party.	82	6.03
2. Matt and Emma are going to the cinema. They are late because Emma is getting ready very slowly. *Matt:* Don't worry, Emma. Take your time.	A. I am a bit annoyed that you are not ready yet. B. It's fine that you're slow—we have got plenty of time. C. I am happy that the film starts so early.	92	5.11
3. *Sally:* Could you wash my plate? *Tom:* Do you want me to tidy your room, too?	A. I definitely will not help you with washing up. B. I like taking care of my sister. C. I feel sorry for you, so I will help you.	92	6.06
4. A: Would you like me to hold the umbrella over you? B: No, I really like getting wet.	A. I am upset that you brought the umbrella with you. B. No, I don't want the umbrella because I want to get wet. C. Of course I want the umbrella, so I think your question is stupid.	87	5.06
5. *Sally:* Do you think I should package up the phone before posting (mailing) it? *Tom:* No, just put a stamp on it and pop it into the post (mailbox).	A. Of course you need to package it up before posting. B. This phone does not need to be packed before posting. C. I am surprised that you want to package the phone.	84	5.97
6. *Tom:* Can you help me cook the dinner? I'm tired. *Sally:* Oh yes, because I have just been sitting around doing nothing at school today	A. I'm cross because you do not understand what a bad day I have had. B. I'm actually saying "no" because I am too tired to cook with you.	84	5.75

(*continued on next page*)

Appendix C (*continued*)

Vignette	Response options	Mean adult accuracy (%)	Mean adult difficulty rating (1 = easy; 9 = difficult)
	C. I will help you to make the dinner because I'm really bored.		
7. Matt and Emma are in the bathroom. *Matt:* Before you brush your teeth, you have to put toothpaste on your toothbrush. *Emma:* No way! Really?	A. I think that everyone knows what you are telling me about brushing teeth. B. I didn't know that I should put the toothpaste on the toothbrush first. C. I am surprised that you know how to brush your teeth.	84	4.78

References

Abbot-Smith, K. (2020). Language disorders and autism: Implications for usage-based theories of language development. In C. Rowland, B. Ambridge, A. Theakston, & K. Twomey (Eds.), *Current perspectives on child language acquisition: How children use their environment to learn (Trends in Language Acquisition Research)*. Amsterdam: John Benjamins.

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*, 1–16.

Aijmer, K. (2019). "Ooh whoops I'm sorry!" Teenagers' use of English apology expressions. *Journal of Pragmatics, 142*, 258–269.

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*, 133–146.

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15*, 31–36.

Arffa, S. (2007). The relationship of intelligence to executive function and non-executive function measures in a sample of average, above average, and gifted youth. *Archives of Clinical Neuropsychology, 22*, 269–278.

Banasik, N. (2013). Non-literal speech comprehension in preschool children—An example from a study on verbal irony. *Psychology of Language and Communication, 17*, 309–324.

Banasik-Jemielniak, N., & Bokus, B. (2019). Children's comprehension of irony: Studies on Polish-speaking preschoolers. *Journal of Psycholinguist Research, 48*, 1217–1240.

Begeer, S., Howlin, P., Hoddenbach, E., Clauser, C., Lindauer, R., Clifford, P., ... Koot, H. (2015). Effects and moderators of a short theory of mind intervention for children with autism spectrum disorders: A randomized controlled trial. *Autism Research, 8*, 738–748.

Bosco, F. M., & Bucciarelli, M. (2008). Simple and complex deceits and ironies. *Journal of Pragmatics, 40*, 583–607.

Bucciarelli, M., Colle, L., & Bara, B. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics, 35*, 207–241.

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.

Chevallier, N., Sheffield, T., Nelson, J., Clark, C., Wiebe, S., & Espy, K. (2012). Underpinnings of the cost of flexibility in preschool children: The roles of inhibition and working memory. *Developmental Neuropsychology, 37*, 99–118.

Demorest, A., Meyer, C., Phelps, E., Gardner, H., & Winner, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development, 55*, 1527–1534.

Diamond, A. (2013). Executive functions. *Annual Reviews of. Psychology, 64*, 135–168.

Elliot, C. D., & Smith, P. (2011). *British Ability Scales-Third Edition (BAS 3)*. London: GL Assessment.

Filik, R., Leuthold, H., Wallington, K., & Page, J. (2014). Testing theories of irony processing using eyetracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 811–828.

Filippova, E. (2014). Irony production and comprehension. In D. Matthews (Ed.), *Pragmatic development in first language acquisition* (pp. 261–278). Amsterdam: John Benjamins.

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*, 126–138.

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

Godbee, K., & Porter, M. (2013). Comprehension of sarcasm, metaphor and simile in Williams syndrome. *International Journal of Language and Communication Disorders, 48*, 651–665.

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? *Development and Psychopathology, 8*, 345–365.

Hancock, J. T., Dunham, P. J., & Purdy, K. (2010). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development, 1*(2), 37–41.
Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*, 101–119.
Happé, F. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129–154.
Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 42*, 327–341.
Jacques, S., & Zelazo, P. D. (2005). On the possible roots of cognitive flexibility. In B. D. Homer & C. S. Tamis-LeMonda (Eds.), *The development of social cognition and communication* (pp. 53–81). Mahwah, NJ: Lawrence Erlbaum.
Kenworthy, L., Anthony, L. G., Naiman, D. Q., Cannon, L., Wills, M. C., Luong-Tran, C., ... Wallace, G. L. (2014). Randomized controlled effectiveness trial of executive function intervention for children on the autism spectrum. *Journal of Child Psychology and Psychiatry, 55*, 374–383.
Kidd, E. (2013). The role of verbal working memory in children's sentence comprehension. *Topics in Language Disorders, 33*, 208–223.
Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development, 74*, 1823–1839.
Leggitt, J. S., & Gibbs, R. W. (2000). Emotional reactions to verbal irony. *Discourse Processes, 29*, 1–24.
Malkin, L., Abbot-Smith, K., Williams, D., & Ayling, J. (2018). When do children with autism spectrum disorder take common ground into account during communication? *Autism Research, 11*, 1366–1375.
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*, 1–13.
Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: A review of associations with formal language, social cognition, and executive functions. *Language Learning and Development, 14*, 186–223.
Mewhort-Buist, T. A., & Nilsen, E. S. (2013). What are you really saying? Associations between shyness and verbal irony comprehension. *Infant and Child Development, 22*, 180–197.
Mewhort-Buist, T. A., & Nilsen, E. S. (2019). Shy children's understanding of irony: Better comprehension does not always mean better socio-emotional functioning. *Infant and Child Development, 28* e2131.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Nilsen, E. S., & Bacso, S. A. (2017). Cognitive and behavioural predictors of adolescents' communicative perspective-taking and social relationships. *Journal of Adolescence, 56*, 52–63.
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.
Olkoniemi, H., Strömberg, V., & Kaakinen, J. K. (2019). The ability to recognise emotions predicts the time-course of sarcasm processing: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 72*, 1212–1223.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13.
Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathologies. *Journal of Child Psychology and Psychiatry, 37*, 51–87.
Pexman, P. M., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony?. *Journal of Neurolinguistics, 20*, 178–196.
Pexman, P., Reggin, L., & Lee, K. (2019). Addressing the challenge of verbal irony: Getting serious about sarcasm training. *Languages, 4*, 23.
Raven, J. C. (1986). *Coloured Progressive Matrices*. London: H. K. Lewis.
Ţurcan, A., & Filik, R. (2016). An eye-tracking investigation of written sarcasm comprehension: The roles of familiarity and context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1867–1893.
Wechsler, D. (2003). *The Wechsler Intelligence Scale for Children* ((4th ed.).). San Antonio: TX: Psychological Corporation.
Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II)*. San Antonio, TX: NCS Pearson.
Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition: Technical manual and interpretive manual*. San Antonio, TX: Psychological Corporation.
Wellman, H. M. (2014). *Making minds: How theory of mind develops*. New York: Oxford University Press.
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.
White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing mentalizing impairments in autism. *Child Development, 80*, 1097–1117.
Wiig, E. H., Semel, E., & Secord, W. (2013). *Clinical Evaluation of Language Fundamentals-Fifth Edition (CELF-5)*. San Antonio, TX: NCS Pearson.
Wilson, D. (2013). Irony comprehension: A developmental perspective. *Journal of Pragmatics, 59*, 40–56.
Zajączkowska, M., Abbot-Smith, K., & Kim, C. (2020). Using shared knowledge to determine ironic intent: A conversational response paradigm. *Journal of Child Language. Advance online publication*. https://doi.org/10.1017/S0305000920000045.
Zelazo, P. D., & Müller, U. (2007). Executive function in typical and atypical development. In *Blackwell handbook of childhood cognitive development* (pp. 445–469). New York: John Wiley.