# Pragmatic, linguistic and cognitive factors in young children's development of quantity, relevance and word learning inferences

**Elspeth WILSON[a]* and Napoleon KATSOS[b]**

**[a]Faculty of Education, University of Cambridge**

**[b]Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge**

***Corresponding author**

ep321@cam.ac.uk

Faculty of Education, University of Cambridge

184 Hills Road, Cambridge, CB2 8PQ, UK

**Acknowledgements**

Elspeth Wilson was funded by an ESRC PhD Studentship and ESRC Postdoctoral Fellowship. We are grateful to the families and schools who took part in the study, and Becky Brooks for assistance with data entry.

**Key words**

Pragmatic development

Relevance implicature

Quantity implicature

Word learning by exclusion

**Abstract**

To better understand the developmental trajectory of children's pragmatic development, studies which examine more than one type of implicature as well as associated linguistic and cognitive factors are required. We investigated three- to five-year-old English-speaking children's (N=71) performance in ad hoc quantity, scalar quantity and relevance implicatures, as well as word learning by exclusion inferences, using a sentence-to-picture-matching story-based task. Children's pragmatic abilities improved with age, with word learning by exclusion acquired first, followed by relevance and ad hoc quantity implicatures, and finally scalar quantity implicatures. In an exploratory analysis (with a subset of the data N=58), we found that structural language knowledge was a predictor of pragmatic performance (but no evidence for an association with socioeconomic status or Theory of Mind, controlling for structural language). We discuss reasons why this developmental pattern emerges with reference to linguistic and extra-linguistic properties of these inferences.

## Introduction

In developing communicative abilities, children have to learn how to make inferences to understand the meaning which the speaker intends to convey, beyond the literal meaning of what was uttered. On Grice's (1989) approach to pragmatics, both the speaker and hearer have expectations about *co-operative* communication, and assume that the other will be truthful, informative, relevant and conventional.

(1) What did you take from the fridge?
I took a strawberry.
(2) What would you like for breakfast?
I'll get the milk.

In (1), a QUANTITY IMPLICATURE, the hearer can infer that the speaker took *only* a strawberry from the fridge, because, had she taken more, she would have said so to provide a fully informative answer to the question. In (2), a RELEVANCE IMPLICATURE, in a context where the available alternatives are cereal or toast, the hearer can infer that the speaker wants cereal, because the world knowledge that milk is required for cereal makes this a relevant answer to the question. Over the past two decades a rich seam of research has been laid down on the interpretation, processing and development of implicatures within Experimental Pragmatics; the majority of studies have examined quantity implicatures, and only one type of implicature in isolation. The aim of the current study was to investigate the developmental trajectory of different implicature types in children aged three to five years, by comparing both quantity and relevance implicatures, as well as WORD LEARNING BY EXCLUSION, a key skill that develops early in child language development. We also wanted to explore other linguistic, cognitive and environmental factors which may play a role. We first present our motivations for this study, both empirical and theoretical, before briefly surveying existing findings on the development of each inference type and the contribution of other factors.

*Examining order of implicature acquisition*

Across different linguistic skills, including phonological, morphological and syntactic competence, the question of the relative order of acquisition of different constructions is a fundamental one: the emerging answers both increase our understanding of reliable patterns of child development, and also reveal more about the linguistic properties of the structures being studied. When it comes to pragmatic development, most studies either use global measures which include a wide variety of different pragmatic inferences (for a review see Matthews, Biney & Abbot-Smith, 2018), or focus on individual types of inference, such as ad hoc quantity implicatures. Although, as we shall see below, there is a growing body of evidence about children's implicature development (see too Table 1), comparing across different studies is problematic. Not only are there potentially significant task differences, even within a single paradigm like sentence-to-picture-matching, but studies are sampling different populations, with different languages, socioeconomic properties and educational experiences. This means that taking, for example, evidence for competence in relevance implicatures at three years from one study, and for competence in ad hoc quantity implicatures at four years from another study, cannot lead us to confidently infer that relevance inferences are acquired before ad hoc quantity inferences. In addition, there is a great heterogeneity and individual difference in the *rate* of acquisition across language skills (Kidd, Donnelly & Christiansen, 2018). Therefore, what is needed to better understand children's pragmatic development are more studies which investigate the relative acquisition of pragmatic skills within a single sample of children, together with other linguistic, cognitive and environmental factors which may play an important role, so that we can examine which skills co-develop with or are prerequisites for pragmatics.

*The role of relevance, the Question Under Discussion, and alternatives*

There are also theoretical reasons to examine different types of implicature together and potentially expect interesting differences in their development. On a CONSTRAINT-BASED view of pragmatic inference, which sits broadly within the Gricean tradition, hearers consider a whole range of sources of information in parallel in order to understand the speaker's meaning (Degen & Tanenhaus, 2014, 2019). One important factor is tracking what is relevant to the discourse, which is often characterised as the degree to which the utterance addresses the Question Under Discussion (e.g. Roberts, 2012). The QUESTION UNDER DISCUSSION (QUD) does not have to be an explicit question, as in examples (1) and (2), but can be implicit in the topic of discourse or the subgoal of conversation mutually agreed by the interlocutors. It is arguably important for all types of implicature, not just relevance (Degen & Tanenhaus, 2019).

In a relevance implicature, the hearer makes an elaborative inference, which forms a cohesive link based on world knowledge about what is typically the case between what is said and what is implicated (Cummings, 2005). In (2), the hearer can infer that what the speaker said is relevant by virtue of the fact (world knowledge) that milk is typically necessary for one of the breakfast options, namely cereal. In a quantity implicature, the hearer generates stronger alternatives, such as a strawberry and an apple in (1) – arguably involving elaborative inference as well, forming a cohesive link between what was said and the situation, based on knowledge of the situation or of linguistic scales – and crucially activated and constrained by the QUD (Benz & Jasinskaja, 2017). These relevant alternatives are negated to arrive at the intended meaning, *only a strawberry*. Indeed, there is empirical evidence that adult hearers do not derive an implicature when it is not relevant to the QUD (e.g. Zondervan, Meroni  & Gualmini, 2008) and that a challenge for children in understanding scalar implicatures is tracking the QUD and generating relevant alternatives (Hurewitz, Papafragou, Gleitman & Gelman, 2006; Skordos & Papafragou, 2016). For example, in (3), the explicit question is informatively answered by the speaker if she means *I took at least a strawberry*; whether or not she took other items is not relevant.

(3) Did you get fruit from the fridge?
I took a strawberry.

The acquisitional challenge for children on a constraint-based view, therefore, involves not just acquiring the inferential process, but also learning to recognise and weight constraints appropriately for a situation. In particular, they have to learn to track the QUD and apply this knowledge within the inferential process. For relevance implicatures this means forming an elaborative inference between what the speaker says and how it relates to the QUD; for quantity, it *additionally* means negating the generated relevant alternatives. Thus one would expect at the very least relevance and quantity implicatures to emerge together in development, and quite probably relevance before quantity.

*Acquisition of quantity implicatures*

To date the vast majority of studies on children's implicature development have focussed on quantity implicatures. A range of measures have been employed, most notably Truth Value or Acceptability Judgement Tasks, and sentence-to-picture-matching tasks. For the sake of comparison, here we will concentrate on findings from picture-matching tasks – see Table 1 for a review of picture-matching studies (for more general reviews see Papafragou & Skordos, 2016; Wilson & Katsos, 2020). Picture-matching tasks have been argued to be more direct measures of children's interpretation of implicature-triggering sentences: alternatives are presented visually and children are asked only to choose a picture. In contrast, judgement tasks may rely on metalinguistic skills, often asking children to explain their decision, and they might be susceptible to a 'yes' bias or pragmatic tolerance (Katsos & Bishop, 2011; Veenstra & Katsos, 2018).

Considering existing studies, it seems that children learn to derive AD HOC QUANTITY IMPLICATURES, as in (1), where the alternatives are contextually salient, from three years (Grosse, Schulze, Noveck, Tomasello & Katsos, under review; Stiller, Goodman & Frank, 2015; Yoon & Frank, 2019) although cross-linguistically there might be considerable variation (e.g. Fortier, Kellier, Flecha & Frank, under review; Zhao, Jie, Frank & Zhou, in press). For SCALAR IMPLICATURES with the quantifier *some*, children display adult-like or above-chance rates of implicatures later, from around five years or even older (Cremers, Kane, Tieu, Kennedy, Sudo, Folli & Romoli, 2018; Hurewitz et al., 2006; Nordmeyer Yoon & Frank, 2016). The three studies which directly compare ad hoc and scalar inferences confirm this difference in developmental trajectory: Foppolo, Mazzagio, Panzeri and Surian (2020) found a difference between ad hocs and scalars in younger Italian-speaking children (aged 3;8-6;0) but not older children (aged 6;0-9;2); Grosse et al (under review) showed that German-speaking five-year-olds perform better than three-year-olds with scalar implicatures, while for ad hocs there is a similar pattern but both groups are above chance; and in American English-speaking four-year-olds, Horowitz, Schneider and Frank (2018) observed significantly worse performance on scalar implicature trials than on ad hocs, for which performance was approaching ceiling.

These studies are typically designed to test or have implications for an ongoing theoretical debate about the nature of scalar versus ad hoc quantity implicatures and their development. On a lexical scales account, scalar implicatures are distinct in that they rely on lexically encoded scales, such as <all, some> (Hirschberg, 1991), and children's difficulty stems from not having acquired or having difficulty accessing these scales (e.g. Barner, Brooks & Bale, 2011; Foppolo, Guasti & Chierchia, 2012). On alternative accounts, more general pragmatic factors might be driving differences, such as expectations of informativeness (e.g. Katsos & Bishop, 2011; Noveck, 2001; Papafragou & Skordos, 2016). For instance, Foppolo et al (2020) set out opposing lexical and pragmatic accounts, as well as "processing" accounts, which tend to implicate "processing resources" or more specific capabilities like developing Executive Functions (e.g. Pouscoulous, Noveck, Politzer & Bastide, 2007), and propose that only lexicalist approaches predict a difference between scalar and ad hoc implicatures, as "pragmatic factors" should affect both types equally. However, it is not difficult to see how pragmatic factors could account for differences as well: for example, there might be contextual factors which make alternatives more relevant and accessible in the ad hoc case, or more low-level factors like the simpler visual scene for ad hoc implicatures. Horowitz, Schneider and Frank (2018), meanwhile, contrast the lexical account (an Alternatives Hypothesis) with a more specific hypothesis of difficulties with quantifiers (see too Hurewitz et al., 2006). While they do provide evidence that children have difficulties with quantifiers (there is no trial order effect, contra the lexical account, and there is a relationship between implicature rates and knowledge of quantifiers), to properly test the quantifier difficulties hypothesis in comparison to the lexical account, comparison with other scales is surely required, and there may be other reasons while other scales are more or less challenging than those with quantifiers (e.g. epistemic modals <must, may> are likely to be acquired still later, Ozturk & Papafragou, 2015). In other words, trying to reduce the difference between scalar implicatures with *some* and ad hocs to a single factor is problematic. Thus, we consider it more informative to approach the acquisition of implicatures within a more holistic constraint-based view, and compare ad hoc and scalar quantity implicatures with relevance implicatures. That said, both the range of current theories and existing comparative data lead us to expect ad hoc quantity implicatures to emerge before scalars in this study too.

*Acquisition of relevance inferences*

The study of the development of relevance implicatures stretches back several decades, thanks to early attention on a particular instantiation, the indirect request (e.g. Bernicot & Legros, 1987). As with quantity implicatures, early studies suggested relatively late acquisition, aged eight years and over, in all likelihood due to the metalinguistic nature of the task, asking children to explain what the speaker meant (e.g. Bucciarelli, Colle & Bara, 2003; de Villiers, de Villiers, Coles-White & Carpenter, 2009). More recently, there have been, to our knowledge, three investigations of children's understanding of relevance implicatures using picture-matching tasks. Tribushinina (2012), Schulze, Grassmann and Tomasello (2013), and Schulze, Endesfelder Quick, Dampe and Gaum (2020) all present evidence that they are available from three years, especially in simple cases such as (4), but also in the case of (2):

(4) Should [child] give you the elephant?
I like elephants / I don't like elephants.

Only one previous study has compared relevance and quantity implicatures: Verbuk and Schultz (2010) compared implicatures with part-whole scales with indirect requests, and did not find evidence for a difference between them. However, there were a number of issues with the design: the wide age-range of children in one group for analysis (5;1-8;1); the heavily metalinguistic task (requiring children to explain their picture choice in order to score as correct); and the inclusion of a 'non-verbal' condition, which could affect expectations about the speaker and task.

*Word learning by exclusion*

In this study, as well as testing children on quantity and relevance implicatures, we included word learning by exclusion as a comparison (we use this as a general term to avoid association with a particular theory such as Mutual Exclusivity bias, Markman et al., 2003). Word learning by exclusion is a robust phenomenon, whereby children presented with a familiar object and a novel object will choose the novel object for a novel label. On many accounts, this is a result of reasoning by exclusion that the label does not refer to the familiar object (for which they already know the label) and so must refer to the novel object (e.g. Clark, 1990, Halberda, 2003). This strategy is evident even in infancy, from the second year of life, and strengthens over development (e.g. Graham, Poulin-Dubois & Baker, 1998; Halberda, 2003; Markman et al., 2003). Some have suggested that it is a pragmatic strategy, with striking parallels to implicature derivation (e.g. Barner, Brooks & Bale, 2011; Clark, 1990; Katsos & Bishop, 2011; Stiller, Goodman & Frank, 2015). On this account, the child can reason that the speaker *intends* to refer to the novel object with the novel label, because, had she wanted to refer to the familiar object, she would have used its label, being co-operative, conventional and informative. Arguably, the need to track the QUD is diminished in this case, though, as the use of the novel label is such a strong cue that an inference is required. Therefore, word learning by exclusion is an interesting comparison to relevance and quantity implicatures, as it involves some of the same reasoning as for quantity implicatures. Even on a minimal account of word learning – without full reference to speaker intentions – reasoning by exclusion (negating the alternative) is common to both, but overall it is a much simpler inference, which we would therefore expect it to be in place early.

*Table 1 Review of previous literature of implicature development with studies using a picture-matching task*

| Study | Implicature type | Other inferences / measures | Ages and N | Trials for critical condition | Language | Main findings |
|---|---|---|---|---|---|---|
| **Bernicot, Laval & Chaminaud, 2007** | Relevance | Indirect request, Idiom, Sarcasm | 6;0-7;11 (N=20); 8;2-9;9 (N=20); 10;3-11;3, (N=20) | 4 | French | Best performance for Relevance (followed by indirect request, idiom and sarcasm), robustly present at 8 years. |
| **Cremers, Kane, Tieu, Kennedy, Sudo, Folli, & Romoli, 2018** | Scalar | Temporal inference; adverbial modifier under negation | 4;0-5;11 (N=38) | 4 | UK English (Northern Ireland) | Least adult-like for scalar implicatures. |
| **Foppolo, Mazzagio, Panzeri & Surian, 2020** | Scalar and ad hoc | Comparison of TVJT and picture-matching for SIs Structural language, ToM, nonverbal IQ | 3;8-6;0 (N=75), 6;1-9;2 (N=66) | 4 | Italian | Difference between ad hocs and SIs for younger but not older children (better with ad hocs). Correlation with structural language. |
| **Fortier, Kellier, Flecha & Frank, under review** | Ad hoc | | 4-6 (N=11); 6-8 (N=30); 8-10 (N=35) | 2 | Shipibo-Konibo | 8-10 year olds understand ad hocs, in a culture with a more holistic orientation. |
| **Grosse, Schulze, Noveck, Tomasello & Katsos, under review** | Scalar and ad hoc | Under-informative condition Between group: control before critical, and vice versa | 3;2-3;8 (N =24), 5;0-5;5 (N =24) | 3 | German | 3-year-olds can derive ad hoc implicatures; difference between 3- and 5-year-olds for SIs. |

| Study | Other | | Ages and N | Trials for critical condition | Language | Main findings |
|---|---|---|---|---|---|---|
| **Horowitz, Schneider & Frank, 2017** | Scalar and ad hoc | 'none' control. Inhibitory control; quantifier knowledge | 4;0-4;6 (N=24), 4;7-4;11 (N=24) (Exp 1) 3;0-3;6 (N=12/18), 3;7-3;11 (N=13/18), 4;0-4;6 (N=14/18), 4;7-4;11 (N=12/18) (Exp 2/3 SI only) | 4 (exp 1); 6 (exps 2 and 3) | American English | Developmental trend with competence increasing with age. Correlation between SIs and 'none' trials. No correlation with inhibition, controlling for age. |
| **Hurewitz, Papafragou, Gleitman & Gelman, 2006** | Scalar | Numerals | 2;9-3;6 (N=12), 3;7-4;0 (N=12) | 3 | American English | Adult-like performance from both age groups for exact interpretation of numerals, but not SIs. |
| **Katsos & Bishop, 2011** | Scalar and ad hoc | | 5;1-6;1 (N=15) (Exp 3) | 6 | UK English | Adult-like performance for ad hocs and SIs. |
| **Miller, Schmitt, Chang & Munn, 2005** | Scalar | | 3;6-5;10 (N=16) (Exp 2; between subjects) | 4 | ? American English | Effect of prosody (contrast stress): children are adult-like where 'some' is stressed. |
| **Nordmeyer, Yoon & Frank, 2016** | Ad hoc | Inhibition; negation. Reaction times | 4 year-olds (N=22), 5 year-olds (N=19), 6 (N=25) | 30 | American English | Developmental trend (implicatures increasing with age). No evidence of a relationship between inhibition and performance on the negation or implicature tasks. |
| **Schulze, Grassmann & Tomasello, 2013** | Relevance | | 2;10-3;1 (N=20) and 3;10-4;1 (N=20 (Exp 3) | 4 | German | Simple relevance inferences derived by three-year-olds. |

| Study | Other | | Ages and N | Trials for critical condition | Language | Main findings |
|---|---|---|---|---|---|---|
| **Stiller, Goodman & Frank, 2015** | Ad hoc | | 2;0-2;11 (N=49/ 3;0-3;11 (N=50/48), 4;0-4;11 (N=48/49) (original / replication; between subject) | 4 | American English | Simple ad hoc implicatures in four-year-olds and some three-year-olds (but not two-year-olds) |
| **Tribushinina, 2012** | Relevance | | 3;1-3;11 (N=20) and 5;1-5;11 (N=20) (Exp 1) | 9*4 | Dutch | Simple relevance inferences derived by three-year-olds. |
| **Yoon & Frank, 2019** | Ad hoc | Double vs single object control; varied number of distractors Reaction Times | 2 year-olds (N=25/25), 3 year-olds (N=29/30), 4 year-olds (N=26/26), 5 year-olds (N=19) (original / replication) | 4 | American English | Developmental trend (implicatures increasing with age). For youngest children, effect of distractors: more distractor features, worse performance. |
| **Zhao, Jie, Frank, & Zhou, in press** | Scalar and ad hoc | Numerals; two different ways of expressing ad hocs. Between subject design. | 4 yos (N=61), 5 yos (N=61), 6 yos (N=40), 7 yos (N=21), 8 yos (N=42) | 12 | Mandarin | Four-year-olds derived ad hoc inferences (and numerals) but only children aged six and over derived scalar implicatures. |

*Linguistics, cognitive and environmental factors in pragmatic development*

A constraint-based view of implicature interpretation, in which the hearer has to take into account a number of linguistic and contextual pieces of information, would naturally lead us to expect that children's pragmatic development is associated with other linguistic, cognitive and environmental factors. In this study we therefore also explore associations between children's performance with implicatures, and their structural language abilities (vocabulary and grammar), socioeconomic background, and THEORY OF MIND. Few developmental pragmatics studies consider how such factors might interact with the experimental manipulation of the task, despite plausible reasons for their importance.

Firstly, there are two ways that structural language could be related to implicatures in development: specifically to implicature-triggering utterances, and generally to pragmatic development. For any particular utterance, the vocabulary, grammatical constructions and prosody used by the speaker will contribute to whether the hearer derives an implicature. As already mentioned, for some implicatures, like scalars, there may be particular lexical items which present a learning challenge for children. In addition, there may be a more general relationship between total vocabulary and grammar knowledge and pragmatic skills: one might expect that the more structural language children have acquired, the more possibility they have to access some meaning in context, practice pragmatic skills, and learn how expectations of co-operativity function in conversation. Conversely, on accounts of language acquisition which view pragmatic skills as fundamental, better pragmatic abilities would facilitate lexical and grammatical acquisition (Bohn & Frank, 2019; Tomasello, 2003). Foppolo et al (2020) and Antoniou and Katsos (2017) both found that structural language was a predictor of implicature performance, in three- to nine-year-olds and six- to nine-year-olds respectively.

Secondly, socioeconomic status (SES) is widely reported to be connected to language development, especially vocabulary (e.g. Hoff, 2006), although problems with test measures favouring middle-class children have been noted. The reasons for a relationship are likely to be complex, and, as Pace, Luo, Hirsh-Pasek & Golinkoff (2017) point out, have received less attention from a psycholinguistic approach; they may, though, include differences in processing, in input, and in available learning materials. Within experimental pragmatics, samples are typically assumed to be fairly homogenous, though Antoniou and Katsos (2017), Antoniou, Veenstra, Kissine and Katsos (2020), and Schulze, Endesfelder Quick, Gampe & Daum (2020) did measure SES and did not find evidence for a correlation.

Thirdly, and very briefly given significant theoretical and empirical debate, Theory of Mind – the ability to represent and reason about others' beliefs and mental states – is a central component to a Gricean approach to pragmatics, in that the hearer recognises the communicative intentions of the speaker, and assumes that they are truthful and knowledgeable on the relevant matter, unless there is evidence to the contrary. Indeed, reasoning about the speaker's epistemic state is an integral part of the pragmatic inferencing which the hearer engages in to arrive at the speaker's intended meaning. On a constraint-based view, the speaker's epistemic state is likewise one of the many factors considered in inferencing (Degen & Tanenhaus, 2019), and, indeed, there is evidence that adult speakers, at least, are able to take the speaker's knowledge into account and derive or not derive an implicature appropriately (e.g. Breheny, Ferguson & Katsos, 2013). There are, though, alternative views of pragmatics, which propose that different strategies may be available for inferencing, which take into consideration the speaker's knowledge more or less (e.g. Andrés-Roqueta & Katsos, 2017; Kissine, 2016). In children, the evidence is more mixed, with some studies finding that they are able to reason about the speaker's knowledge in implicature inferencing (Kampa & Papafragou, 2020), and others suggestive of children deriving implicatures before they can integrate the speaker's epistemic state (e.g. Barner, Hochstein, Rubenstein & Bale, 2018).

*The current study*

To take stock: empirical investigations so far have provided evidence for the early acquisition of relevance implicatures, and, separately, ad hoc quantity implicatures, which seem to emerge before scalar implicatures. Word learning by exclusion, which could be a simple pragmatic inference, is likely to be in place even earlier. We have also argued that developing an understanding of relevance and ability to track the QUD for elaborative inferencing is important for both relevance and quantity implicatures. In addition, quantity implicatures require generating and negating relevant alternatives, an inference plausibly similar to reasoning by exclusion in word learning. Thus, all else being equal, one might expect word learning by exclusion to be grasped first, followed by relevance implicatures, and finally quantity implicatures. Additional semantic or pragmatic challenges in the acquisition of quantifiers – and possibly other scales – also mean that scalar quantity implicatures are likely to be acquired after ad hocs. It is also likely that children's implicature development is associated with other aspects of their linguistic and cognitive development.

In this study, we aimed to investigate the developmental trajectory of implicatures, and explore some of the factors that may be associated with this development. We conducted a story-based picture-matching task with British English-speaking three- to five-year-olds to test their ability to derive relevance, ad hoc and scalar quantity implicatures and do word learning by exclusion. We therefore extend the findings of previous studies, by directly comparing the developmental trajectories of both relevance and quantity implicatures in a single experiment, across three age groups (three-, four- and five-year-olds). We also build on other child-friendly picture-matching tasks by designing an interactive 'story', in which there is an explicit QUD in each trial before the critical utterance: children had to choose which of two pictures matched what the puppet-protagonist said he did, and put it on their story board. In addition, we add an exploratory analysis of the association of structural language, SES and Theory of Mind (using standard measures for each) with implicature interpretation.

## Method

We designed a picture-matching task, inspired particularly by Stiller, Goodman and Frank's (2015), Grosse et al's (under review) and Schulze, Grassmann and Tomasello's (2013) studies, which were available when we were commenced this study (in pre-print form or as conference proceedings). However, we created a story-based task to make it more naturalistic and child-friendly, and because a rich discourse context has been suggested to facilitate children's inference-making (Hurewitz et al., 2006). We also added a word learning by exclusion condition, based on one standard version of the task (Markman & Wachtel, 1988). The aim was to test children's derivation of quantity, relevance and word learning inferences in a supportive context, as well as to gather correlational measures of structural language knowledge, SES and Theory of Mind, using standard tests. The full protocol and stimuli can be accessed at osf.io/75uv4/.

### *Participants*

Participants aged 2;8–5;11 were recruited from Foundation classes in two local primary schools in UK, from nurseries and preschools, and from personal contacts. Parents provided consent for children to participate, via an opt-in or opt-out procedure depending on the setting's policy. The study received approval from the University of Cambridge Psychology Ethics Committee.

In total, 135 children were recruited. Some participants were excluded from analysis because of too noisy an environment (N = 2), failure to finish the task (N = 8), or declared developmental disorder (N = 2). In addition, some children were recruited (given parental consent) but chose not to take part in the study or were absent from school or nursery at the time of testing (N = 17). We also collected information on the languages spoken by the children, and for this study present results only for monolingual children, excluding 35 bilingual children who also completed the tasks: the question of the effect of multilingual acquisition on pragmatic skills is an interesting one which merits investigation on its own terms (Antoniou et al., 2020; Antoniou & Katsos, 2017). The responses from 71 monolingual children were included in the final analysis – see Table 2. For the exploratory analysis of the association of structural language, SES and Theory of Mind, we included only those children who had completed all tests and the parental background questionnaire, which left 58 children.

In addition, 28 children were recruited from two other local primary schools for pretesting and piloting of this study. The adult control group (N=15) were recruited via Prolific Academic, an online recruitment platform for research.

*Table 2 Information about participants*

| Age group | Participants | Females | Mean age (months) | Standard Deviation |
|---|---|---|---|---|
| 2;8–3;11 | 25 | 13 | 40.9 | 4.2 |
| 4;0–4;11 | 25 | 11 | 54.0 | 3.6 |
| 5;0–5;11 | 21 | 10 | 63.8 | 2.7 |
| Total | 71 | 34 | | |

*Table 3 Information about participants for exploratory analysis of subset of participants*

| Age group | Participants | Females | Mean age (months) | Standard Deviation |
|---|---|---|---|---|
| 2;8–3;11 | 17 | 10 | 40.4 | 4.2 |
| 4;0–4;11 | 21 | 10 | 54.7 | 3.4 |
| 5;0–5;11 | 20 | 9 | 63.7 | 2.8 |
| Total | 58 | 29 | | |

***Stimuli***

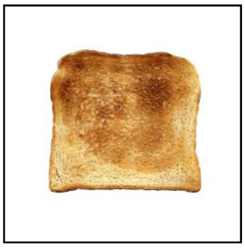
The picture-matching task was presented as physical story books in a small folder, with laminated pictures attached by magnets so that they could easily be removed by participants and placed on their magnetic ‘story board’. Each item consisted of a) a context sentence, b) a question, and c) the critical or control utterance (an answer to the question). The context sentence and question were uttered by the experimenter and accompanied by a single picture in the book; the critical utterance was given by a puppet (the protagonist in the story) with pre-recorded voice and accompanied by two pictures side by side in the book. The puppet was always a male, and the experimenter a female; having pre-recorded utterances has the advantage that all children hear the critical utterance in the same way. Pictures in the picture-book were photographs sourced from the BOSS Database (Brodeur, Dionne-Dostie, Montreuil & Lepage, 2010), Pixabay, a database of CC0 licensed images (Braxmeier & Steinberger, 2017), or via an online search for images labelled for non-commercial reuse. They were edited using GIMP (Kimball, Mattis & The Gimp Development Team, 2016).

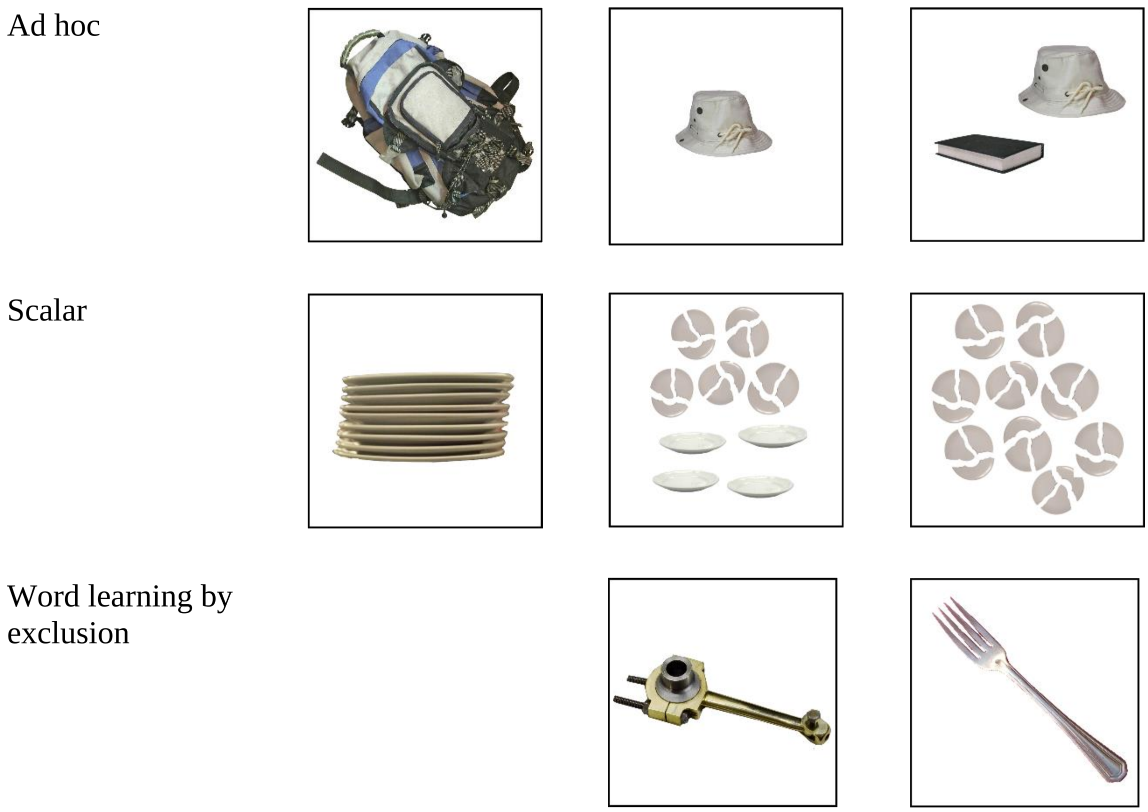
We tested four inference types – relevance, ad hoc quantity, scalar quantity and word learning by exclusion – in two conditions: critical (where an implicature was intended by the speaker) and control (where no implicature was intended by the speaker and the answer to the QUD was addressed by the literal meaning of the utterance) – see Tables 4 and 5 for examples. Relevance, ad hoc quantity and scalar quantity were mixed across 4 stories, each with 6 trials, one in critical and one in control condition for each implicature type; children therefore heard 4 trials for each condition for each implicature type overall (32 trials). The word learning by exclusion trials (again, four in critical and four in control conditions) were always presented in a block as the final story: this was so that the puppet's use of novel words did not affect the participant's perception of him as a cooperative speaker. For word learning, there was also only a minimal context phase (e.g. 'I went into the shop and…') so that the discourse did not provide any competing cues to the intended referent.

*Table 4 Experiment example items*

| | Context sentence and question | Critical utterance | Control utterance | Critical picture choice | Control picture choice |
|---|---|---|---|---|---|
| Relevance | It was breakfast time. Bob's dad asked, 'What would you like for breakfast?' | And I said, 'I'll get the milk.' | And I said, 'I'd like toast.' | Cereal | Toast |
| Ad hoc | Bob was getting ready for school. His mum asked, 'What have you packed in your bag?' | And I said, 'I packed a hat.' | And I said, 'I packed a book and a hat.' | Hat | Book and hat |
| Scalar | Bob made a crash in the kitchen. His dad asked, 'What have you done with the pile of plates?' | And I said, 'I broke some of the plates.' | And I said, 'I broke all of the plates.' | Some (not all) plates broken | All plates broken |
| Word learning by exclusion | He went further inside and… | 'I picked a dax.' | 'I picked a fork.' | Novel object | Fork |

*Table 5 Examples of visual stimuli for each inference type and condition*

| | Context picture | Critical picture choice | Control picture choice |
|---|---|---|---|
| Relevance |  |  |  |




For relevance, the question was always about an activity or object the puppet wanted, e.g. ‘What would you like for breakfast?’, and the puppet answered either directly (in the control condition), e.g. *I’d like toast*, or indirectly, triggering a relevance implicature: *I’ll get the milk*. The two pictures to choose from showed a different item that represented the activity (e.g. eating cereal or toast). In the control condition, only one of the pictures depicted the utterance’s meaning; in the critical condition, on the literal meaning, neither picture seemed relevant, so the choice was ambiguous; on the implicated meaning, one of the pictures matched. The items were devised via pre-tests to make sure that children knew the association between the relevant object (e.g. milk) and activity (e.g. eating cereal).

For ad hoc quantity, the puppet said, for instance, *I packed a hat* in the critical condition, and *I packed a book and a hat*, in the control condition. One picture showed a hat, and the other a hat and a book, so that in the critical condition both were semantic matches for the utterance, but only one matched the implicature, ‘I packed only a hat’. Likewise, in the scalar quantity condition, the puppet said, for example, *I broke some of the plates* (critical condition) or *I broke all of the plates* (control condition), and the pictures showed either some (but not all) or all of the plates broken. We used *some of* rather than *some*, in line with other developmental studies (e.g. Horowitz et al., 2018) and as it is known to facilitate scalar implicature derivation (Degen & Tanenhaus, 2014). In addition, all pictures displayed a number of objects well above the subitizing range, so that numerals were not competing alternatives.

Finally, for word learning by exclusion, the puppet said *I picked a dax* or *I picked a fork*, and one picture displayed a novel object, while the other a familiar object for the familiar label. The novel words were taken from other studies and consisted of 4 monosyllabic and 4 bisyllabic words with English phonotactics (Barner & Snedeker, 2008; Diesendruck et al., 2003; Diesendruck & Markson, 2001; Halberda, 2003). The novel objects were pretested with adults to make sure that a majority of adults did not recognise them. Known items were also pretested with children to make sure they were clearly identifiable.

Participants saw only the critical or control condition for any one item; items within each story were rotated across participant lists, and arranged such that no two of any utterance type appeared one after the other and no more than two of the critical or control condition appeared together; and the first four stories themselves were rotated. This counter-balanced design produced 48 lists. In addition, across lists, the position of the pictures (left or right) was counter-balanced.

### *Procedure*

Children were tested individually in their school, nursery or home. They sat at a table with the picture-book in front of them on a book rest, and the magnetic story board on the table in front. The experimenter sat to the side, so that the puppet, picture book and computer (to play the pre-recorded utterances) could all easily be operated. After the experimenter explained the activity, there was a warm-up phase with a short story consisting of four unambiguous trials; then the experimenter asked the children whether they would like to go on to the next story. During the context sentence and question, the experimenter looked between the children and pictures to establish join attention, but during the critical utterance, she looked at the puppet so that the children's choice would not be influenced by the experimenter's gaze. If the child was unsure and asked the experimenter for help, the experimenter looked straight at the children, and encouraged them to *choose the picture that goes with the story*. If children tried to choose both pictures, the experimenter gave a reminder to choose just one. At the end of the session, which took about 20 minutes, children were given a sticker as a thank you. Their responses were recorded as a photograph of the story boards showing their selected pictures. The adult control group completed an online version of the task, using Qualtrics (*Qualtrics*, 2016)

In a second testing session, children were given the structural language and Theory of Mind measures. The British Picture Vocabulary Scale-3 (Dunn, Dunn, Sewell, Styles, Brzyska, Shamsan & Burge, 2009) was used to test receptive vocabulary, and a reduced version of the Test of Receptive Grammar II (Bishop, 2003) was used to test grammar, with 20 items instead of 80, one from each block of the full TROG II (this reduced testing time for the children; the abbreviated version tested each of the twenty sentence types of the full TROG II but with a single trial per sentence type). To measure Theory of Mind, two false belief tasks were used: the Change of Location, or Sally-Anne, task (Baron-Cohen, Leslie & Frith, 1985; Wimmer & Perner, 1983), which was acted out with puppets and props, and the Unexpected Contents task (Perner et al., 1987). Parents were asked to fill in a background questionnaire which asked about language exposure (based on the Alberta Language Environment Questionnaire, Paradis, 2011), and about SES via the Family Affluence Scale (Boyce, Tosheim, Currie & Zambon, 2006) and parental education.

## Results

### Coding

For the implicature task, the picture choices were coded as matching the implicature or control utterance (e.g. the picture with one object or with two, for ad hocs), and this was then converted to 'correct' or 'incorrect' depending on the condition for each item. For the BPVS-3 and TROG II, raw scores were calculated and used in analyses. In the Theory of Mind tasks, children could score a maximum of three: one in the Change of Location task, and two in the Unexpected Contents task. From the background questionnaire, SES scores for each component (Family Affluence Scale, and parental education) were first centred and scaled, and then a mean calculated for each participant combining them, so that the two were equally weighted.

### Analysis

There is a clear developmental trend for ad hoc, scalar and relevance implicatures, which improve with age, but not for word learning by exclusion inferences which are already approaching ceiling in the youngest group. Children also perform worse with scalar trials compared to other inference types. Accuracy on control trials is always better than on critical inference trials. This overall pattern is consistent with previous research (e.g. Foppolo et al., 2020; Grosse et al., under review; Horowitz et al., 2018), which suggests the paradigm is an appropriate measure for implicature comprehension. The proportion of correct responses for all inference types, condition and age is shown in Figure 1. Adults were at ceiling (over 95% correct) across all trial types (Figure 2) and are not included in further analysis.

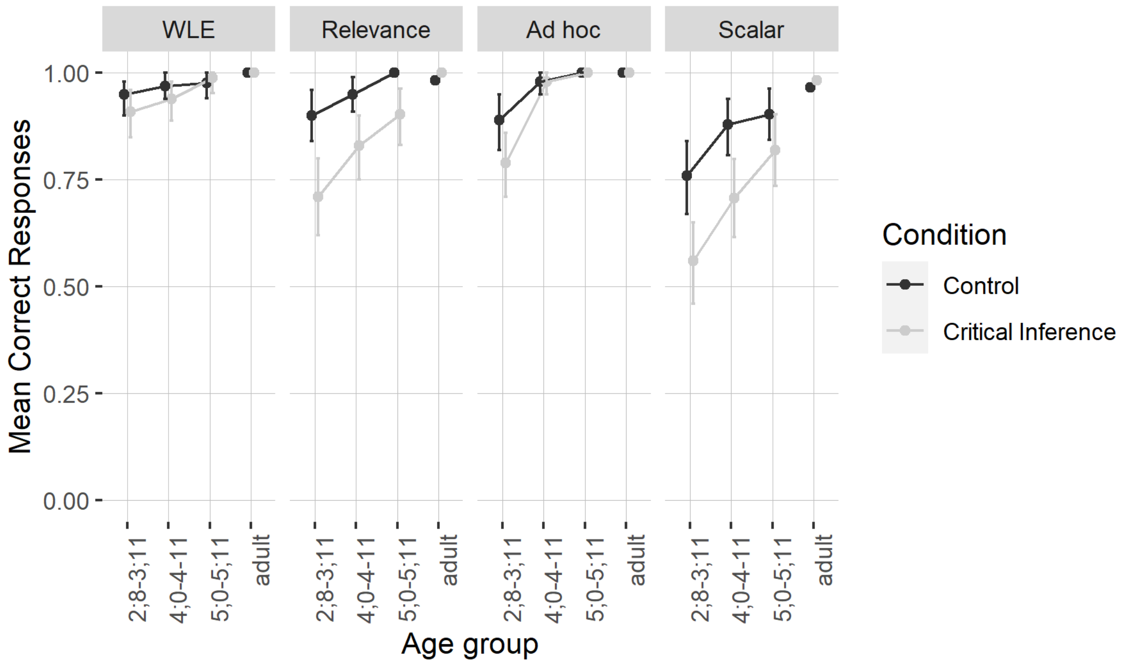



*Figure 1 Proportion of correct responses for word learning by exclusion (W), relevance (R), ad hoc quantity (A) and scalar quantity (S) inferences. Error bars show bootstrapped 95% confidence intervals for between-subject comparison*

*Table 6 Proportion of correct responses by condition, inference type and age group*

| Age group | Trial type | Word learning | Relevance | Ad hoc | Scalar |
|---|---|---|---|---|---|
| 2;8–3;11 | Critical | 0.91 | 0.71 | 0.79 | 0.56 |
| | Control | 0.95 | 0.9 | 0.89 | 0.76 |
| 4;0–4;11 | Critical | 0.94 | 0.83 | 0.98 | 0.71 |
| | Control | 0.97 | 0.95 | 0.98 | 0.88 |
| 5;0–5;11 | Critical | 0.99 | 0.9 | 1 | 0.82 |
| | Control | 0.98 | 1 | 1 | 0.9 |

To examine the developmental trajectories of the different inference types, we ran a mixed-effects logistic regression model, using the *lme4* package in R (Bates, Mächler, Bolker & Walker, 2015; R Core Team, 2016). The maximal model with all random effects would not converge, and so, following Barr, Levy, Scheepers and Tily (2013), we fitted separate models with by-item and by-subject random effects, and here present the more conservative model with by-item random effects. A model with condition, inference type and age group as fixed effects (with sum coding), and item by condition, age group and story order, indicates a main effect of condition, such that the control condition is higher than the grand mean ($\beta = .53$, $p < .001$); a main effect of scalar inference type, such that the scalar type is lower than the grand mean ($\beta = -1.25$, $p < .001$); and an effect of the age group 2;8–3;11, such that it is lower than the grand mean ($\beta = -1.02$, $p < .001$) – see Table 7.

*Table 7 Mixed-effects logistic regression model: Response ~ Condition + Type + Age group + (1 + Condition + Age group + Block | Item), using glmer, family = binomial, optimizer = bobyqa, backward difference coding*

| | Estimate | SE | z | p |
|---|---|---|---|---|
| Intercept | 2.8 | .16 | 17.1 | < .001 |
| Control | .53 | .13 | 4.19 | < .001 |
| Ad Hoc | .37 | .22 | 1.69 | .08 |
| Relevance | -.14 | .19 | -.78 | .44 |
| Scalar | -1.25 | .12 | -6.43 | < .001 |
| 2;8–3;11 | -1.02 | .16 | -6.34 | < .001 |
| 4;0–4;11 | .015 | .14 | .11 | .92 |

To test in particular whether the order of acquisition of inference types was as we predicted, we fitted a second, theoretically-informed model, with the factors coded with successive difference contrasts, so that each level within a factor is compared to the previous one. The comparison order was control–critical for condition, word learning–relevance–ad hoc–scalar for type, and decreasing age groups. This indicates a difference in condition, such that the rate of correct responses for critical trials is lower than for control trials ($\beta = -1.06$, $p < .001$); a difference between relevance and word learning by exclusion, such that rate of correct response is lower for relevance ($\beta = -1.18$, $p = .0024$); no difference between relevance and ad hocs; but a difference between ad hocs and scalars, with scalars lower than ad hocs ($\beta = -1.63$, $p < .001$). There is also a difference between age groups: 4-year-olds perform worse overall than 5-year-olds ($\beta = -.99$, $p = .0024$), and 3-year-olds worse than 4-year-olds ($\beta = -1.04$, $p < .001$) – Table 8.

*Table 8 Mixed-effects logistic regression model: Response ~ Condition + Type + Age group + (1 + Condition + Age group + Block | Item), using glmer, family = binomial, optimizer = bobyqa, backward difference coding*

| | Estimate | SE | z | p |
|---|---|---|---|---|
| Intercept | 2.80 | .16 | 17.1 | < .001 |
| Critical – Control | -1.06 | .25 | -4.20 | < .001 |
| R – WLE | -1.18 | .39 | -3.03 | .0024 |
| AH – R | .052 | .32 | 1.64 | .10 |
| S - AH | -1.63 | .33 | -4.89 | < .001 |
| 4;0–4;11 – 5;0–5;11 | -.99 | .33 | -3.04 | .0024 |
| 2;8–3;11 – 4;0–4;11 | -1.04 | .20 | -5.05 | < .001 |

In a post hoc exploration of the data, we first examined the distribution of scores, as previous studies have observed a bimodal distribution particularly for scalar implicatures, such that children tend to consistently derive or not derive *some but not all* inferences (Foppolo et al., 2020; Horowitz et al., 2018). In our study, though, histograms suggest no evidence for a bimodal distribution for any age group, and in particular for the youngest age group with scalars, the modal value is .5, and for all other ages the distribution is skewed towards ceiling performance – Figure 3. Secondly, we considered whether there were any practice effects, such that children's performance improved over the task, through model comparison, with and without story order – this was for relevance, ad hoc and scalar inferences only across the first four stories, as word learning trials were always presented in the final story. Overall, there was no effect of adding story order to the model – either in general or considering only scalar inferences (Tables 9 and 10). Finally, we looked at the relationship between performance for relevance and quantity implicatures by conducting partial correlations for scores in the critical condition, controlling for language (the control condition) and age in months. For scalar implicatures, there is a significant positive relationship of small to moderate size with relevance ($\tau$ = .21, z = 2.5, p = .012); for ad hocs, there is no significantly positive relationship ($\tau$ = .078, z = .94, p = .35).

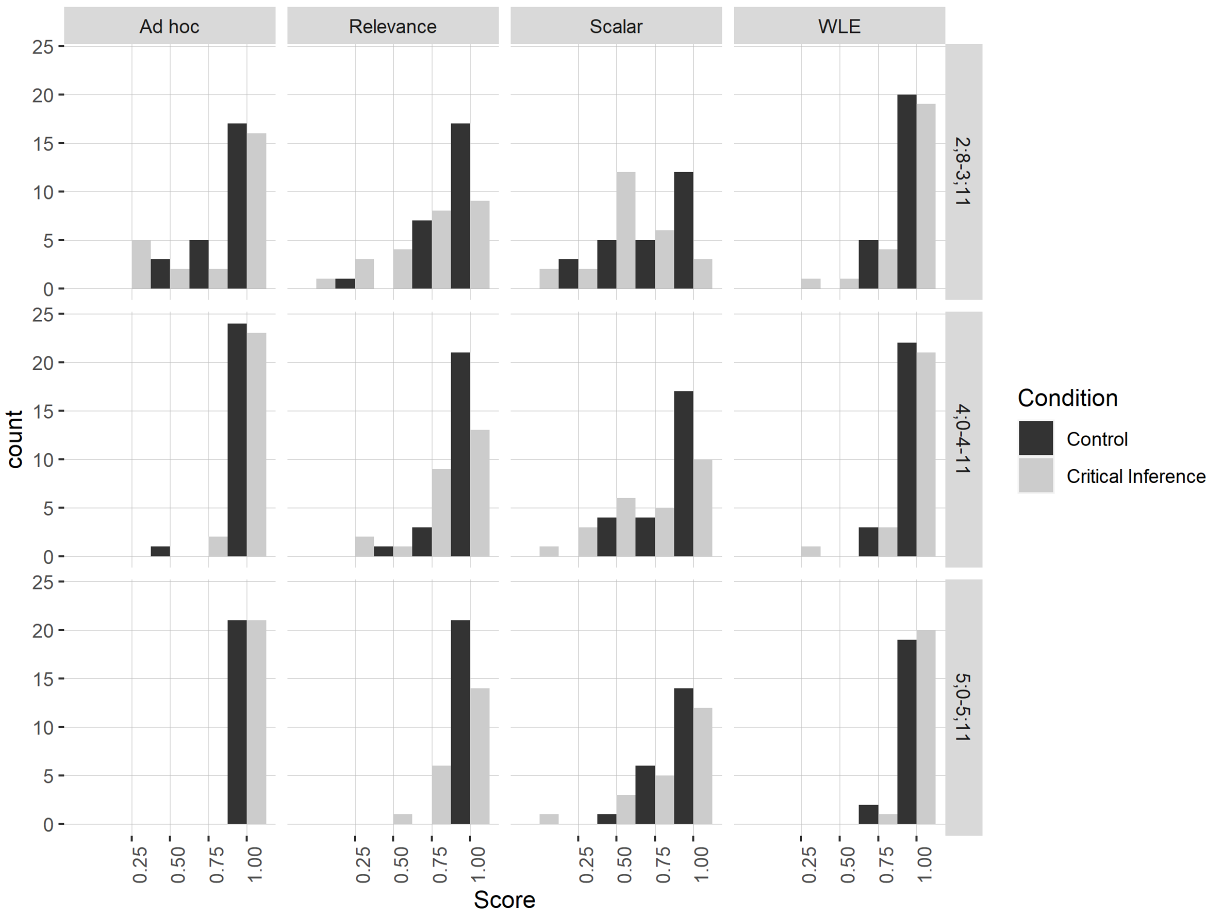



*Figure 2 Distribution of participant scores by age, inference type and condition*

*Table 9 ANOVA model comparison for effect of block order, using glmer, family = binomial, optimizer =*

*bobyqa, sum coding*

| Model | Df | AIC | Log Lik | Deviance | $\chi^2$ | p |
|---|---|---|---|---|---|---|
| Score ~ 1 + (1 + Critical + Age group + Trial_block \| Item) | 29 | 1256.8 | -599.4 | 1198.8 | | |
| Score ~ Critical + (1 + Critical + Age group + trial_block \| Item) | 30 | 1249.5 | -594.7 | 1189.5 | 9.35 | .002 |
| Score ~ Critical + Type + (1 + Critical + Age group + Trial_block \| Item) | 32 | 1241.8 | -588.9 | 1177.8 | 11.69 | .003 |
| Score ~ Critical + Type + Age group + (1 + Critical + Age group + Trial_block \| Item) | 34 | 1215.8 | -573.9 | 1147.8 | 30.01 | < .001 |
| Score ~ Critical + Type + Age group + Trial_block + (1 + Critical +Age group + Trial_block \| Item) | 37 | 1218.1 | -572.0 | 1144.1 | 3.68 | .3 |

*Table 10 ANOVA model comparison for effect of block order for scalar trials, using glmer, family = binomial, optimizer = bobyqa, sum coding*

| Model | Df | AIC | Log Lik | Deviance | $\chi^2$ | p |
|---|---|---|---|---|---|---|
| Score ~ 1 + (1 + Critical + Age group + Trial_block \| Item) | 29 | 636.5 | -289.3 | 578.5 | | |
| Score ~ Critical + (1 + Critical + Age group + trial_block \| Item) | 30 | 635.2 | -287.6 | 575.2 | 3.29 | .07 |
| Score ~ Critical + Type + (1 + Critical + Age group + Trial_block \| Item) | 32 | 629.7 | -282.9 | 565.7 | 9.5 | .009 |
| Score ~ Critical + Type + Age group + Trial_block + (1 + Critical +Age group + Trial_block \| Item) | 35 | 633.1 | -281.6 | 563.1 | 2.8 | .46 |

In an exploratory analysis, we investigated the associations of structural language, SES and Theory of Mind with performance on the implicature task. Not all children completed both sessions or returned the parental background questionnaire, so this analysis was conducted on a subset of 58 children for whom all data was available. We conducted model comparison using the anova function with mixed-effects logistic regression models, using implicature scores in the critical condition (for relevance, ad hoc and scalar implicatures) as the outcome variable. The BPVS-3 and the TROG II scores were centred and scaled, and then a mean for each participant calculated, to provide a composite structural language score. Age (in months), structural language, Theory of Mind and SES scores were each centred and scaled; gender was coded with sum contrasts. We added the factors in the following order: gender, structural language, SES and Theory of Mind. This was because we wanted to control for the effect of structural language in assessing the contribution of Theory of Mind, as it is arguably related to mentalising (Milligan, Astington & Dack, 2007); likewise, given the association of vocabulary with SES, we wanted to see whether SES independently predicted pragmatic performance (Pace et al., 2017). Structural language was the only factor which significantly improved the model, once age gender and SES are taken into account ($\chi^2(1) = 6.85$, $p = .009$) – Table 11.

*Table 11 ANOVA model comparison for Age, Gender, structural language, SES and ToM for monolinguals*

| Model | Df | AIC | Log Lik | Deviance | $\chi^2$ | p |
|---|---|---|---|---|---|---|
| Score ~ 1 + (1 + Age + Gender + SES + Language + ToM \| Item.no) | 22 | 609.62 | -282.81 | 565.62 | | |
| Score ~ Age + (random effects) | 23 | 582.00 | -268.00 | 536.00 | 29.62 | < .001 |
| Score ~ Age + Gender + (random effects) | 24 | 583.79 | -267.90 | 535.79 | .21 | .65 |
| Score ~ Age + Gender + Structural Language + (random effects) | 25 | 578.95 | -264.47 | 528.95 | 6.85 | .009 |
| Score ~ Age + Gender + Structural Language + SES + (random effects) | 26 | 579.60 | -263.80 | 527.60 | 1.35 | .25 |
| Score ~ Age + Gender + Structural Language + SES + ToM + (random effects) | 27 | 580.97 | -263.49 | 526.97 | .63 | .43 |

**Discussion**

In our study, we found evidence that the preschool years, aged three to five, are important ones for pragmatic development: the ability to derive some implicatures, like ad hoc quantity and simple relevance, emerges reliably in the fourth year of life, and continues to improve over the following years. Overall, children's performance increased with age, and each age group performed better than the previous one, and it was better overall in control trials (which required no pragmatic inference) compared to critical trials (which required an implicature to be derived). We also observed different developmental trajectories across inference types, with word learning by exclusion in place first, followed by relevance and ad hoc quantity, and finally scalar quantity implicatures.

These findings complement others which have found that children aged three are able to derive ad hoc quantity and, separately, relevance implicatures (Grosse et al., under review; Schulze et al., 2013; Stiller et al., 2015; Tribushinina, 2012; Yoon & Frank, 2019), and extend them by showing this competence in a single sample of children and in a task which requires both kinds of inference to be made. Similarly, scalar implicatures with *some* prove to be more challenging than ad hoc quantity implicatures, again complementing existing findings (Foppolo et al., 2020; Grosse et al., under review; Horowitz et al., 2018), but for the first time indicating how this pattern develops over three successive years.

Based on the notion that both relevance and quantity implicatures crucially involve understanding relevance and tracking QUD, but quantity in addition involves generating and negating alternatives, we tentatively predicted that we might see relevance implicatures emerging first. Contrary to this expectation, we did not find evidence for a difference between relevance and ad hoc performance. There could be multiple possible reasons for this: the task may have not been sensitive enough to capture any difference, for example if the relevance items were harder than ad hoc items for an independent reason, such as the background knowledge they required; or it may be that once children can appreciate relevance and track the QUD they are relatively easily able to integrate this with generating and negating relevant alternatives in a quantity implicature – certainly the basic exclusion inferential mechanism seems to be in place early, based on ceiling performance in the word learning by exclusion condition. In other words, these results do not yet constitute evidence against the key role of developing an ability to understand relevance and track the QUD, but rather invite further research. Similarly, given these shared requirements between quantity and relevance inferencing, we expected to see a relationship between performance across SIs, ad hocs and relevance implicatures. However, the results of the exploratory correlational analyses with the youngest age group were mixed: relevance and scalar inferences were correlated, but relevance and ad hoc inferences were not. It could be that the correlation of performance on relevance and scalar inferences reflects the shared components, while the lack of correlation between ad hocs and relevance is due to the lack of variation in ad hocs. Alternatively, it could be that the correlation we did observe merely reflects unrelated similarities and differences in the stimuli across the implicature types; future task improvements, discussed below, could elucidate this.

As in other studies, we observed scalar implicatures to be the latest in which children become competent. The youngest children, in particular, are not at ceiling in the control condition, with *all*, which suggests that learning the semantics of quantifiers per se – let alone learning scales or accessing the relevant alternative – might be one particular challenge, in line with Horowitz, Schneider and Frank's (2018) findings that quantifier knowledge is one key challenge for scalar implicatures. Explaining the difference between control and critical conditions, though, is not possible with this kind of design, i.e. for those children who know the semantics of *some* and *all*, one cannot tease apart with a simple picture-selection task whether the remaining challenge is learning that they are scalemates, or learning to generate *all* as a relevant alternative to *some*; this would require further experimental manipulation (e.g. Barner et al., 2011).

Interestingly, we did not observe a bimodal distribution for scalar implicatures, contrary to some previous studies where children are consistently correct or incorrect (Foppolo et al., 2020, Experiment 1; Guasti et al., 2005; Horowitz et al., 2018; Skordos & Papafragou, 2016). For the youngest age group, the modal score was .5, while for all other age groups it was 1, with the distribution skewed towards ceiling performance. One possible reason for this might be task differences: Foppolo et al (2020, Experiment 1), Guasti et al (2005) and Skordos & Papafragou (2016) all employ a Truth Value Judgement task, with a single inference type. Horowitz, Schneider and Frank (2018) do use a picture-matching task, but they test only quantity implicatures (ad hoc and scalar in Experiment 1, and only scalar in Experiments 2-4); it could be that switching between relevance and quantity in our task meant that quantity was not highlighted as an important part of the QUD so much. Furthermore, the stimuli in Horowitz, Schneider and Frank (2018) contained either four of one object type (e.g. four cats) or two of one type and two of another (e.g. two cats and two birds), whereas in our study a larger number of objects had some property or not (e.g. all plates were broken or not); in the case where children do not derive a scalar implicature, and therefore have to guess between the two pictures, as both match the literal *at least some* interpretation, it could be that the picture matching *all* was more salient and more likely to be chosen in Horowitz, Schneider and Frank's design. In addition, if children were simply ignoring the quantifier, they would arrive at the wrong picture consistently in their design, by way of an ad hoc implicature ('some of the animals are cats' would be interpreted as 'the animals are cats and nothing else'), whereas for our design object type does not provide any further strategy for disambiguating the utterance. This highlights the potentially significant difference apparently small changes in design can make in the way that they affect the communicative context.

Finally, we did not find evidence for a practice effect, either in general or for scalar inferences in particular: adding in the story order (with each story containing one critical and one control for each implicature type) did not improve the fit of the model. Existing studies are mixed in their findings on order effects: Horowitz, Schneider and Frank (2018) also did not observe an effect, while Grosse et al (under review) and Skordos and Papafragou (2016) did see an advantage in hearing the stronger alternative *all* before a critical *some* implicature trial, in a picture-matching and judgement task, respectively. It is likely that in our case the switching between three implicature types may have removed any effect of lower-level priming or activation of the alternative; indeed, Horowitz and Frank (2015) observed worse performance when ad hoc and scalar trials were mixed together, compared to just testing scalars.

In our exploratory analysis of linguistic, sociocognitive and environmental factors which may affect children's pragmatic development, we found that only structural language (a composite of receptive vocabulary and grammar) predicted children's pragmatic performance (their score on relevance, ad hoc and scalar implicature trials), once gender and age were controlled for. Again this complements emerging findings in the literature of the association between pragmatic and linguistic skills in older children (Antoniou & Katsos, 2017; Foppolo et al., 2020) and with global pragmatics measures (Matthews et al., 2018). Theoretically this association could be expected in either direction (structural language contributing to pragmatic skills or vice versa) or, most likely, bidirectional: for any particular utterance, the vocabulary and grammatical constructions used trigger or constrain any implicature derived, and the more linguistic experience that has contributed to vocabulary and grammatical knowledge, the more opportunities to practice pragmatic skills as well; on the other hand pragmatic inferencing is a key way that children can learn the meaning of new words or constructions (Bohn & Frank, 2019; Horowitz & Frank, 2016) and semantic and pragmatic skills are difficult to disentangle, especially developmentally (Matthews et al., 2018). Interestingly, this pattern has also emerged in a related but functionally distinct line of research: children's development of reading inferences. While the type of inference tested is typically different, longitudinal studies have found bidirectional associations, such that vocabulary skills predict later inferencing skills, which in turn predict later vocabulary skills (Language and Reading Research Consortium, Currie & Muijselaar, 2019). Future work could adopt such longitudinal designs for implicatures as well, to begin to understand the directionality of influence; in addition, more investigation is needed of the contribution of other factors such as the similarity of tasks (in our study, both the structural language and implicature tasks were essentially sentence- or word-to-picture-matching).

We did not observe evidence for an effect of SES on implicature performance (controlling for language). This stands in contrast to the strong associations between structural language and SES but echoes the findings of other studies on children's implicature development (Antoniou et al., 2020; Antoniou & Katsos, 2017; Schulze et al., 2020). However, given that none of the studies on implicatures, including this one, were explicitly designed to test the association of SES and pragmatic skill, more research in this area is clearly needed to ascertain whether SES only has an affect on pragmatic development as mediated by structural language skills, whether it contributes independently, or not at all. If pragmatic skills like implicature derivation turn out to be less influenced by differences in SES than structural language skills like vocabulary, this raises interesting questions to do with the prerequisites of pragmatic development and the role played by the input.

We also did not observe any effect of Theory of Mind, controlling for language and SES, which is unexpected given a Gricean approach to pragmatics which implicates reasoning about the speaker's knowledge and beliefs, and a constraint-based approach in the same spirit, where tracking a mutual QUD is important (Degen & Tanenhaus, 2014; Grice, 1989). Alternative pragmatic accounts (e.g. Andrés-Roqueta & Katsos, 2017; Kissine, 2016) propose that some pragmatic inference types, including some quantity implicatures, are available without sophisticated mentalising in some communicative situations. For instance, simple scalar or ad hoc implicatures could be derived through an egocentric search for relevance, based on an awareness that more informative descriptions are preferred: by reasoning that, for instance, *I broke some of the plates* is an underinformative description of a picture in which all the plates are broken, and so matching the less informative term (e.g. *some*) to the correct picture, without attributing any intentions to communicate this enriched meaning on the part of the speaker. There is a small but growing range of evidence to support these alternative views (e.g. Andrés-Roqueta & Katsos, 2020; Wilson et al., under revision). Some reflection, though, shows that correlating Theory of Mind tests with performance on implicature tasks is problematic for a number of reasons: they have their own linguistic and cognitive demands which may obscure children's actual ability with False Belief, or at least present additional challenges to the implicature task (Rubio-Fernández & Geurts, 2013, 2016). In addition, with a range of possible scores of 0-3 for the Change-of-Location and Unexpected Contents tasks, there is not much variance for correlational analyses. Moreover, while these tasks are often taken as a "gold standard" for Theory of Mind, they measure False Belief, which is only one aspect of mentalising, and may not be required for implicatures in a simple communicative situation such as in our picture-matching task. An approach which could offer clearer interpretation of results would involve experimental manipulation of Theory of Mind within a pragmatic inferencing task, such as manipulating whether or not the speaker is knowledgeable (for adults see Breheny et al., 2013; and for paradigms suitable for children see Kampa & Papafragou, 2020, and Wilson et al., under review).

One strength of this study was the way in which several inference types were combined in a single task, with a more naturalistic story task with context sentence and explicit QUD. Future studies could further improve this combination of a more naturalistic task with experimental control: in particular, the relationship of the explicit QUD to the critical utterance could be more tightly controlled across inference types. For ad hocs, a question of the type, *what did you take from the fridge?* made an exhaustive, ad hoc implicature interpretation highly relevant; for scalars, a question of the type *what did you do with the pile of plates?* may have made a scalar *some but not all* interpretation less relevant compared to an action (*I* broke *some/all of them*), even though the question was similar in form to the question for ad hocs. Likewise, as in Horowitz, Schneider and Frank's (2018) design, having the same visual stimuli across all inference types would be an improvement, reducing possible differences between types due to item effects. Further, while the relevance items were based closely on previous studies (Schulze, Grassmann and Tomasello, 2013), one potential concern with them is that the correct picture could be chosen purely based on a semantic association between the key word in the utterance and the picture. That is, instead of using semantic and world knowledge in a pragmatic inference to derive the speaker's intended relevant meaning, the association, such as 'milk goes with cereal' (rather than toast) or 'brushes go with paint' (rather than crayons) is used to solve the task without reference to the speaker. In our study, the majority of items were arguably open to this interpretation; one exception, for instance, was:

(5) What fruit do you want to pick?

I'll get a ladder.

(Choice: apple or strawberries)

Future studies could use these kinds of items, while also making sure that children possess the relevant world knowledge, in order to rule out the possibility of using a simple association strategy.

While in our study we treated age group as a main predictor and compared performance across age groups, in line with previous studies, the different developmental trajectories of different inferences, and the association with at least one other developmental factor (structural language), suggests that a fruitful way forward in future research could be to examine children's development of pragmatic inferences primarily in relation to other skills. In other words, the driving question becomes not, 'at what age can children derive a certain implicature?', but instead 'which developing skills are associated with or necessary for a certain implicature?'. Given that there is great variation in age of acquisition for many linguistic skills (Kidd et al., 2018), this could enhance our understanding more than only comparing children by age groups. That said, this study also raises the question of what it is that develops around the fourth year of life which enables implicature comprehension to improve, when word learning by exclusion is grasped much earlier. Indeed, studies which have tested two-year-olds with ad hoc implicatures, even with specially adapted designs, have not found evidence for competence at that age (Horowitz et al., 2018; Stiller et al., 2015). It could be that completely different experimental paradigms which are more social and interactive in nature could reveal the beginnings of implicature understanding: Schulze and Tomasello (2015), for instance, found that even 18-month-olds are able to interpret an intentional non-verbal indirect request in the context of a game (in contrast to the same action performed unintentionally).

In sum, the findings of our study suggest that the preschool years, ages three to five, are crucial for children's developing understanding of implicatures: children aged three years are able to derive some types of implicature, like relevance and simple ad hoc quantity, and this continues to improve through to age four or five. Scalar implicatures with quantifiers, though, are more challenging, while word learning by exclusion inferences are in place early. Within a constraint-based approach to implicatures, we argued theoretically for a key role in learning to understand relevance and track the QUD for all implicature types. Our results neither contradict this hypothesis nor provide strong support – relevance and ad hoc implicatures emerged together, and a correlation was only found between relevance and scalar implicatures, but not relevance and ad hocs – and so invite further research. Finally, it seems that developing structural language skills are closely linked to pragmatic skills, but the directionality of this relationship requires further investigation.

# References

Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with Autistic Spectrum Disorders. *Frontiers in Psychology*, *8*, 996.

Andrés-Roqueta, C., & Katsos, N. (2020). A Distinction Between Linguistic and Social Pragmatics Helps the Precise Characterization of Pragmatic Challenges in Children With Autism Spectrum Disorders and Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1494–1508. https://doi.org/10.1044/2020_JSLHR-19-00263

Antoniou, K., & Katsos, N. (2017). The effect of childhood multilingualism and bilectalism on implicature understanding. *Applied Psycholinguistics.*, 1–47.

Antoniou, K., Veenstra, A., Kissine, M., & Katsos, N. (2020). How does childhood bilingualism and bi-dialectalism affect the interpretation and processing of pragmatic meanings? *Bilingualism: Language and Cognition*, *23*(1), 186–203. https://doi.org/10.1017/S1366728918001189

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*(1), 87–96.

Barner, D., Hochstein, L. K., Rubenson, M. P., & Bale, A. (2018). Four-year-old children compute scalar implicatures in absence of epistemic reasoning. In *Semantics in Language Acquisition* (Vol. 24, pp. 325–349).

Barner, D., & Snedeker, J. (2008). Compositionality and Statistics in Adjective Acquisition: 4-Year-Olds Interpret Tall and Short Based on the Size Distributions of Novel Noun Referents. *Child Development*, *79*(3), 594–608. https://doi.org/10.1111/j.1467-8624.2008.01145.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Benz, A., & Jasinskaja, K. (2017). Questions Under Discussion: From Sentence to Discourse. *Discourse Processes*, *54*(3), 177–186. https://doi.org/10.1080/0163853X.2017.1316038

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1450.

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.

Bernicot, J., & Legros, S. (1987). Direct and indirect directives: What do young children understand? *Journal of Experimental Child Psychology*, *43*(3), 346–358.

Bishop, D. V. M. (2003). *TROG-2—Test for reception of grammar-2* (2nd ed.). Harcourt.

Bohn, M., & Frank, M. C. (2019). The Pervasive Role of Pragmatics in Early Language. *Annual Review of Developmental Psychology*, *1*(1), 223–249. https://doi.org/10.1146/annurev-devpsych-121318-085037

Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The family affluence scale as a measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research*, *78*(3), 473–487.

Braxmeier, H., & Steinberger, S. (2017). *Pixabay*. www.pixabay.com

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition*, *126*(3), 423–440.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

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*(2), 417–431.

Cremers, A., Kane, F., Tieu, L., Kennedy, L., Sudo, Y., Folli, R., & Romoli, J. (2018). Testing theories of temporal inferences: Evidence from child language. *Glossa*, *3*(1). https://doi.org/10.5334/gjgl.604

Cummings, L. (2005). *Pragmatics: A multidisciplinary perspective*. Edinburgh University Press.

de Villiers, P. A., de Villiers, J. G., Coles-White, D., & Carpenter, L. (2009). Acquisition of Relevance Implicatures in Typically-Developing Children and Children with Autism. *Proceedings of the 33rd Annual Boston University Conference on Language Development*, *33*, 121–132. Linguistics and Language Behavior Abstracts (LLBA). http://search.proquest.com/docview/85716304?accountid=9851

Degen, J., & Tanenhaus, M. (2019). Constraint-based pragmatic processing. In C. Cummins & N. Katsos (Eds.), *The Oxford handbook of experimental semantics and pragmatics* (pp. 21–38).

Degen, J., & Tanenhaus, M. K. (2014). Processing Scalar Implicature: A Constraint-Based Approach. *Cognitive Science*, 667–710. https://doi.org/10.1111/cogs.12171

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630.

Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science*, *14*(2), 164–168.

Dunn, L., Dunn, L., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., & Burge, B. (2009). *The British picture vocabulary scale* (3rd ed.). GL Assessment. https://scholar.google.com/scholar_lookup?publication_year=2009&author=L.+Dunn&author=L.+Dunn&author=J.+Sewell&author=B.+Styles&author=B.+Brzyska&author=Y.+Shamsan&author=B.+Burge&title=The+British+picture+vocabulary+scale

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar Implicatures in Child Language: Give Children a Chance. *Language Learning and Development*, *8*(4), 365–394. Linguistics and Language Behavior Abstracts (LLBA).

Foppolo, F., Mazzaggio, G., Panzeri, F., & Surian, L. (2020). Scalar and ad-hoc pragmatic inferences in children: Guess which one is easier. *Journal of Child Language*, 1–23. https://doi.org/10.1017/S030500092000032X

Fortier, M., Kellier, D., Flecha, M. F., & Frank, M. C. (under review). *Ad-hoc pragmatic implicatures among Shipibo-Konibo children in the Peruvian Amazon* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/x7ad9

Graham, S. A., Poulin-Dubois, D., & Baker, R. K. (1998). Infants' disambiguation of novel object words. *First Language*, *18*(53), 149–164.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Grosse, G., Schulze, C., Noveck, I., Tomasello, M., & Katsos, N. (under review). *Three-year-olds make some, but not all inferences based on informativeness*.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*(5), 667–696.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*(1), B23–B34.

Hirschberg, J. (1991). *A theory of scalar implicature*. Garland Press.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2018). The Trouble With Quantifiers: Exploring Children's Deficits in Scalar Implicature. *Child Development*, *89*(6), E572–E593. https://doi.org/10.1111/cdev.13014

Horowitz, A., & Frank, M. C. (2015). Sources of developmental change in pragmatic inferences about scalar terms. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.* http://langcog.stanford.edu/papers/ACHMCF-cogsci-underreview.pdf

Horowitz, A., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development*, *87*, 807–819.

Hurewitz, F., Papafragou, A., Gleitman, L., & Gelman, R. (2006). Asymmetries in the acquisition of numbers and quantifiers. *Language Learning and Development*, *2*(2), 77–96.

Kalashnikova, M., Mattock, K., & Monaghan, P. (2014). Disambiguation of novel labels and referential facts: A developmental perspective. *First Language*, *34*(2), 125–135. https://doi.org/10.1177/0142723714525946

Kampa, A., & Papafragou, A. (2020). Four-year-olds incorporate speaker knowledge into pragmatic inferences. *Developmental Science*, *23*(3), e12920. https://doi.org/10.1111/desc.12920

Katsos, N., & Bishop, D. V. M. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, *120*(1), 67–81.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kimball, S., Mattis, P., & The GIMP Development Team. (2016). *GNU Image Manipulation Program* (2.8.18) [Computer software].

Kissine, M. (2016). Pragmatics as Metacognitive Control. *Frontiers in Psychology*, *6*, 2057. https://doi.org/10.3389/fpsyg.2015.02057

Language and Reading Research Consortium, Currie, N. K., & Muijselaar, M. M. L. (2019). Inference making in young children: The concurrent and longitudinal contributions of verbal working memory and vocabulary. *Journal of Educational Psychology*, *111*(8), 1416–1431. https://doi.org/10.1037/edu0000342

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language & Communication Disorders*, *37*(1), 3–15.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

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.

Miller, K., Schmitt, C., Chang, H.-H., & Munn, A. (2005). Young children understand some implicatures. *Proceedings of the 29th Annual Boston University Conference on Language Development*, 389–400.

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. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Nordmeyer, A. E., Yoon, E. J., & Frank, M. C. (2016). Distinguishing processing difficulties in inhibition, implicature, and negation. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 2789–2794. http://langcog.stanford.edu/papers_new/nordmeyer-2016-underrev.pdf

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, *78*(2), 165–188.

Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, *11*(3), 191–214.

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying Pathways Between Socioeconomic Status and Language Development. *Annual Review of Linguistics*, *3*(1), 285–308. https://doi.org/10.1146/annurev-linguistics-011516-034226

Papafragou, A., & Skordos, D. (2016). Scalar Implicature. In J. Lidz, W. Snyder, & J. Pater (Eds.), *Oxford Handbook of Developmental Linguistics* (pp. 611–632). Oxford University Press.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language Acquisition*, *14*(4), 347–375.

*Qualtrics* (Version 2016). (2016). [Computer software]. Qualtrics. www.qualtrics.com

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015, URL h ttp. *Www. R-Project. Org*.

Reetzke, R., Zou, X., Sheng, L., & Katsos, N. (2015). Communicative development in bilingually exposed Chinese children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*, *58*(3), 813–825.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, *5*, 6–1.

Rubio-Fernández, P., & Geurts, B. (2013). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science*, *24*(1), 27–33.

Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, *7*(4), 835–850.

Schulze, C., Endesfelder Quick, A., Gampe, A., & Daum, M. M. (2020). Understanding verbal indirect communication in monolingual and bilingual children. *Cognitive Development*, *55*, 100912. https://doi.org/10.1016/j.cogdev.2020.100912

Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-Year-Old Children Make Relevance Inferences in Indirect Verbal Communication. *Child Development*, *84*(6), 2079–2093. https://doi.org/10.1111/cdev.12093

Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition*, *136*, 91–98. https://doi.org/10.1016/j.cognition.2014.11.036

Scrafton, S., & Feeney, A. (2006). Dual processes, development, and scalar implicature. *28th Annual Conference of the Cognitive Science Society*, 774–779.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition*, *153*(6–18).

Stiller, A. J., Goodman, N. D., & Frank, M. C. (2015). Ad-hoc Implicature in Preschool Children. *Language Learning and Development*, *11*(2), 176–190. https://doi.org/10.1080/15475441.2014.927328

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tribushinina, E. (2012). Comprehension of relevance implicatures by pre-schoolers: The case of adjectives. *Journal of Pragmatics*, *44*(14), 2035–2044. https://doi.org/10.1016/j.pragma.2012.09.018

Veenstra, A., & Katsos, N. (2018). Assessing the comprehension of pragmatic language: Sentence judgment tasks. In A. H. Jucker, K. P. Schneider, & W. Biblitz (Eds.), *Methods in Pragmatics* (pp. 257–279). de Gruyter Mouton.

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.

Wilson, E., & Katsos, N. (2020). Acquiring implicatures. In K. Schneider & E. Ifantidou (Eds.), *Developmental and Clinical Pragmatics*. de Gruyter Mouton.

Wilson, E., Lawrence, R., & Katsos, N. (under revision). *The role of perspective-taking in children's quantity implicatures*.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Yoon, E. J., & Frank, M. C. (2019). The role of salience in young children's processing of ad hoc implicatures. *Journal of Experimental Child Psychology*, *186*, 99–116. https://doi.org/10.1016/j.jecp.2019.04.008

Zhao, S., Jie, R., Frank, M. C., & Zhou, P. (in press). Mandarin Children's Interpretation of Implicatures and Inference. *Language Learning and Development*. https://doi.org/10.17605/OSF.IO/SYBMJ

Zondervan, A. J. (2010). Scalar implicatures or focus: An experimental approach. *LOT Dissertation Series*, *249*.

Zondervan, A., Meroni, L., & Gualmini, A. (2008). Experiments on the role of the question under discussion for ambiguity resolution and implicature computation in adults. *Proceedings of SALT*, *18*, 765–777.