How do 3-year-olds use relevance inferencing to interpret indirect speech?

Kirsten Abbot-Smith (iD)
University of Kent, UK

Cornelia Schulze (iD)
Leipzig University, Germany

Nefeli Anagnostopoulou
Maria Zajączkowska
University of Kent, UK

Danielle Matthews
The University of Sheffield, UK

Abstract

If a child asks a friend to play football and the friend replies, 'I have a cough', the requesting child must make a 'relevance inference' to determine the communicative intent. Relevance inferencing is a key component of pragmatics, that is, the ability to integrate social context into language interpretation and use. We tested which cognitive skills relate to relevance inferencing. In addition, we asked whether children's lab-based pragmatic performance relates to children's parent-assessed pragmatic language skills. We tested 3.5- to 4-year-old speakers of British English (Study 1: $N =$ 40, Study 2: $N = 32$). Children were presented with video-recorded vignettes ending with an utterance requiring a relevance inference, for which children made a forced choice. Study 1 measured children's Theory of Mind, their sentence comprehension and their real-world knowledge and found that only real-world knowledge retained significance in a regression analysis with children's relevance inferencing as the outcome

Corresponding author:
Kirsten Abbot-Smith, Kent Child Development Unit, School of Psychology, Keynes College, University of Kent, Canterbury CT2 7NP, UK.
Email: K.Abbot-Smith@kent.ac.uk

variable. Study 2 then manipulated children's world-knowledge through priming but found this did not improve children's performance on the relevance inferencing task. Study 2 did, however, reveal a significant correlation between children's relevance inferencing and a measure of morpho-syntactic production. In both studies parents rated their children's pragmatic language usage in daily life, which was found to relate to performance in our lab-based relevance inferencing task. This set of studies is the first to empirically demonstrate that lab-based measures of relevance inferencing are reflective of children's pragmatic abilities 'in the wild'. There was no clear association between relevance inferencing and Theory of Mind. There was mixed evidence for the role of formal language, which should be further investigated. Finally, real-world knowledge was indeed associated with relevance inferencing but future experimental work is required to test causal relations.

Keywords

Implicature, inference, relevance, real-world knowledge, Theory of Mind, pre-school, individual differences, priming, typical development

Introduction

Children encounter indirect use of language on a daily basis. Imagine a child asks a friend to play football with him and the friend replies that he has a cough. To interpret this reply as an indirect refusal, the child must infer that 'I have a cough' is somehow relevant to the current 'question under discussion' in the conversation and use this to determine the implied meaning (e.g. Benz & Jasinskaja, 2017). For this reason, some theorists have referred to successful interpretation of this type of indirect language as 'relevance inferencing' (Sperber & Wilson, 2002). Although theorists differ regarding their views on the specific mechanics of this sense-making process (e.g. Grice, 1975; Searle, 1975; Sperber & Wilson, 2002; Tomasello, 2008), they agree that the basis is an assumption by the listener that speakers communicate co-operatively.

Children hear – and appear to successfully interpret – some forms of indirect speech from a fairly early age (e.g. Shatz, 1978). For example, a parent may ask a child 'Can you shut the door' rather than imperatively demanding her to 'Shut the door'. Children below the age of 2 years respond appropriately to the interrogative request by carrying out the action (as opposed to interpreting the parent as querying their ability to shut the door). More implicit indirect speech that requires a relevance inference (e.g. replying 'You've had a lot of biscuits already' to a child's request for a biscuit) is understood only later in development. That is, children appear to have difficulty with such context-dependent interpretation of indirect speech until around the age of 6 years (e.g. Bernicot et al., 2007; Bucciarelli et al., 2003; see also Verbuk & Shultz, 2010). However, with a more goal-directed design (i.e. giving a puppet what she/he wants), Schulze et al. (2013: Study 3) found that even at 36 months, children were able to compute the required relevance inferences and successfully interpreted others' communicative acts 66% of the time overall (see also Schulze & Buttelmann, 2021, regarding the trajectory up until and including 5 years of age). In Schulze et al.'s (2013: Study 3) paradigm, children saw vignettes of two puppets engaging in a short dialogue, as in (1) below:

(1)   Puppet Anna: 'Oh, I slept really well'.
        Puppet Becky: 'I'm really hungry. I'd like to have breakfast'
        E1 [sitting with child]: "Look, I have cereal or toast. You can give one to the puppets. I'll ask them what they'd like to have.
        E1 [to puppets]:   'Do you want cereal or toast?'
        Puppet Anna: 'The milk's all gone'
        E1: [to child]:   "Which one will you give to the puppets?

Three-year-old children assumed Anna's utterance was relevant to the object choice, and inferred that Anna did not want the cereal (given that there was no milk) and thus handed her the toast.

Our focus

The main focus of this study is on investigating which cognitive skills relate to children's ability to successfully compute relevance inferences. Specifically, we were interested in investigating the role of world knowledge, which has previously been stated as an important factor (e.g. in terms of common ground) but has not previously been investigated in relation to relevance inferencing or comprehension of communicative intent. In addition, we also ask whether lab-based measures of relevance inferencing relate to children's parent-reported pragmatic language abilities. This is a crucial step since we need to know which cognitive resources are related to the ability to interpret indirect language in daily interactions. Lab-based judgements are likely to underestimate children's inferencing ability 'in the wild' since these types of lab tasks require children to process and remember conversation between two protagonists and are never likely to be as intrinsically motivating for a child as would be naturalistic dialogue (see also Papafragou & Tantalou, 2004, for additional points).

Cognitive and socio-cognitive correlates of relevance inferencing

Of the many cognitive resources that might be important for processing relevance inferences, theoretical attention has been focused on (1) Theory of Mind (which is the ability to understand the desires, perspectives, emotions, knowledge and beliefs of others and how they may differ from one's own; see Wellman, 2014); (2) formal language ability (i.e. the comprehension and production of vocabulary and grammar); and (3) real-world knowledge. One difficulty with testing relationships between these cognitive resources and inferencing is that they, in turn, tend to be correlated with each other at least to some extent. For example, Theory of Mind tests are positively correlated – usually with a moderate to large effect size – with general language ability (see, for example, Milligan et al., 2007, for a meta-analysis). This relationship is likely to be at least in part due to the fact that Theory of Mind measures require the child to process and remember the verbal information of a vignette, including information about which protagonist likes or knows which things. But it is also likely to reflect the fact that we often learn about other minds in linguistic exchanges. We bear this in mind while reviewing the background empirical literature.

With regard to children's Theory of Mind, studies examining the relationship with the ability to compute relevance inference have yielded mixed results. On one hand,

Huang et al. (2015) found that school-aged children's performance on 'indirect reproach' interpretation (e.g. 'Are you leaving without tidying up'?) was significantly higher for children who passed Theory of Mind tasks than for children who had failed (see also Whyte & Nelson, 2015). On the other hand, De Mulder (2015) found that the pre-schoolers' comprehension of indirect requests (e.g. 'It's really cold outside' said by a mother to a child who is about to go play outside) does not relate to Theory of Mind when formal language is controlled for. One of our current aims was to follow up these findings by investigating the relationship between Theory of Mind and children's ability to make relevance inferences.

Second, for language proficiency, there seems to be ample evidence that formal language skills are related to a range of pragmatic language functions in both typically and atypically developing children (see Matthews et al., 2018, for a review). In one sense, such relationships are not surprising since it would be impossible to take a speaker's perspective and communicative intent into account in order to interpret language if one had not yet acquired the words and morpho-syntax involved. However, it is also highly plausible that the relationship between formal language and pragmatics is much deeper than this. Indeed, a number of theorists have suggested that formal language itself can only be acquired if a child has the ability to consider communicative intent (e.g. H. H. Clark, 1996; Tomasello, 2008) or even that there exists no real distinction between formal language and pragmatic language since linguistic forms (speech sounds, syntactic structures, etc) are used to perform both semantic and pragmatic functions (E. V. Clark, 1990; see Matthews et al., 2018, for a discussion).

Certainly, in the literature on children's inferencing abilities more broadly, formal language has been found to play a hugely important role. This is particularly the case for the literature on the inferences required when comprehending narratives, such as bridging inferences, coherence inferences and anaphor resolution (Currie & Cain, 2015; Davies et al., 2020; Lucas & Norbury, 2015). Furthermore, De Mulder (2015) found that syntax comprehension predicted children's comprehension of indirect requests.

However, a number of studies have not found significant relationships between formal language and implicatures or relevance inferencing. Antoniou and Katsos (2017) found no relation between children's performance on a range of pragmatic phenomena and their performance in a test of expressive vocabulary (see also Antoniou et al., 2019 for similar results). Huang et al. (2015) tested children's receptive vocabulary and found no relation to their interpretation of indirect reproaches. Similarly, Schulze et al. (2020) found no relationship between children's interpretation of relevance implicatures and their receptive vocabulary. Therefore, in these current studies, we also included measures of formal language to investigate whether there is a relationship with relevance inferencing skills in 3-year-olds.

Third, real-world knowledge is logically necessary for most inferencing (e.g. Kintsch, 1988). That is, to correctly interpret 'I have a cough' as a refusal, the child needs to know that people generally do not feel like running around energetically when they have coughs. There is evidence suggesting that relevant background knowledge is crucial for inferencing skills more generally (e.g. Marr & Gormley, 1982; see also Ackerman et al., 1990). However, other studies on coherence inferencing in narratives found that although requisite knowledge is necessary, it is not sufficient for inferencing (Barnes et al., 1996). Therefore, another key aim of these current studies was to investigate whether children

with a higher level of world knowledge would find it easier on the whole to compute relevance inferences.

Thus, we wished to examine the relationship between these three factors (Theory of Mind, formal language and real-world knowledge) and lab-based measures of relevance inferencing. In addition, we also wished to confirm the validity of a lab-based measure of relevance inferencing by determining its relationship with parent-assessments of pragmatic language.

Current studies

We carried out two studies to assess how young children compute relevance inferences. For both studies, we used an adapted version of Schulze et al.'s (2013: Study 3) Relevance Inferencing paradigm. This paradigm has been successfully used with German-speaking and Swiss–German speaking children aged 36 months and above (see also Schulze & Buttelmann, 2021; Schulze et al., 2020). To respond, participants merely have to observe and choose between two objects – or between two pictures on a screen in more recent versions of the paradigm (e.g. Schulze & Buttelmann, 2021) – by making an inference as to which object the protagonist wants based on her indirect speech (i.e. relevance implicature). This method was preferred since a pilot study with school-aged children, summarised in Appendix A in supplemental material, suggested that sentence-length response options might over-burden formal language. We thus presented children with video-recorded vignettes of a short dialogue, ending in an indirect use of language which required a relevance inference for successful interpretation. Both studies used forced-choice behavioural measures of children's relevance inferencing.

> *RQ1*. Our first research question concerned the cognitive correlates of child relevance inferencing. Study 1 addressed the question of which key cognitive abilities were involved in children's communicative abilities, testing the relation with formal language, real-world knowledge and Theory of Mind. Study 2 further explored the role of real-world knowledge by experimentally manipulating children's knowledge through priming.

> *RQ2*. Our second research question was whether lab-based judgement measures of relevance inferencing are ecologically valid. To this end, in both studies, we asked parents to complete a brief questionnaire about their children's conversation skills and examined the relationship with the lab-based assessment.

Study 1

Method

Participants. We tested 48 monolingual English-speaking, typically developing children. Children were recruited through the Kent Child Development Unit in southern England. We pre-excluded any children who either had hearing or possible speech and language difficulties. Parents brought their children to the Kent Child Development Unit's testing

suite. Some parents remained in the waiting room whereas others sat behind their children and were requested not to speak. Five children were tested but excluded because they showed a side bias; that is, they pointed to the same side for all six trials. Three children were excluded due to technical error. The final sample comprised 40 children aged between 41 and 48 months (Mean age = 44.6 months, SD = 2.1), of whom 16 were boys. Ethical approval was obtained from the School of Psychology ethics committee at the University of Kent, UK. Children were told that they could take a break whenever they wished and could end testing entirely if they desired.

Overall procedure

All children saw a relevance inferencing task consisting of six trials, followed by a Theory of Mind measure, the 'Sentences Structures' sub-test of the CELF-P (Wiig et al., 2006), and the 'Information' sub-test of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI, Wechsler, 2012). Parents completed the Mindful Conversational Difficulties Scale (De Rosnay et al., 2014) to assess children's everyday communication capacities. The testing session lasted 45 minutes in total, including a 10-minute 'free play' break in the middle. The test performance was video-recorded.

Relevance inferencing measure

The relevance inferencing task was adapted from Schulze et al. (2013). Following Schulze et al., (2020), we only used six rather than eight vignettes in order to have time to run our additional measures. We therefore selected five of the eight original vignettes (milk, dog's lead, cups, knife, toothpaste) and added a vignette in which children chose between sunglasses and a scarf (see Appendix B in supplemental materials), which was also included in the study by Schulze and Buttelmann (2021).[1] Instead of using live puppets, we followed Schulze and Buttelmann (2021) by presenting children with video-recorded interactions between a 'King' and 'Princess' puppet followed by a screen in which the child was asked to point to the item (out of two options) that the final speaker wanted (e.g. 'Can you touch the one the King wants?'). Each vignette was presented in PsychoPy v. 1.83.04 (Peirce, 2007) and the audio-recordings were of a male and a female speaker of southern British English using child-directed speech.

We developed four script orders for the purposes of counterbalancing. The position of the target object and whether the king or princess uttered the key statement was counterbalanced both within and across script orders. Across script orders we also counterbalanced which object (e.g. cereal vs toast) was the target for a particular vignette. The target object was never on the same side of the screen more than twice in a row. The four script orders were fairly evenly distributed over the final sample of children (order 1 = 22.5%, order 2 = 27.5%, order 3 = 30%, and order 4 = 20%).

Relevance inferencing procedure

Warm-up. Each child first participated in four binary-choice warm-up trials during which they were given feedback if they pointed to the incorrect option. (Children were

not excluded based on incorrect performance during the warm-up.) In the first two warm-up trials, children had to complete two easy forced-choice trials (e.g. 'Can you touch the one where she is eating the banana/cutting the banana' or 'Can you touch the one where she is kicking/catching the ball'). For the second two warm-up trials, children were introduced to the two puppet characters (King and Princess) and these two trials were structured identically to the test trials except that they involved literal language interpretation with no ambiguity. For example, in the final warm-up trial, the King asked, 'Do you want your car or your bouncy ball' and the Princess replied, 'I want the car'. The child was then asked to touch the object that the Princess wanted (car vs ball).

Relevance inferencing test trials. Following this, the children were told that they would see some short films about a day in the life of the two puppet characters (Princess and King) and that they would sometimes have to help them out. Thus, each child first saw the vignette about breakfast and the sixth and final vignette always involved 'going to bed'. No feedback was given for the test trials.

Theory of Mind assessment. The quintessential measure of Theory of Mind is often considered to be whether children understand that another person may hold a false belief. For our target age (3.5–4 years), it is well-established that the majority of children fail measures of false belief understanding (see, for example, Wellman & Liu, 2004). However, the ability to understand that others have different desires, perspectives, and knowledge has its roots much earlier in development.

   Wellman and Liu Theory of Mind Scale. We administered Wellman and Liu's (2004) scale, since it includes measures of the earlier-developing abilities of an understanding of diverse desires and knowledge (see Tomasello, 2008, for discussion). This scale starts with simpler forms of Theory of Mind, namely the understanding that others may like different things to oneself (Diverse Desires). It also assesses Knowledge Access (i.e. the understanding that you cannot 'know' the location of an object unless you have seen it), as well as Diverse Beliefs, where the protagonist believes an object is in a different location to the child but neither knows the true location. We also included the 'Contents False Belief' task from Wellman and Liu's (2004) scale.
   In our assessment, we included all four of the above sub-tests (although the procedure for Knowledge Access was modified slightly in line with Pratt & Bryant, 1990: Study 1).[2] We did not use Wellman and Liu's (2004) False Belief Location nor their Apparent Emotion tasks since they would be extremely difficult for our age group.

   Theory of Mind 'Penny-Hiding' task. To assess children's ability to understand that another person may have differing perspectives from one's own, we also administered the 'Penny-hiding' task, which has previously been used to assess Theory of Mind in preschoolers (Hughes & Ensor, 2005) and autistic children (Baron-Cohen, 1992).
   For this task, the experimenter hid a coin behind her back and then brought both hands forward keeping the coin hidden and asked the child to guess which hand the coin was in. This was carried out three times in total. Then the experimenter told the child that it was his or her turn to hide the coin. The child received one point for putting both hands

behind his or her back, an additional point for bringing both hands forward and a further point if the coin remained hidden until the experimenter guessed which hand it was in.

Additional measures

Language abilities. To assess children's formal language abilities, the 'Sentences Structures' sub-test of the CELF-P (Wiig et al., 2006) was administered. Here, children were required to point to the correct picture (out of four) that corresponded to the experimenter's statement. For example, 'Point to: the boy is being chased by his cat'.

Real-world knowledge. To assess real-world knowledge, we administered the 'Information' sub-test of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 2012), during which children were asked a series of increasingly complex questions such as 'What do people use to stay dry in the rain?'

Mindful Conversational Difficulties Scale. Parents completed the eight-item Mindful Conversational Difficulties Scale (De Rosnay et al., 2014). None of the items assess understanding of implicature or indirect speech but this questionnaire does assess pragmatic language more broadly. It includes items such as 'Does the child adapt appropriately to conversing with different people in varied social situations (e.g. speaks differently to a classmate than the School Principal)?' Parents respond on a five-point scale ranging from 'very much less difficulty/[skill]' to 'very much more difficulty/[skill] than a typical child this age'. Four items are reverse-scored. All eight items are included in the main text of de Rosnay et al. (2014).

Coding and reliability

In the relevance inferencing task, a trial was scored as one if the child pointed to the correct object and zero if the child pointed to the incorrect object. If a child pointed to both objects, this item was removed from analysis[3] (and for this reason, proportion scores were used). The data from five children (12.5% of the data) were coded by a second rater, blind to the original codes. There was perfect agreement between the raters for all trials for all five children ($\kappa = 1.00$).

The children's performance on the additional measures was coded following the test script. For all Theory of Mind measures, the data from five children (12.5% of the data) were coded by a second rater, blind to the original codes. There was perfect agreement between the raters for all trials for all five children.

Results

The full anonymised datasets are available on the Open Science Framework web pages here: https://osf.io/wcg5p/?view_only=3d6f9f5eb33742959bbc2516d7119405. For an overview of the descriptive statistics pertaining to each task, see Table 1.

Overall, children pointed to the correct response 69% of the time, which was above chance, $t(39) = 6.39$, $p < .001$. Our main research question was whether individual differences in relevance inferencing can be explained by children's Theory of Mind, their

Table 1. Overview of the children's performance in the tasks of Study 1.

Construct	Task	Mean (SD)	Max poss. score
Relevance inferencing		.69 (.19)	1.00
Theory of mind	Wellman and Liu Scale Sum-score	2.5 (.8)	4
	Penny-Hiding Task	1.9 (1.1)	3
Formal language	CELF-P 'Sentence structures'	14.8 (3.7)	22
Real-world knowledge	WPPSI 'Information' sub-test	14.7 (3.9)	29
Daily pragmatic skills	Mindful Conversational Difficulties Parental questionnaire	28.4 (3.6)	40

CELF-P: Clinical Evaluation of Language Fundamentals Preschool; SD: standard deviation; WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

formal-language abilities and/or the children's real-world knowledge. As can be seen in the correlation matrix presented as Table 2, children's performance in the relevance inferencing task related to their real-world knowledge ($r = .498$, $p = .001$) and their performance in the 'Penny-Hiding'-ToM-task ($r = .331, p = .037$), but no other measure of Theory of Mind.

RQ1: cognitive correlates of relevance inferencing. We then entered all variables of the correlational analyses into a direct entry linear regression model with children's performance in the relevance inferencing task as the outcome variable (see Table 3). This led to a significant model, $F(5, 31) = 3.75, p = .009$, and accounted for 27.7% of the variance.

As can be seen from Table 3, only real-world knowledge was a significant predictor; it accounted for 18% unique variance as assessed by sr^2. Multi-collinearity was not a concern (tolerance for all $> .7$). The same pattern of results was found when comparing the full model with models in which one of these factors was removed. That is, when age was removed from the full model, this did not lead to a significant difference, $F = 1.60, p = .22$, and the model itself retained significance, $p = .007$, adjusted $R^2 = .26$. When formal language (CELF-P 'Sentence Structures' sub-test – a measure of sentence comprehension) was removed, this also did not lead to a significant difference, $F = 1.02, p = .32$, and the model itself retained significance, $p = .006$, adjusted $R^2 = .27$. Similarly, when the Wellman and Liu Theory of Mind measure was removed, this also did not lead to a significant difference, $F = 1.29, p = .27$, and the model itself retained significance, $p = .007$, adjusted $R^2 = .27$. When the Theory of Mind 'Penny Hiding' task was removed, this also did not lead to a significant difference, $F = 2.73, p = .12$ and the model was significant ($p = .012$), accounting for 24% of variance. In contrast, when real-world knowledge was removed from the model, this did lead to a significant difference, $F = 8.73, p = .006$, and the model itself was no longer significant, $p = .12$, adjusted $R^2 = .10$.

RQ2: ecological validity of lab-based measures. Our second research question addressed the ecological validity of our relevance inferencing task. Using Pearson's R, we found that the correlation of children's relevance inferencing performance and the parental report of children's everyday communication (Mindful Conversational Difficulties Scale) approached significance, $r = .301, p = .059$. In line with the findings for our relevance

Table 2. Correlation matrix showing relationships between the variables assessed in Study 1.

	1	2	3	4	5	6
1. Relevance inferencing	–					
2. Age in months (N = 40)	.11	–				
3. Theory of mind (Wellman & Liu) (N = 40)	−.18	−.09	–			
4. Theory of mind (Penny-Hiding) (N = 40)	.33*	.01	−.02	–		
5. Formal language (CELF-P 'Sentence Structures') (N = 39)	.21	.17	.10	−.05	–	
6. Real-world knowledge (WPPSI 'Information') (N = 38)	.50**	.50**	−.05	.23	.20	–

CELF: Clinical Evaluation of Language Fundamentals-Preschool.
WPPSI: Wechsler Preschool and Primary Scale of Intelligence.
$^{\dagger}p < .1$; $*p < .05$; $**p < .01$; $***p < .001$.

Table 3. Regression analysis for Study 1.

Variable	Beta	SE	t	sr^2
Age (in months)	−.21	.02	1.27	.03
Theory of mind Wellman and Liu scale	−.16	.03	−1.14	.03
Theory of mind Penny-Hiding Task	.24	.02	1.65	.05
Formal language (CELF-P 'Sentence Structures')	.15	.01	1.01	.02
Real-world knowledge (WPPSI 'Information')	.50	.01	2.96**	.18

CELF-P: Clinical Evaluation of Language Fundamentals – Preschool; SE: standard error; WPPSI: Wechsler Preschool and Primary Scale of Intelligence.
sr^2 = squared part correlation.
$^{\dagger}p < .1$; $*p < .05$; $**p < .01$; $***p < .001$.

inferencing measure, the Mindful Conversational Difficulties Scale did not correlate with age, $r = .20, p = .21$, the Wellman and Liu Theory of Mind measure, $r = .034, p = .833$, nor the Theory of Mind 'Penny Hiding' task, $r = −.07, p = .65$. However, it did correlate both with real-world knowledge, $r = .42, p = .009$, and formal language (CELF-P 'Sentence Structures'; $r = .32, p = .045$).

Discussion

Study 1 found that individual differences in children's relevance inferencing abilities were not reliably related to their Theory of Mind-capacities (as measured with tasks 1–4 from the Wellman and Liu scale, Wellman & Liu, 2004, and the penny hiding task) or their formal language abilities (sentence comprehension). In contrast, real-world knowledge (as measured by the 'Information' sub-test of the WPPSI, Wechsler, 2012) accounted for 18% of the variance of children's relevance inferencing performance. Finally, a medium-sized correlation (of borderline statistical significance) between the relevance inferencing task and parental ratings of children's everyday communication suggested the lab-based relevance inferencing tasks may be a reasonably valid assessment of childhood pragmatic ability.

While the correlation between real-world knowledge and inferencing in this study certainly suggests it might be a limiting factor for many children, the items on the

Table 4. Means (SD in brackets) for participant characteristics in Study 2.

	Priming (n = 16, 7 male)	Control (n = 16, 8 male)		
	Mean (SD)	Mean (SD)	*p*	d
Chronological age (months)	44.06 (2.05)	44.88 (1.93)	.26	0.40
Formal language (CELF-P 'Word Structure') scaled scores	11.13 (2.13)	11.75 (1.88)	.39	0.31
Parent report (Mindful Conversational Difficulties)	28.53 (3.94)	30.09 (3.76)	.27	0.41

CELF: Clinical Evaluation of Language Fundamentals; SD: standard deviation.

standardised tests of world knowledge did not relate directly to the experimental items on the inferencing test. Therefore, our primary aim in Study 2 was to examine the influence of the specific world knowledge required in the relevance inferencing task by experimentally manipulating it prior to assessment of inference (RQ1). Since we had not found a relationship between sentence *comprehension* and relevance inferencing in Study 1, we explored whether there might be a relationship between relevance inferencing and a different measure of formal language; in Study 2, we replaced the 'Sentence Structures' sub-test with the 'Word Structure' sub-test of the CELF-P, which measures expressive morpho-syntax, a domain in which there is rapid growth in this age group.

To answer our second key research question (RQ2), we also repeated the use of the parent questionnaire to assess validity.

Study 2

Method

Participants. We tested 39 monolingual English-speaking children in the Kent Child Development Unit. Five children were excluded in total. Four of these were excluded because they showed a side bias in the test trial phase and the final child was excluded because the parents indicated a referral for a suspected developmental disorder. Two additional children indicated they did not wish to complete testing. The final sample was thus 32 children aged between 41 and 47 months (Mean age = 44.5 months, SD = 2.0), of whom 15 were boys. The children were assigned to one of two conditions ('Priming' vs 'Control'). The two conditions did not differ from one another in terms of chronological age, gender, formal language scores (performance on the 'Word Structure' sub-test of the CELF-P) or the parent-completed Mindful Conversational Difficulties measure, as can be seen in Table 4.

Design, materials and set-up

All children were engaged in a warm-up and when they were comfortable talking to the experimenter, completed the 'Word Structure' sub-test of the CELF-P (Wiig et al., 2006). Based on this measure, children were assigned to one of the two conditions, so that the

children in both conditions had equivalent formal language abilities. The Priming and the Control condition each contained 12 short silent video-clips. This was followed by the relevance inferencing task as described in Study 1 and then a 'specific world knowledge' post-test quiz.

Parents completed the Mindful Conversational Difficulties Scale (de Rosnay et al., 2014). The testing session lasted 35 minutes in total. Video and audio stimuli were presented in the same manner as for Study 1 (through PsychoPy and using the same screen, videos and sound files).

Procedure

Priming. Each child was shown 12 silent videos clips, with a total length of 5 minutes 20 seconds. In the Priming condition, children saw videos relating to the items in the relevance inferencing task (e.g. a man pouring milk on cereal and then a different man putting butter on toast). In the Control condition, all the videos bore no relation to the relevance inferencing task. Rather, they involved colours, shapes and numbers (e.g. women running in a race and later holding up their position numbers; a woman placing colour cards on colour slots in a board game).

Relevance inferencing task. We used the same task as in Study 1.

Specific world knowledge. After the end of the relevance inferencing task, each child was asked the following 'Real-world knowledge' questions, which related to the actual items in the relevance inferencing task:

1. Do people like to drink out of dirty cups or out of clean cups?
2. Do people wear sunglasses when it is sunny or when it is cloudy?
3. Do people put milk on toast or on cereal?
4. Do people put leads on dogs or on cats?
5. Do people put toothpaste on hairbrushes or on toothbrushes?
6. Do people use knives to cut cakes or to cut biscuits?

Coding and reliability

In the relevance inferencing task, a trial was scored as one if the child pointed to the correct object and zero if the child pointed to the incorrect object. For 20% of the data, a second rater coded blind to the primary rater's coding. Cohen's Kappa showed perfect agreement among the coders ($\kappa = 1.00$).

The specific world knowledge was scored as one if the child gave the correct (conventional) answer (i.e. said that people like to drink out of clean cups).

Results

The full anonymised dataset is available on OSF here: https://osf.io/wcg5p/?view_only= 3d6f9f5eb33742959bbc2516d7119405

Table 5. Means (SD in brackets) for relevance inferencing and specific world knowledge task (Study 2).

	Priming	Control		
	Mean (SD)	Mean (SD)	*p*	*d*
Relevance inferencing	76.04% (17.18)	72.92 (20.07)	0.64	0.16
Specific world knowledge task	79.17 (25.46)	83.33 (19.25)	0.61	0.18

SD: standard deviation.

Main analyses

RQ1: does priming real-world knowledge improve relevance inferencing? Overall, children pointed to the correct response 74.5% ($SD = 18.5$) of the time, which was above chance, $t(31) = 7.51$, $p < .001$. Our first research question was whether priming real-world knowledge would enhance children's relevance inferencing abilities. Children in the Priming condition did not perform significantly better on relevance inferencing overall than the children in the Control condition (see Table 5). Moreover, priming did not appear to lead to an increased activation of the requisite real-world knowledge itself, as there was no significant difference between the Priming condition and the Control condition in the specific knowledge post-test (see Table 5).

RQ2: ecological validity of lab-based measures. Our second research question was whether children's scores on a parent-rated questionnaire of their everyday communication related to their performance on relevance inferencing (i.e. ecological validity). In Study 2, we found that children's relevance inferencing abilities were indeed strongly correlated with the parent-rated Mindful Conversational Difficulties Scale, $r = .514$, $p = .003$.

Secondary analyses

Since there were no differences between our priming and control group, we pooled the data across groups to investigate relationships between age and relevance inferencing and also between expressive language (as assessed by the CELF-P 'Word Structure' subtest) and relevance inferencing. As can be seen in Table 6, expressive language was related and this showed a strong positive relationship overall with performance on the relevance inferencing task, $r(32) = .47$, $p = .006$.[4]

Combining Studies 1 and 2

To further investigate, the relationship between our relevance inferencing measure and the Mindful Conversational Difficulties Scale, we combined the data for these two variables across our two studies. (Recall that the age groups are identical, recruited from and

Table 6. Correlation matrix for Study 2.

	Age in months	Formal language (CELF word structure) raw scores
Relevance inferencing	.06	.47**
Formal language (CELF 'Word Structure') raw scores	.31†	

CELF: Clinical Evaluation of Language Fundamentals.
†$p < .1$; *$p < .05$; **$p < .01$; ***$p < .001$.

tested in the same developmental lab and the two variables were presented in the same way to participants in the two studies.) Again, using Pearson's R, we found a significant relationship between the two variables, $r = .41, p < .001$.

Discussion

In this study, we asked whether priming the world knowledge required in the relevance inferencing task would enhance children's performance in this task given that Study 1 found that real-world knowledge correlated with children's inferencing abilities. We found that the children's performance in the Priming condition did not differ from that of children in the Control condition. Thus, supporting world knowledge seemingly did not influence children's relevance inferencing in this study. However, we also found that priming this knowledge did not result in differences in the knowledge of children in either condition. That is, the priming did not lead greater accuracy on a test of the relevant real-world knowledge. Since the prime was not effective as intended, we cannot know whether promoting real-world knowledge would indeed promote inferencing. Given that the vignettes of the relevance inferencing task were designed to age-appropriately match children's knowledge, future research should address this question further by constructing vignettes that call on knowledge only just within or slightly beyond the children's current state of knowledge and then manipulating this by giving the children the relevant information through priming.

A second question addressed the ecological validity of our paradigm. In Study 2, we found that parental ratings of children's communication strongly correlated with their relevance inferencing abilities. The fact that parent ratings were associated with the lab measure in both studies suggests that the communication task developed by Schulze et al. (2013) does tap important real-life pragmatic skills of young children.

General discussion

We carried out two studies using video-based vignettes followed by forced-choice measures to examine how young monolingual, typically developing preschool children compute relevance inferences. Our first research question was which key cognitive abilities are related to individual differences in the relevance inferencing abilities in young children. In Study 1, we found that real-world knowledge explained variance in children's performance on the relevance inferencing task. While one measure of the

perspective-taking component of Theory of Mind (the 'Penny Hiding Task') was correlated with relevance inferencing, neither this, the other theory of mind measure nor the language comprehension measure (CELF-P 'Sentence Structures') were significant predictors in our regression analysis. In Study 2, a different measure of formal language (CELF-P Word Structure) did predict relevance inferencing.

Our second research question was whether parental assessment of children's conversation skills in daily life, as assessed by the Mindful Conversational Difficulties Scale, is related to their in-lab performance in the relevance inferencing task. In Study 1, the relationship between the Mindful Conversational Difficulties Scale and relevance inferencing approached significance. In Study 2, this relationship was strong. Moreover, when data were collapsed over both studies, the relationship was significant, suggesting that lab-based relevance inferencing measures reflect real-life pragmatic abilities.

Regarding the relative role of cognitive and socio-cognitive factors, from the perspective of Relevance Theory (Sperber & Wilson, 2002), it is perhaps surprising that we found no robust evidence that Theory of Mind relates to relevance inferencing. According to Relevance Theory, to understand indirect speech, a child must consider the communicative intention. This certainly involves mentalising or Theory of Mind at some level as the correct interpretation requires the listener to consider the perspective of the speaker in some manner.

That said, as Sperber and Wilson (2002) themselves point out, a consideration of communication intentions may not require that a listener accesses all aspects of Theory of Mind. There is certainly no logical reason why relevance inferencing should require an understanding of false belief, for example. Furthermore, there are many reasons why a child may be aware that there is an intended, relevant meaning but nonetheless not alight on what the exact intended meaning is. To give one example, at the most basic level, if a child asks a forced-choice question such as 'Do you want cereal or toast?' and receives in response neither a direct reference to cereal nor a direct reference to toast, this might lead to the realisation that a response such as 'there's no milk' is likely to relate in some manner to the child's question. If, however, the child's family does not eat cereal with milk, he or she may still nonetheless be at a loss as to the intended meaning (see Schulze & Buttelmann, 2021, for discussion).

It is worth noting that the Theory of Mind 'Penny hiding' measure, while not a significant predictor in the regression analysis, was correlated with relevance inferencing. However, unexpectedly, it was not correlated with the other measure of Theory of Mind. Future research should explore – with a larger sample size – relationships between Theory of Mind and pragmatic development using multiple measures of perspective-taking, understanding of knowledge, false belief understanding and perhaps also non-verbal measures of communicative intention reading (for the latter, see, for example, Bohn et al., 2019; Schulze & Tomasello, 2015).

When it comes to formal language (vocabulary, morpho-syntax), we would also need further research to clarify mixed findings. On one hand, in Study 1, we did not find a relation between relevance inferencing and children's sentence *comprehension*, as measured by the 'Sentence Structures' sub-test of the CELF-P (although the latter did correlate with the parental questionnaire about the child's naturalistic verbal social communication more generally). On the other hand, in Study 2, we did find a significant

relationship between *expressive* language (morpho-syntax) and relevance inferencing. Our pilot study (see Appendix A in supplemental materials) also found a significant relationship between expressive language (CELF-5 'Formulated Sentences') and relevance inferencing, albeit, using a much more complex and linguistically demanding relevance inferencing task.

One difficulty with interpreting these conflicting findings for formal language is that all these standardised measures are to reflect several interrelated aspects of formal language which are presumably implicated in relevance inferencing. These include sentence processing speed, depth of semantic knowledge and the speed with which relevant associations can be accessed. We, therefore, suggest that further exploration of the role of all these aspects of language processing is necessary before firm conclusions can be drawn. A first step might be to test whether expressive language measures are reliably more closely associated with relevance inferencing than receptive language measures.

Finally, children's general level of real-world knowledge was a significant predictor of relevance inferencing in Study 1. This seems reasonable since, for most relevance inferences, specific real-world knowledge is required. That is, one cannot determine the communicative intent of 'there's no milk' in response to 'do you want cereal or toast for breakfast' without knowing that we put milk on cereal. Moreover, the findings regarding the important role of real-world knowledge are in line with those for other domains of pragmatic development, including metaphor comprehension, which depends heavily on understanding of the semantic domains being aligned (Winner, 1988). For example, even 3-year-olds show a surprising grasp of novel metaphors, if they have already acquired the world knowledge and vocabulary needed to access the intended meaning (e.g. Pouscoulous & Tomasello, 2020; see also Winner, 1988).

Of course, it may also be that the significant influence of real-world knowledge found in Study 1 could be due to some other – as yet untapped – characteristic. However, real-world knowledge not only related to our lab-based measure of relevance inferencing in Study 1 but also to the more general parent report of children's everyday pragmatic abilities. Thus, Study 1 suggests that a lack of general knowledge across the board is a substantial stumbling block for some children when making relevance inferences.

In Study 2, we sought to test the causal role of real-world knowledge in drawing (or failing to draw) relevance inferences by manipulating the availability of the specific real-world knowledge required. Unfortunately, though, this experimental manipulation appears not to have been successful. This could be because the vignettes that we used to assess children's relevance inferencing were designed to match their real-world knowledge and thus, the children were already at ceiling in terms of potential priming effects. Alternatively, it could be that priming might be more effective if we had used language to highlight relevant aspects of the priming videos and/or if the priming videos had been a little more exciting and engaging. In the absence of a successful manipulation of real-world knowledge in this study, future studies should attempt to assess the role of real-world knowledge through an alternative design, perhaps by also including a range of vignettes for which children this age have not yet usually required the requisite knowledge (so that they need to rely on the experimentally provided world knowledge to draw the correct inferences).

Conclusion

This article is the first to empirically demonstrate that lab-based measures of relevance inferencing are reflective of children's pragmatic abilities 'in the wild' (c.f. Holtgraves, 1997: Study 6, for adults). Individual differences in the lab-based measure were not clearly associated with individual differences in Theory of Mind. The role of formal language warrants further investigation. Performance in the relevance inferencing task was associated with individual differences in children's general real-world knowledge. Future experimental work is required to test the direction of causality here.

ORCID iDs

Kirsten Abbot-Smith https://orcid.org/0000-0001-8623-0664

Cornelia Schulze https://orcid.org/0000-0001-9151-4792

Supplemental material

The supplemental material for this article is available online.

Notes

1. Certain vignettes (e.g. tram vs car) could not be used for cultural reasons. Other vignettes in the original study had objects (e.g. jigsaw vs board game) which were difficult to recognise on the screen.
2. The knowledge-access task was modified slightly (see Pratt & Bryant, 1990: Study 1), such that one figure 'lifts' a box whereas the other 'looks in' the box and the child is then asked who knows what is in the box.
3. Only one child did this for one item in Study 1.
4. A linear regression analysis found the same pattern of results. CELF 'Word Structure' raw score, Beta = .501, SE = .01, t = 2.93, p = .007.

References

Ackerman, B., Silver, D., & Glickman, L. (1990). Concept availability in the causal inferences of children and adults. *Child Development*, *61*, 230–246.

Antoniou, K., & Katsos, N. (2017). The effect of childhood multilingualism and bilectalism on implicature understanding. *Applied Psycholinguistics*, *38*(4), 787–833.

Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2019). How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings? *Bilingualism: Language and Cognition*, *23*(1), 186–203.

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology*, *61*(3), 216–241.

Baron-Cohen, S. (1992). Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology and Psychiatry*, *33*, 1141–1155.

Benz, A., & Jasinskaja, K. (2017). Questions Under Discussion: From Sentence to Discourse. *Discourse Processes*, *54*(3), 177–186.

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, *39*(12), 2115–2132.

Bohn, M., Kachel, G., & Tomasello, M. (2019). Young children spontaneously recreate core properties of language in a new modality. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(51), 26072–26077.

Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, *35*(2), 207–241.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*(02), 417–431.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, *137*, 57–75.

Davies, C., McGillion, M., Rowland, C., & Matthews, D. (2020). Can inferencing be trained in preschoolers using shared book reading? A randomised controlled trial of parents' inference-eliciting questions on oral inferencing ability. *Journal of Child Language*, *47*(3), 655–679.

De Mulder, H. (2015). Developing communicative competence: A longitudinal study of the acquisition of mental state terms and indirect requests. *Journal of Child Language*, *42*, 969–1005.

De Rosnay, M., Fink, E., Begeer, S., Slaughter, V., & Peterson, C. (2014). Talking theory of mind talk: Young school-aged children's everyday conversation and understanding of mind and emotion. *Journal of Child Language*, *41*(05), 1179–1193.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). Academic.

Holtgraves, T. (1997). Styles of language use: Individual and cultural variability in conversational indirectness. *Journal of Personality and Social Psychology*, *73*(3), 624–637.

Huang, S. F., Oi, M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: The role of theory of mind and receptive vocabulary. *Clinical Linguistics and Phonetics*, *29*(8–10), 764–775.

Hughes, C., & Ensor, R. (2005). Executive function and theory of mind in 2 year olds: A family affair? *Developmental Neuropsychology*, *28*(2), 645–668.

Kintsch, W. (1988). The role of knowledge in discourse comprehension. *Psycholinguistics: Critical Concepts in Psychology*, *95*(2), 369.

Lucas, R., & Norbury, C. (2015). Making inferences from text: It's vocabulary that matters. *Journal of Speech, Language, and Hearing Research*, *58*, 1224–1232.

Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly*, *18*, 89–104.

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*(3), 186–223.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646.

Papafragou, A., & Tantalou, N. (2004). Children's computation of implicatures. *Language Acquisition*, *12*(1), 71–82.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13.

Pouscoulous, N., & Tomasello, M. (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, *156*, 160–167.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*, 973–982.

Schulze, C., & Buttelmann, D. (2021). Children understand communication intuitively, but indirect communication makes them think twice – Evidence from pupillometry and looking patterns. *Journal of Experimental Child Psychology*, *206*, Article 105105.

Schulze, C., Endesfelder Quick, A., Gampe, A., & Daum, M. (2020). Understanding verbal indirect communication in monolingual and bilingual children. *Cognitive Development*, *55*, Article 100912.

Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-year-old children make relevance inferences in indirect verbal communication. *Child Development*, *84*(6), 2079–2093.

Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, *136*, 91–98.

Searle, J. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (*Vol. 3*, pp. 83–106). Academic Press.

Shatz, M. (1978). Children's comprehension of their mothers' question-directives. *Journal of Child Language*, *5*(1), 39–46.

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

Tomasello, M. (2008). *Origins of human communication*. MIT Press.

Verbuk, A., & Shultz, T. (2010). Acquisition of relevance implicatures: A case against a rationality-based account of conversational implicatures. *Journal of Pragmatics*, *42*(8), 2297–2313.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence – Fourth edition technical manual and interpretive manual*. The Psychological Corp.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541.

Whyte, E., & Nelson, K. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders*, *54*, 2–14.

Wiig, E. H., Secord, W., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals – Preschool 2 (CELF-P)*. Pearson.

Winner, E. (1988). *The point of words: Children's understanding of metaphor and irony*. Harvard University Press.