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

Kirsten Abbot-Smith

Cornelia Schulze

Nefeli Anagnostopoulou

Maria Zajączkowska

Danielle Matthews

Word count = 5260

Contact: Kirsten Abbot-Smith
Dr Kirsten Abbot-Smith
Kent Child Development Unit
School of Psychology,
University of Kent, Keynes College
Canterbury CT2 7NP, UK
Email: K.Abbot-Smith@kent.ac.uk
Tel: +44 (0)1227 823016

Acknowledgments
A special 'thank you' to the children and to the parents who brought their children into the Kent Child Development Unit. An additional thank you to Katherine Billinghurst, Kerrie Bragg, Annabel Carter, Lauren Jenner, Chelna Karnavat, Melanie Lopez-Perez, Dr Louise Malkin, Nancy Shokeen, and Anna Xygkou for assistance with stimuli creation, data collection, data storage or coding. A huge thank you to Frank Gasking for assistance with PsychoPy.

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. Additionally, we asked whether children’s lab-based pragmatic performance relates to children’s parent-assessed pragmatic language skills. We tested 3½- to 4-year-olds (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 variable. Study 2 then manipulated children’s world-knowledge via priming but found this did not improve children’s performance on the relevance inferencing task. Study 2 did, however, find 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’. We argue that real-world knowledge is a necessary (but not sufficient) for relevance inferencing.

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.

Although children hear – and appear to successfully interpret- some forms of indirect speech from a fairly early age (e.g. Schatz, 1978), in most experimental studies that have examined this question, children appear to have difficulty with context-dependent interpretation of indirect speech until around the age of six years (e.g. Bernicot, Laval & Chaminaud, 2007; Bucciarelli, Colle & Bara, 2003; see also Verbuk & Schulz, 2010). With a more goal-directed design (i.e. giving a puppet what s/he wants), however, Schulze, Grassmann and Tomasello (2013: Study 3) found that even at 36 months children were able to compute relevance inferences and successfully interpret others' communicative acts. 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.

Cognitive and socio-cognitive correlates of relevance inferencing

The main focus of the current study is on investigating which cognitive skills relate to children’s ability to successfully compute relevance inferences. In addition, we also ask whether lab-based measures of relevance inferencing relate to children’s naturalistic 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, particularly as lab-based judgements are likely to underestimate children’s inferencing ability ‘in the wild’ (see Papafragou & Tantalou, 2004, for discussion).

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 (that is, 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 is correlated with language ability (e.g. Milligan, Astington & Dack, 2007). We bear this in mind while reviewing the background empirical literature.

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

Oi and Taguichi (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.

Secondly, 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, Biney & Abbot-Smith, 2018, for a review). 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, McGillion, Rowland & Matthews, 2019; Lucas & Norbury, 2015).

However, 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 and colleagues (2015) tested children's receptive vocabulary and found no relation to their interpretation of indirect reproaches. Similarly, Schulze, Endesfelder Quick, Gampe and Daum (2020) found no relationship between children's interpretation of relevance implicatures and their receptive vocabulary. However, it is possible that proficiency with sentence processing is more critical than vocabulary skills; de Mulder (2015) found that syntax comprehension but not receptive vocabulary predicted children's comprehension of indirect requests. Therefore, in the current study, we followed up de Mulder's (2015) finding by investigating whether there is a

relationship between sentence comprehension and relevance inferencing skills in three-year-olds.

Thirdly, 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 from the reading acquisition literature suggesting that relevant background knowledge is crucial for inferencing skills more generally (e.g. Marr & Gormley, 1982; see also Ackerman, Silver & Glickman, 1990). However, other studies on coherence inferencing in narratives found that although requisite knowledge is necessary, it is not sufficient for inferencing (Barnes, Dennis & Haefele-Kalvaitis, 1996). Therefore, another key aim of the 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.

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 up (see also Schulze et al., 2020; Schulze & Buttelmann, 2021) and avoids the problems associated with tasks that involve wordy response options. (Our pilot study with school-aged children, summarised in Appendix A, suggested wordy responses might over-burden formal language). We 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 via 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 and tested in a university developmental unit in southern England. We pre-excluded any children who either had hearing or possible speech and language difficulties. Five children were tested but excluded because they showed a side bias and 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 first authors' university. Children were told that they could take a break whenever they wished and could end testing entirely if they desired.

Design, materials & set-up

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 (Peterson et al., 2009) to assess children's every-day 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.

The relevance inferencing task was adapted from Schulze et al. (2013). We 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). Instead of using live puppets, we presented 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 female speaker of southern British English using child-directed speech.

As in the original study, for half the trials the precondition for the proposed action was unfulfilled; that is, when King and Princess had breakfast and were about to choose between cereal and toast, the King said that the milk was all gone. The children had to infer that because most people do not eat cereal without milk, the King wanted toast instead. The other half of the test trials seen by a given child involved a fulfilled pre-condition; that is, in the example above, the final statement would have been 'There's milk' and the children would thus have to infer that the King wanted the cereal.

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 whether a particular vignette contained a 'fulfilled' or 'unfulfilled' pre-condition. 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%).

To assess children's Theory of Mind capacities, the following measures were administered. The first Theory of Mind measure was an adaptation of the scale developed by Wellmann & Liu (2004). The first two tasks (Diverse Desires and Diverse Beliefs) and the fourth task (Contents False Belief) followed the script in Wellman and Liu (2004; Appendix). The third task (Knowledge-Access) was modified slightly in line with Pratt and Bryant (1990: Study 1). We did not use Wellmann and Liu (2004)'s fifth and sixth tasks since they are too difficult for our age group. For the second Theory of Mind measure, we then carried out the 'Penny-hiding' task, which has previously been used to assess Theory of Mind in pre-schoolers (Hughes & Ensor, 2005) and autistic children (Baron-Cohen, 1992).

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. 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 task. 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 task.

We followed the script by Wellman and Liu (2004) for their scale's test on diverse desires and diverse beliefs. The knowledge-access task was modified slightly (see Pratt & Bryant, 1990: Study 1), that is, 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. This was followed by a contents false belief task (Wellman & Liu, 2004: Task 4). We then carried out the 'Penny-hiding' task (Baron-Cohen, 1992; Hughes & Ensor, 2005). In 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 / her turn to hide the coin. The child received one point for putting both hands behind his / 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.

Language abilities. To assess children 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 (Peterson et al., 2009). This questionnaire 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.

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. The data from five children (12.5% of the data) was 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 in the additional measures were coded following the test script. For all Theory of Mind measures, the data from five children (12.5% of the data) was 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 Sciences Framework web pages here: [BLINDED DURING REVIEW]. For an overview of the descriptive statistics pertaining to each task, see Table 1.

INSERT TABLE 1 ABOUT HERE

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	Parental questionnaire	28.4 (3.6)	40

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 formal-language abilities and / or the children's real world knowledge. 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 (see Table 2).

INSERT TABLE 2 ABOUT HERE

Table 2.

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

	Age in months (N = 40)	ToM (Wellman & Liu) (N = 40)	ToM (Penny-Hiding) (N = 40)	Formal Language (N = 39)	Real-World Knowledge (N = 38)
Relevance Inferencing	.11	-.18	.33*	.21	.50**
Theory of Mind (Wellman & Liu)	-.09		-.02	.10	-.05
Theory of Mind (Penny-Hiding)	.01			-.05	.23
Formal Language	.17				.20
Real-World Knowledge	.50**				

Note. $^{\forall}$ = $p < .1$, * = $p < .05$, ** = $p < .01$, *** = $p < .001$

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.

INSERT TABLE 3 ABOUT HERE

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 'Sentence Structures')	.15	.01	1.01	.02
Real World Knowledge (WPPSI 'Information')	.50	.01	2.96**	.18

Note. $^{\forall}$ = $p < .1$, * = $p < .05$, ** = $p < .01$, *** = $p < .001$, sr^2 = squared part correlation

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 '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. We found that the correlation of children's relevance inferencing performance and the parental report of children's every-day communication (Mindful Conversational Difficulties Scale) approached significance ($r = .301$, $p = .059$).

Discussion

Study 1 found individual differences in children's relevance inferencing abilities were not related to their Theory of Mind-capacities (as measured with task 1-4 from the Wellman and Liu scale, Wellman & Liu, 2004) 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 every-day communication suggested the lab tasks are reasonably ecological valid.

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 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 language *production*; in Study 2, we replaced the 'Sentence Structures' sub-test with the 'Word Structure' sub-test of the CELF-P.

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

Study 2

Method

Participants

We tested 39 monolingual English-speaking children in the same university developmental lab as the previous study. Five children were excluded because they showed a side bias in the test trial phase ($n = 4$) or because the parents indicated referral for a suspected developmental disorder ($n = 1$). 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) nor on the parent-completed Mindful Conversational Difficulties Scale, as can be seen in Table 4 below.

INSERT TABLE 4 ABOUT HERE

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 scaled scores	11.13 (2.13)	11.75 (1.88)	.39	0.31
Parent Report	28.53 (3.94)	30.09 (3.76)	.27	0.41

Design, materials & 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.

Parents completed the Mindful Conversational Difficulties Scale (Petersen et al., 2009). The testing session lasted 35 minutes in total. Video and audio stimuli were presented in the same manner as for study 1 (via 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

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 (M = 76.0% correct, SD = 17.2) than the children in the Control condition (M = 72.9% correct, SD = 20.1) ($t(30) = 0.47$, $p = .64$, $d = 0.16$). However, 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 (M = 79.2% correct, SD = 25.5) and the Control condition (M = 83.3% correct, SD = 19.2) in the specific knowledge post-test ($t(30) = .52$, p = .61).

RQ2: Ecological validity of lab-based measures

Our second research question was whether children's scores on a parent-rated questionnaire of their every-day communication related to their performance on relevance inferencing (i.e., ecological validity). We found that children's relevance inferencing abilities 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' sub-test) and relevance inferencing. Only expressive language was related and this showed a strong positive relationship overall with performance on the relevance inferencing task ($r(32) = .47$, p = .006).

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 to higher levels 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 require an amount of knowledge that is slightly above the children's current knowledge and then manipulate this factor by giving the children the relevant information via 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 and colleagues (2013) does tap important real-life 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 children compute relevance inferences. We asked which key cognitive abilities would relate to individual differences in the relevance inferencing abilities in young children. In Study 1, we found that real-world knowledge (but not Theory of Mind or sentence comprehension) explained variance of children's performance in the relevance inferencing task.

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 can be considered Theory of Mind in a broad sense although arguably not one that requires a deep level of mentalising. That said, even though the Theory of Mind 'Penny hiding' measure did not retain significance as a predictor of children's relevance inferencing in the regression analysis, it still correlated with relevance inferencing. Indeed, it would seem likely that certain aspects of social communication might be important for the interpretation of indirect language at some level. Future research should explore this useful a wider variety of socio-cognitive measures and with a larger sample size.

The findings with regards to formal language (vocabulary, morpho-syntax) are less clear-cut. On the 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). This is in line with Huang and colleagues (2005) and Schulze and colleagues (2020) who tested receptive vocabulary. On the other hand, in Study 2 we did find a significant relationship between expressive language and relevance inferencing. The latter finding ties in with a number of studies have found strong relations between formal language and other pragmatic phenomena (see Matthews et al., 2018 for a review). Moreover, our pilot study (see Appendix A) also found a significant relationship between expressive language and relevance inferencing, albeit using a much more complex and linguistically demanding task. One difficulty with interpreting these findings is that all these standardised measures are presumably serving as a proxy for those aspects of formal language which are 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.

Finally, children's general level of real-world knowledge was a significant predictor of relevance inferencing in Study 1. Of course, 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. Study 1 suggests that 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. Finally, it could also be that the influence of real-world knowledge found in Study 1 could in fact be due to other – as yet untapped – characteristics.

Our second research question was whether parental assessment of children's conversation skills in daily life, as assessed by the Mindful Conversational Difficulties Scale, would relate to their in-lab performance in the relevance inferencing task. In Study 1, we found a marginally significant relationship between the Mindful Conversational Difficulties Scale and relevance inferencing. In Study 2, this relationship was strong, suggesting that lab-based relevance inferencing measures reflect real-life pragmatic abilities.

Conclusions

The current paper is the first to empirically demonstrate that lab-based measures of relevance inferencing are reflective of children's pragmatic abilities 'in the wild' (cf. Holtgraves, 1997: Study 6, for adults). Individual differences in the lab-based measure were not 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. Since we were unable

to experimentally manipulate the specific knowledge required for our relevance inferencing items in study 2, future experimental work is required to test causal relations.

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,

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.

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.

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.

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

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.

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

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, 6, 1200

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

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.

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

Horst, J. (2013). Context and repetition in word learning. *Frontiers in Psychology*, vol 4 article 149.

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.

Peterson, C. C., Garnett, M., Kelly, A., & Attwood, T. (2009). Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *European Child & Adolescent Psychiatry, 18*(2): 105-115.

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

Searle, J. (1975). Indirect speech acts. In P. Cole & J. L.Morgan (Eds.), Syntax and semantics. Vol. 3. New York: Academic Press.

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

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

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*, 105105.

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

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.

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

Wechsler, D. (2003). The Wechsler intelligence scale for children (4th ed.). San Antonio: TX: The Psychological Corporation.

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. Oxford Univ. 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., Secord, W. & Semel, E. (2006). Clinical Evaluation of Language Fundamentals – Preschool 2 (CELF-P). London, UK: Pearson.

Wiig, E.H., Semel, E., & Secord, W. (2013) Clinical Evaluation of Language Fundamentals – Fifth Edition (CELF-5). San Antonio, TX: NCS Pearson Inc.

Supplementary materials

Appendix A (Pilot-Study)

Participants: 57 typically-developing monolingual British-English speaking children (*N* = 25 boys; Mean age = 90.56 months, range 73-109 months)

Task: Five video-based mini interaction between two protagonists, ending with an indirect statement. The child is asked to select one of the three options which best summarises what the speaker means.

Example: Sam and Tom play in the same football team.

Sam: "Tom, we desperately need you to play in the football match today"

Tom: "I don't feel well. I'm coughing and sneezing".

DV (forced choice, reaction time): "What does Tom mean?"

A. I can't play because the weather is not good for playing football.

B. I am warning you of my cold so you don't catch it.

C. I can't play in your football match because I'm ill.

Additional measures: Nonverbal IQ (NVIQ, assessed using the 'Matrices' sub-test of the British Ability Scale – 3, Elliot & Smith, 2011), sentence production (assessed using the 'Formulated Sentences' sub-test of the Clinical Evaluation of Language Fundamentals – Fifth Edition, Wiig, Semel & Secord, 2013), real-world knowledge (assessed using the 'Information' sub-test of the WISC, Wechsler, 2003), and Theory of Mind (assessed using four items from Happé's (1994) 'Strange Stories').

Results:

	Age in months (N = 57)	NVIQ raw score (N = 53)	Expressive Language raw score (N = 57)	Real-world Knowledge raw score (N = 57)	Theory of Mind (N = 52)
Relevance Inferencing	.21	-.09	.38**	.22$^{\forall}$	.25$^{\forall}$
NVIQ raw score	.48**		.32*	.56***	.29*

Formal Language raw score	.59***	.32*		.66***	.52***
Real-world Knowledge raw score	.64***	.56**	.66***		.45**
Theory of Mind	.43**	.29*	.52***	.45**	

∀ = p < .1 * = p < .05 ** = p < .01 *** = p < .001

In a linear regression analysis (N = 51) only core language (CELF 'Formulated Sentences') was a significant predictor for Relevance Inferencing (β = .39, SE = .004, *t* = 2.044, *p* = .047), accounting for 7% unique variance as assessed by the squared part correlation. The model itself (F(5, 45) = 2.43, *p* = .05) accounted for 13% of variance.

Appendix B (Materials for Study 1 and Study 2)

1.

a. Precondition fulfilled

Girl puppet: [BIG YAWN] Oh, I slept really well.

Boy puppet: [BIG YAWN] I'm really hungry. I'd like to have breakfast.

Girl puppet: Do you want cereal or toast?

Boy puppet: There's milk.

Target picture: cereal.

Foil picture: toast.

b. Precondition unfulfilled

Girl puppet: [BIG YAWN] Oh, I slept really well.

Boy puppet: I'm really hungry. I'd like to have breakfast.

Girl puppet: Do you want cereal or toast?

Boy puppet: The milk's all gone.

Foil picture: cereal.

Target picture: toast.

2.

a. Precondition fulfilled

Boy puppet: We need to look after our pets.

Girl puppet: Do you want to walk the dog or feed the cat?

Boy puppet: I've got a lead.

Target picture: dog.

Foil picture: cat.

b. Precondition unfulfilled

Boy puppet: We need to look after our pets.

Girl puppet: Do you want to walk the dog or feed the cat?

Boy puppet: The lead is broken.

Target picture: cat.

Foil picture: dog.

3.

a. Precondition fulfilled

Boy puppet: Let's get dressed.

Girl puppet: Do you want to wear your scarf or your sunglasses?

Boy puppet: It is hot and sunny outside.

Target picture: Sunglasses

Foil picture: Scarf

b. Precondition unfulfilled

Boy puppet: Let's get dressed.

Girl puppet: Do you want to wear your scarf or your sunglasses?

Boy puppet: It isn't hot and sunny outside.

Target picture: scarf.

Foil picture: sunglasses

4.

a. Precondition fulfilled

Boy puppet: Snacktime!

Girl puppet: What should we eat? Cake or biscuits?

Boy puppet: I've got a knife.

Target picture: cake

Foil picture: biscuits.

b. Precondition unfulfilled

Boy puppet: Snacktime!

Girl puppet: What should we eat? Cake or biscuits?

Boy puppet: We don't have a knife.

Target picture: biscuits.

Foil picture: cake.

5.

a. Precondition fulfilled

Boy puppet: Let’s have a healthy snack.

Girl puppet: What should we have? Orange juice or an apple?

Boy puppet: Our cups are clean.

Target picture: orange juice.

Foil picture: apple.

b. Precondition unfulfilled

Boy puppet: Let’s have a healthy snack.

Girl puppet: What should we have? Orange juice or an apple?

Boy puppet: Our cups are dirty.

Target picture: apple.

Foil picture: juice.

6.

a. Precondition fulfilled

Boy puppet: Let’s get ready for bed

Girl puppet: Are you going brush your teeth or your hair?

Boy puppet: We have toothpaste

Target picture: Toothbrush.

Foil picture: Hairbrush

b. Precondition unfulfilled

Boy puppet: Let's get ready for bed

Girl puppet: Are you going to brush your teeth or your hair?

Boy puppet: We don't have any toothpaste left.

Target picture: Hairbrush

Foil picture: Toothbrush.