Journal of Child Language (2021), **48**, 350–372
doi:10.1017/S030500092000032X

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Scalar and ad-hoc pragmatic inferences in children: guess which one is easier

Francesca FOPPOLO[a*], Greta MAZZAGGIO[b*], Francesca PANZERI[a], and Luca SURIAN[b]

[a]Department of Psychology, University of Milan-Bicocca and [b]Department of Psychological and Brain Sciences, University of Trento
Corresponding author: Francesca Foppolo, Department of Psychology, University of Milano-Bicocca, Ed. U6 Room 3136 3rd Floor, Piazza dell'Ateneo Nuovo, 1, 20126 Milano, Italy; Email: francesca.foppolo@unimib.it;

(Received 13 May 2019; revised 18 February 2020; accepted 13 May 2020;
first published online 10 June 2020)

Abstract

Several studies investigated preschoolers' ability to compute scalar and ad-hoc implicatures, but only one compared children's performance with both kinds of implicature with the same task, a picture selection task. In Experiment 1 ($N = 58$, age: 4;2-6;0), we first show that the truth value judgment task, traditionally employed to investigate children's pragmatic ability, prompts a rate of pragmatic responses comparable to the picture selection task. In Experiment 2 ($N = 141$, age: 3;8-9;2) we used the picture selection task to compare scalar and ad-hoc implicatures and linked the ability to derive these implicatures to some cognitive and linguistic measures. We found that four- and five-year-olds children performed better on ad-hoc than on scalar implicatures. Furthermore, we found that morphosyntactic competence was associated with success in both kinds of implicatures, while performance on mental state reasoning was positively associated with success on scalar but not ad-hoc implicatures.

Keywords: acquisition of pragmatics; scalar implicatures; ad-hoc implicatures

Introduction

Children's ability to cope with conversational inferences has been the matter of a lively debate both in the linguistic and the psychological literature of the past decade. Following some seminal works on children's interpretation of the quantifier *some*, the modal *might* and the disjunction *or* (Chierchia, Crain, Guasti, Gualmini & Meroni, 2001; Noveck, 2001), a thriving body of research has been devoted to

*Francesca Foppolo and Greta Mazzaggio are joint first authors of this paper.

Authors' contribution is as follows: Greta Mazzaggio conceived the experimental question, under the supervision of Luca Surian. Francesca Foppolo and Francesca Panzeri developed the tasks. Greta Mazzaggio recruited the children and supervised the testing. Francesca Foppolo performed the statistical analyses. Francesca Foppolo and Greta Mazzaggio drafted the manuscript which was critically revised by all authors.

experimental pragmatics, with a special focus on children's and adults' interpretation of the quantifier *some* (cf. Chemla & Singh, 2014a,b; Foppolo & Marelli, 2017; Skordos & Papafragou, 2016 for overviews).

To introduce the topic of our investigation, imagine a typical situation in which two hungry boys are looking for a snack in the kitchen cupboard. If boy A utters sentence (1) to boy B after opening the cookie box, then boy B will infer that there are some cookies left in the box, i.e., (2). This inference is known as a SCALAR IMPLICATURE. When the strictly literal meaning of (1) is enriched with the scalar implicature in (2), the strengthened meaning (3) is derived.

(1) Mommy ate *some* of the cookies.
(2) Mommy did not eat *all* of the cookies.
(3) Mommy ate *some but not all* of the cookies.

The scalar implicature in (2) arises by virtue of the fact that the speaker chose to utter (1) instead of a possible – equally plausible – alternative, namely (4), which would be compatible with a situation in which the box is empty.

(4) Mommy ate *all* of the cookies.

Assuming that the speaker is cooperative, the hearer will thus infer (3) on the basis of the speaker's choice to utter (1) over (4), following a standard Gricean reasoning about the maxims of conversations that rule our conversational exchanges (Grice, 1975). In particular, the Gricean First Maxim of Quantity urges the speaker to provide as much information as required by the goal of the exchange: in the example at hand, the quantity of cookies left in a box is relevant to the hearer, and the use of a sentence like (1) in the case Mommy ate *all* of the cookies would be UNDERINFORMATIVE given that a more informative alternative, i.e., (4), could be uttered.

Pragmatic inferences that ensue from lexical scales like <*some*, *all*> are one kind of generalized (conversational) implicature and depend on the scalar ordering of items on a scale of informativity in which the stronger element in the scale (in this case, *all*) entails the weaker term (*some*). In such kinds of implicatures, alternatives are linguistically determined by their position on this scale (Horn, 1972).

There is also another kind of conversational implicature, one in which alternatives are not linguistically pre-determined but are made available due to special features of the context. Sticking to our example above, suppose our hungry children find two cookie boxes in the cupboard, one with a ribbon and one with a ribbon and a flag. Then boy A peeps in the boxes and utters (5).

(5) The box with the ribbon is empty!

In this context, boy B would surely understand that the empty box is the one with (only) the ribbon and thus he will look for his snack in the box with the ribbon and the flag. The underlying reasoning is analogous to the one made above: sentence (5) is one of the possible alternatives that the speaker could have used. In particular, the speaker could have uttered (6), which would be optimally informative in the case the empty box were the one with both the ribbon and the flag; the fact that he didn't entitled the hearer to infer that (7) is what the speaker meant, which guides the hearer towards the correct choice.

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

(6) The box with the ribbon and the flag is empty!
(7) The box with only the ribbon is empty!

While the mentalistic steps beyond the derivation of the pragmatic inferences in (3) and (7) are the same, the way alternatives are brought to salience is radically different: while in the case of generalized scalar implicatures alternatives are linguistically encoded in a scale, in the case of particularized implicatures alternatives are construed ad-hoc in the context of utterance. A difference along this dimension makes a crucial difference for some theoretical accounts, but not for others, as we will discuss.

As for children's performance in the derivation of pragmatic inferences, previous works in the acquisition literature showed that pre-school aged children have difficulties in deriving the scalar implicature associated to *some* and other scalar terms, and that difficulties remain for some of the children even in the most facilitating and ecological settings (cf. Foppolo, Guasti & Chierchia, 2012 and Skordos & Papafragou, 2016 for an overview). On the other hand, a recent study by Stiller, Goodman and Frank (2015) showed that three and a half year-old children are already adult-like in interpreting particularized inferences like the one in (7), which are labeled ad-hoc implicatures (see also Bredart, 1984; Jackson & Jacobs, 1982 and Surian & Job, 1987 for related findings).

Different hypotheses have been formulated to account for children's difficulty with scalar implicatures. Under one approach, that we might label the *lexicalist* approach, children's failure with the *some but not all* implicature stems from the fact that children have not lexicalized the scale yet or have problems in retrieving the scalar alternatives to *some*: though knowing the meaning of *some* and *all*, they might not know that the two are part of a scale, or they might still be in the process of drawing stable links amongst quantifiers belonging to the same scale. Under Foppolo *et al.*'s (2012) proposal, children's failure with the scalar implicature associated to *some* reflects children's immaturity at the lexical level, in which two layers of meaning should be associated to the quantifier: the existential meaning (according to which 'some P Q' instantiates the existence of at least one P that Q) – and the scalar or enriched meaning (according to which this quantifier is ordered with other alternatives on a scale so that 'some P Q' leads to the negation of the stronger scalar alternatives 'most P Q' and 'all P Q'). While almost all children aged 6 and 7 are adult-like in the derivation of the enriched meaning of *some*, four- and five-year-olds children split, and only some of them derive the scalar implicature, despite the fact that all of them recognize that *all* is a better description than underinformative *some* (Foppolo *et al.*, 2012: Experiment 1 and Experiment 5). Furthermore, five-year-old children improve when optimal lexical alternatives are presented alongside underinformative uses of *some* (Foppolo *et al.*, 2012: Experiment 6). Similarly, Barner, Brooks and Bale (2011) argued that children have difficulty in accessing the relevant scalar alternatives in the lexicon, and showed that such difficulty is not grounded in memory limitation, as demonstrated by children succeeding in accessing relevant contextual alternatives. Thus, under the LEXICALIST hypothesis, there is a developmental stage in which children have not completed the additional learning step that links scalar quantifiers in a scale and consequently fail in rapidly and automatically accessing lexical items as scalar alternatives (see also Tieu, Romoli, Zhou & Crain, 2015 for an explanation along these lines for the scalar implicature associated with the scale <*or, and*>).

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

A different approach is what we might label the PROCESSING account: within a Relevance-Theoretic framework, Pouscoulous, Noveck, Politzer and Bastide (2007) argued that children's problems with scalar implicatures are due to difficulties in the optimization process between the cognitive gains and the processing costs of deriving the most informative interpretation of the utterance. Their argument is that, in most cases, the non-enriched interpretation of a scalar term will often suffice as a relevant-enough interpretation of the utterance in which it occurs; in the light of a balance between cognitive benefits and processing effort, children might fail to access the enriched meaning in more complex scenarios and tasks.

A more general PRAGMATIC approach, however, links children's failure with scalar inferencing with their yet immature pragmatic system. This general idea is set out differently in a variety of accounts. Under Katsos and Bishop's hypothesis (Katsos & Bishop, 2011), children fail to reject underinformative *some* because they are more tolerant of pragmatic anomalies than adults: though recognizing that underinformative *some* is not optimal, they tolerate it, as if they were conforming to non-adult-like pragmatic norms. More recently, Skordos and Papafragou (2016) argued in favor of an important role of conversational relevance in the derivation of scalar implicatures and proposed that children's problem with scalar implicatures might lie in their failure to recognize that scalar terms constitute relevant alternatives in certain contexts. They tested children's performance in three different experiments. In Experiment 1 and 3, they modulated the availability of scalar alternatives, testing whether the mere presence of the stronger lexical member of the quantifier scale (*all*) or another scalar quantifier (*none*) during the experiment facilitated the generation of a scalar implicature when the weak alternative (*some*) was used. The hypothesis that the relevance of lexical alternatives plays a role in scalar implicature generation was tested in Experiment 2, by manipulating the degree to which the stronger lexical item could be easily recognized as a relevant alternative by children in a given context. Their results show that children were more prone to reject underinformative uses of *some* when alternatives were made accessible and relevant in the course of the experiment (see also Foppolo *et al.*, 2012: Experiment 6 for a similar finding).

With respect to the distinction introduced above between particularized (such as ad-hoc implicatures) and generalized implicatures (such as scalar implicatures), these three general accounts make different predictions. In principle, no difference should be expected between types of implicatures within a *pragmatic* or a *processing* account. According to these approaches, children's non-adult-like behavior derives from differences in pragmatic tolerance or in the processes required to access or recognize relevant alternatives (due to processing demands, as in Pouscolous *et al.*'s account, or to low saliency, as in Skordos and Papafragou's account). If this is the case, then these factors should similarly affect all kinds of implicatures. In fact, under some theoretical accounts, like the Relevance Theory (Sperber & Wilson, 1986/1995), the derivation of all pragmatic inferences rests on the same underlying mechanism, one in which context, together with the evaluation of costs and benefits, plays a key role in determining when enrichment is relevant or not.

Under LEXICALIST approaches, on the other hand, a difference between ad-hoc and scalar implicatures is expected: while in the case of ad-hoc implicatures the alternatives that are activated depend solely on context, in all generalized implicatures (including scalar implicatures), the set of alternatives is a feature of the language relying on the lexical representation of the scalar item itself. It is worth noting that also in this approach context might intervene in favoring (or suspending)

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

the inference, in obedience to Gricean maxims of conversations (Chierchia, 2013). The crucial difference is in the access to the alternatives, which depends on a linguistic representation and a lexical retrieval mechanism in the case of scalar quantifiers, while it is purely context-driven in the case of ad-hoc scales.

As we said, previous results indicate that children as young as three and a half are successful in deriving ad-hoc implicatures (Stiller *et al*., 2015), while a more complex picture emerges in the case of scalar implicatures, for which children's success has been shown to vary considerably across ages, materials and tasks; in general, not all children at age 5 have yet reached an adult-like stage.

In a recent study, Horowitz, Schneider and Frank (2017) compared ad-hoc and scalar implicatures directly by means of a Picture Selection task modelled after Stiller *et al*. (2015; see also, Jackson & Jacobs, 1982 and Surian & Job, 1987 for similar tasks), in which the child had to select a target (among different pictures) by following oral instructions. The authors tested children aged 4 and 5, reporting a better performance with ad-hoc than scalar implicatures. They also found a correlation between children's rate of interpretation of the scalar quantifier *some* as *some but not all* and their performance on the negative quantifier *none*. In order to account for these findings, they suggested that part of children's problems with scalar implicatures might be rooted in some difficulties at the semantic level, rather than in difficulties at the pragmatic level or in a lack of general processing resources.

Albeit their intriguing results and the clever experimental setting that allowed them to compare the two type of inferences directly, we believe that the task employed by Horowitz *et al*. (2017) deserves some further attention. We acknowledge the fact that the Picture Selection task has some advantages: it is ecological; the scalar alternatives to *some* are also presented as a visual alternative, thus enhancing their relevance; also, no metalinguistic judgment is required. However, we argue that relying on the Picture Selection task as such, without an additional control, might result in an underestimation, or overestimation, of the children's performance. The following example will clarify the reasons for this concern. In the task used by Horowitz and colleagues, the child had to match one sentence (e.g., "Some of the pictures are cats") by selecting one of three pictures that represented book covers. The visual alternatives for the above sentence, for example, included one book cover with four cats (the *all* target) and one book cover with two birds and two cats (the *some but not all* target), together with one distractor with no cats. The authors interpreted the child's selection of the *some but not all* target as evidence for their derivation of the scalar inference. However, even a child who is not in the stage of being able to derive such an inference might have selected this picture picking randomly one of the two pictures that depict cats. If the child is interpreting the sentence as meaning "(at least) Some (even all) of the pictures are cats", i.e., she is not enriching the meaning of *some* with the scalar implicature, then she might randomly select one of the two pictures with cats in it, independently of the fact that only some or all of the pictures are cats since, possibly, she is not paying attention to this distinction (yet).

Second, there is another methodological concern pertaining to this study, which is related to the fact that the negative quantifier *none* was also included in the experimental sentences. In fact, *none* does not belong to the scale <*some*, *all*>, and its presence might have affected children's computation of scalar inference with *some*, favoring a lower-bound interpretation of this quantifier. Indeed, in a visual world eye-tracking study on adults, Foppolo and Marelli (2017) found that listening to

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

some after a sentence with *none* induced slower convergence towards the pragmatic target compared to the case in which *some* followed a sentence with *all*. We believe that this factor deserves further investigation that goes beyond the purposes of this paper.

In this paper we aim at contributing to the debate about children's derivation of scalar and ad-hoc inferences by presenting two experimental studies. In a first study (Experiment 1), we tested children's performance on scalar quantifiers in two tasks, administered within subjects: a Power Point version of the classical Truth Value Judgment task (similar to the one used by Katsos & Bishop, 2011) and a referential Picture Selection task similar to the one used by Horowitz *et al.* (2017). The aim of this preliminary study was to compare the two tasks directly, in order to validate the Picture Selection task as a valuable tool to investigate pragmatic inferences, in the light of the possible criticism raised above. Differently from the Picture Selection task, the Truth Value Judgment task allows to (i) check the child's basic knowledge of the relevant quantifiers involved in the scale and (ii) discriminate pragmatic and logical responders more effectively. In this task, children are asked to say if a sentence like "Davide put *some* of the rabbits in the box" is a good or a bad description of a scenario in which Davide put *all* of the rabbits in a box, and if a sentence like "Davide put *some* of the balls in the box" is a good or a bad description of scenarios in which Davide put *only some* or *none* of the balls in a box (similar control conditions were created for felicitous/infelicitous uses of the quantifier *all*). They are also asked to always provide justification for their rejection. If the children accept both descriptions with *some* in the case in which some or all objects are in the box (but rejects it when none of the objects are in the box), then there is reason to conclude that (i) they know the meaning of *some* and (ii) they only have access to the logical, non-strengthened meaning of *some*. On the contrary, if the children reject the underinformative description in the *all* scenario, and provide the relevant justification, then they are indisputably accessing the strengthened meaning of *some* and deriving the scalar implicature. By administering the two tasks to the same group of children and comparing the individual children's responses across the two tasks, we are able to validate their responses in the Picture Selection task as truly logic or truly pragmatic. To our knowledge, no study so far compared the rate of pragmatic responses in the same group of children tested with a Truth Value Judgment task and a Picture Selection task, analyzing task as a within subject variable.

This first experiment was also meant to set the ground for our second study, in which we compared scalar implicatures and ad-hoc implicatures directly by means of the same task, as done by Horowitz *et al.* (2017). In this study (Experiment 2) we tested a different group of children with the same Picture Selection task used for scalar implicatures in Experiment 1 and compared it with a parallel task designed to test ad-hoc implicatures. This task is similar to the one used by Horowitz *et al.*'s (2017) and it was developed independently by us before the publication of that study. Differently from them, we did not include the negative quantifier *none* in our experiment for the concerns raised above. In addition to them, we also investigated the correlations between the rate of implicatures generated by each child and a set of standardized measures of cognitive and linguistic development (such as non-verbal IQ, lexical and grammatical abilities and Theory of Mind). So far, no study has systematically investigated the correlation between these factors and implicature computation in typically developing monolingual children. These analyses had two

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

purposes. First, from a developmental perspective, we aimed at understanding the developmental factors that may underpin children's performance in pragmatic tasks. Second, our study also aimed at comparing ad-hoc and scalar implicatures with respect to these factors: from a theoretical perspective, this analysis can foster the debate about the nature – and computational alleged differences - between types of inferences.

Experiment 1

In this experiment we employed two scalar implicature tasks administered to the same group of participants in different experimental sessions, a classical Truth Value Judgment task and a Picture Selection task. The aim of this study was twofold: first, we aimed at verifying whether the children that provided pragmatic responses in the Truth Value Judgment task were also consistently selecting the pragmatic target in the Picture Selection task, so as to validate the Picture Selection task as a sensitive task for scalar implicature derivation; second, we aimed at setting a baseline for our Experiment 2.

Method

Participants

Fifty-eight Italian children aged 4 and 5 were tested (age: 4;2-6;0; $M = 5;2$, $SD = 0;6$). Children were recruited from two kindergarten schools in Northern Italy and were tested after both parents signed a consent form for participation, in accordance with the requirements of the Ethical Committee of Trento University that approved this study.

Materials and procedure

With this first experiment our goal was to directly compare a Truth Value Judgment task with a Picture Selection task for scalar implicature computation. For this reason, we decided to maintain similar structural characteristics: both tasks were presented as games and were administered through a Power Point presentation on a laptop personal computer in a quiet room of the kindergarten. The tasks lasted approximately five minutes each. All the target sentences were pre-recorded to control for prosody. Children's responses were transcribed by the experimenter on an answer sheet. The tasks' order of presentation was randomized. In the case in which children were tired or distracted, the session was stopped.

Truth Value Judgment task for scalar implicatures

This task was adapted from Foppolo *et al.* (2012) and Katsos and Bishop (2011). Children were presented with a box and two characters that appeared on the screen: Davide, a boy, and Lucy, a foreign girl who was introduced as a learner of Italian. The children's task was to help Lucy to improve her Italian by judging her descriptions of different scenarios. For each trial, the scenario consisted of an array of six objects of the same kind (e.g., apples, dolls, hats); by means of his magic wand, Davide put either none, some (i.e., 4 out of 6) or all of the six objects inside the box. The child saw the objects moving from their initial position to the box,

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

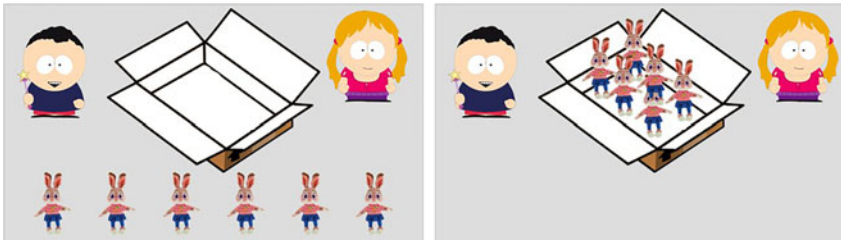

Figure 1. Example of an underinformative item for *some* for the Truth Value Judgment task. The left panel shows the initial array of object; the right panel shows the final outcome of Davide's action, after he moved the objects from the array to the box by means of his magic wand. Lucy's sentence in this trial was: "Davide put some of the bunnies in the box".

while hearing a 'magic' sound produced by Davide's wand triggering the movement. At the final stage, Lucy described what happened by saying "Davide put some of/all the S in the box" and children were asked to judge whether Lucy said things "right" or "wrong", and to correct her if she said something wrong. Please note that we used six objects in the Truth Value Judgment task to avoid that children used the description with numerals to reject the description in the some-TRUE condition: in this condition, the proportion used was 4/6 objects. If a smaller set of objects were used (e.g., five), the number of objects in the box in the case of some-TRUE would have been three, which falls within the subitizing range. This would have possibly favored a rejection of the description with *some* because the description with *three* would have been more appropriate (cf. Foppolo *et al.*, 2012 and Foppolo & Marelli, 2017 for a discussion of this issue).

To familiarize children with the task, the session started with two warm-ups in which a clearly true and a clearly false description of the scene were provided. The test phase comprised four target sentences with *some* used in an underinformative way (Some-UI) to describe a scenario in which Davide put all six objects in the box (Figure 1), and eight control items: four with the quantifier *all* (two true and two false), and four with the quantifier *some* (two true and informative, and two false). The experimental conditions are summarized in Table 1.

Picture Selection task for scalar implicatures

This task is a classical picture selection in which participants had to find the correct target – among 4 pictures – by exploiting a sentential clue; this task is modelled after Surian and Job (1987) and Stiller *et al.* (2015). In this task, children were introduced to a boy (Daniele) and were told that Daniele wanted to play a game with the children by giving them a clue to find specific objects or individuals on the screen. For example, children saw a scenario with 4 birthday cakes (Figure 2) and had to find the one that Daniele is addressing by hearing this clue: "Guess which one is my cake, I give you a clue: On my birthday cake, some of the candles are burning".

All the test sentences were previously pre-recorded to control for prosody. The test phase started after a warm-up trial that did not contain quantifiers and in which children were corrected in case of no answer or wrong answer. In the test phase, children were presented with four sentences containing *some*, two sentences containing *all* and one control sentence with no quantifier. Items were shown in a

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

Table 1. Description of the different types of items in the Truth Value Judgement task, including the sentence that children had to evaluate, the number of objects moved in the box and the expected answer.

Item type	Sentence	Objects in the box	Expected answer
Some-True	Davide put *some of* the S in the box	4 / 6	right
Some-False	Davide put *some of* the S in the box	0 / 6	wrong
Some-UI*	Davide put *some of* the S in the box	6 / 6	wrong (= scalar implicature) right (= no scalar implicature)
All-True	Davide put *all* the S in the box	6 / 6	right
All-False	Davide put *all* the S in the box	4 / 6	wrong

*Some-UI corresponds to the condition in which some is used underinformatively

Figure 2. Example of one of the items for *some* in the Picture Selection task: "Guess which one is my cake, I give you a clue: on my birthday cake, some of the candles are burning". Pictures in the display include: the pragmatic target (bottom-right), the underinformative competitor (bottom-left) and two distractors (top row).

pseudo-randomized order so that the quantifier *all* always appeared first in the sequence.

The structure of the scenario was kept constant across trials of the same kind. In the *all*-scenario, there was a target (for example, a garden with 5/5 red flowers), two competitors (for example, a garden with 3/5 red flowers and a garden with 3/5 blue flowers) and one distractor (a garden with no flowers). In the *some*-scenario, there were two types of distractors: for example, a cake with no candles and a cake with no burning candles (Figure 2, top row). There were also two types of possible targets: a pragmatic target, in which the array of objects was only compatible with the pragmatic meaning of *some,* namely *some but not all* (i.e., a cake with 3/5 burning candles, Figure 2, bottom-right picture) and an underinformative competitor, in which the array of objects was also compatible with the more informative quantifier (*all*) (i.e., a cake with 5/5 burning candles, Figure 2, bottom-left picture). We decided to use a set of 5 in this case in line with previous studies, also considering the fact that the *some*-TRUE condition was not tested in

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

this case (and a set of six objects for each of the four pictures would have rendered the scenario much more crowded). Note that participants who derived the scalar implicature should converge on the pragmatic target (and should do so consistently across trials); participants who stick to the logical interpretation (at least some) could opt for the underinformative competitor, but in principle they could also select the pragmatic target.

As discussed in Horowitz *et al.* (2017), this task was designed to enhance contextual relevance of the *all*-alternative: in fact, this alternative is both presented as a visual alternative in the scenario and it is also presented orally during the experiment. Moreover, the task is in principle simpler than the Truth Value Judgment task in that no metalinguistic judgment is required, thus lowering the computational resources required to solve the task. As highlighted before, on the other hand, when the Picture Selection task is presented in isolation the results are not easily interpretable: if the children select the underinformative competitor, this means that they are assessing the logical meaning of *some*, without deriving the implicature. In contrast, selections of the pragmatic target could be due to the derivation of the implicature (pragmatic responders) or, in principle, to the random choice between target and competitor for a logical responder. A direct comparison of the two tasks within the same group of children can shed some light on children's choices.

Results

In the Truth Value Judgment task, the responses on controls were coded as 'correct' if the child correctly accepted or rejected the true and false statements respectively. Responses on the test statements were coded as 'correct' if the child rejected the underinformative statement with *some* and mentioned *all* in their justification for rejection. This was to ensure that the base for their rejection was the activation of the scalar alternative and the consequent derivation of the pragmatic implicature. In the Picture Selection task, the responses were coded as 'correct' if the child selected the target picture for *all* and the pragmatic-target for *some*.

The two tasks yielded similar results: while children's accuracy on controls was above 94% in both tasks (94.8% in the Picture Selection task and 94.2% in the Truth Value Judgment task), children were not adult-like in deriving the scalar implicatures, regardless of the task: they rejected the underinformative *some* sentences in the Truth Value Judgment task 55.6% of the times, and they selected the pragmatic target in the Picture Selection task 57.6% of the times (Figure 3). By inspecting the cases in which children did not select the pragmatic target in the Picture Selection task, we found that they selected the underinformative competitor (for example, the cake on which all the candles were burning) more than 92% of the times.

Following Skordos and Papafragou (2016), we coded children as 'passers' or 'failers' if they consistently (i.e., at least 3 out of 4 times) accepted or rejected the underinformative *some* sentences in the Truth Value Judgment task or if they consistently selected the pragmatic or the non-pragmatic target in the Picture Selection task. As attested in other studies (Guasti, Chierchia, Crain, Foppolo, Gualmini & Meroni, 2005), children displayed a bimodal distribution: 25 (43%) children were 'failers' and 33 (57%) were 'passers' in the Truth Value Judgment task, and none of them showed an inconsistent pattern of responses (2 rejections and 2 acceptance); 30 (52%) children were 'passers' and 28 (48%) were 'failers' in the Picture Selection task. Moreover, 43 (74%) children were consistently 'passers' or

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

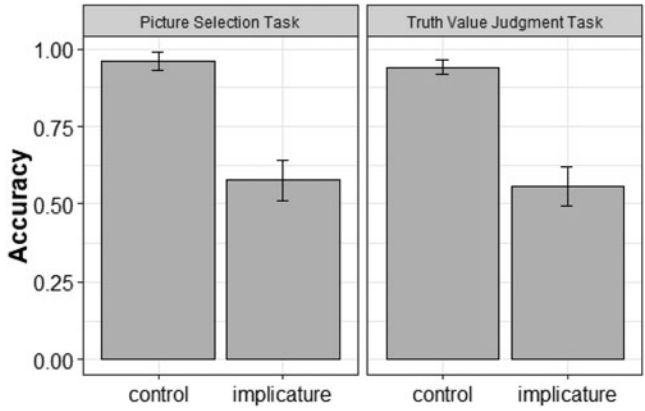

Figure 3. Children's proportion of accuracy on controls and underinformative *some* across tasks.

Table 2. Distribution of passers and failers in the Picture Selection task and in the Truth Value Judgement task (Experiment 1).

		Picture Selection task	
		passers	*failers*
Truth Value Judgement task	*passers*	24	9
	failers	6	19

'failers' in both tasks, as reported in Table 2. A McNemar test comparing the distribution of passers and failers revealed no significant difference between the two tests ($\chi^2(1) = .267$, p = .606). Also, we found a highly significant correlation between the two tasks in the proportion of children's pragmatic responses ($r = .48$, t = 4.0496, $df = 56$, $p < .001$).

We ran a series of logistic regression analyses considering accuracy as the dependent variable. For a better fit of the model, all the variables were centered prior to analysis. Following Horowitz *et al.* (2017), we started with the most complete model that converged and adopted a backward stepwise model selection-procedure to eliminate all the non-significant factors and interactions by comparing each of the models with the ANOVA function in R to select the best model fit. Please note that adding random slopes for subjects and items always resulted in convergence failure, and thus these were never included. Consistently with previous analyses, task (Picture Selection task vs. Truth Value Judgment task) was not found to be a significant predictor and did not significantly improve the model fit, and thus was removed from the final model. The model that best fits the data included subjects and items as random factors, Type (implicature vs. controls) and Age in months as fixed effects, as well as their interactions. This model revealed a significant effect of Type (Est = -4.136, $SE = .580$, $z = -7.129$, $p < .0001$) and Age (Est = .153 $SE = .041$, $z = 3.775$, $p < .001$): accuracy on implicatures is significantly lower than controls and performance is worse for younger children.

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

Discussion

Children showed excellent knowledge of the semantics of the quantifiers involved in the scale: they were at ceiling in the Truth Value Judgment task in judging sentences with *some* and *all* in which these quantifiers applied to true and false situations (cf. Foppolo *et al*., 2012: Experiment 6 for a similar result). They were also at ceiling in selecting the correct target in the *all* sentences in the Picture Selection task. Nonetheless, children's performance was less than optimal in judging underinformative *some* sentences, independently of the task. In fact, we demonstrated that the two tasks were highly correlated, and the type of task did not significantly affect children's accuracy.

Our study is the first one to compare the Truth Value Judgment task and the Picture Selection task directly on the same group of children: thus, the lack of a significant difference between the two tasks helps to clarify an interesting methodological point. As highlighted above, children's selection of the pragmatic target in the Picture Selection task in response to *some* sentences could in principle be due to the computation of the implicature or to the random choice of the logical interpretation of the quantifier (at least some). The high correlation of children's performances across tasks ensures that the two tasks tap the same pragmatic ability: in most cases, the children that were categorized as "pragmatic responders" in the Truth Value Judgment task (i.e., those who rejected the underinformative *some*) were also those who selected the pragmatic target in the Picture Selection task; and the "logical" children who accepted the underinformative *some* in the Truth Value Judgment task were also those who selected the underinformative competitor in the Picture Selection task.

Once we attested that what were considered as 'pragmatic answers' in the two tasks do in fact identify the pragmatic responders, i.e., children who derive the implicature, the fact that children's performance on scalar implicatures did not differ significantly across the two tasks can also shed some light on the role of alternatives in the computation of pragmatic inferences, and the role of task. First of all, our results show that the Truth Value Judgment task, in which children have to listen to a story and have to provide a metalinguistic (binary) judgment, does not prove more difficult than a Picture Selection task. Furthermore, no difference was revealed between the Picture Selection task and the Truth Value Judgment task in the rate of derivation of scalar implicature, despite the fact that in the Picture Selection task the child simply had to select the image that represents the pragmatic interpretation of *some* in a scenario in which the scalar alternative *all* was provided as a visual alternative. *Contra* Skordos and Papafragou (2016), this factor does not seem to boost the derivation of scalar implicatures: regardless of task, children's performance with scalar implicatures was far from optimal.

However, we should highlight the fact that the tasks employed require an off-line judgment and are based on accuracy measures. It would be an important goal for the future to assess whether a difference in the steps or in the process of the derivation of the pragmatic inference between tasks might be found when on-line measures of processing, such as eye-tracking, are used.

As set out in the introduction, the main goal of this first experiment was that of assessing the validity of using the Picture Selection task to test scalar implicatures in order to compare children's performance with scalar and ad-hoc implicature within the same group of children by means of this task. This has been done in Experiment 2.

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

Experiment 2

The aim of the second experiment was to compare children's derivation of scalar and ad-hoc implicatures using the same paradigm, and to link children's performances in these pragmatic tasks with other measures of cognitive and linguistic development. The Picture Selection task has already been used and validated for ad-hoc implicatures, and our first experiment confirmed that this task is also sensitive to children's ability to derive scalar implicatures. For these reasons, we decided to use the Picture Selection task for both ad-hoc and scalar implicatures in Experiment 2.

Method

Participants

A different group of 141 Italian children aged 3 to 9 were tested (3;8-9;2, $M = 6;2$, $SD = 1;5$). Of these, 75 children were enrolled in the kindergarten (3;8-6;0, $M = 5;1$, $SD = 0;10$), and 66 in the primary school (grade 1st to 3rd; 6;2-9;2, $M = 7;5$, $SD = 0;10$). Children were recruited in different kindergartens and primary schools in Northern Italy. All children were tested after both parents signed a consent form for participation.

Materials and procedure

Children were administered two Picture Selection tasks: one was the same Picture Selection task used for scalar implicatures in Experiment 1. The second was a novel Picture Selection task designed for ad-hoc implicatures, modeled after Surian and Job (1987) and adapted from Stiller *et al.* (2015). Like the Picture Selection task for scalar implicatures, children had to point at the correct target – among 4 pictures – by exploiting a sentential clue. The story is similar to that for scalar implicatures: participants should follow Daniele's hint to find the correct referent of his expression. For example, Daniele says "Guess which one is my bed, I give you a clue"; then the four pictures appear on the screen and Daniele says: "On my bed there is a teddy bear". On the basis of this sentence the child has to find the correct target among the four (Figure 4). As for scalar implicatures, the scenario displays two distractors (i.e., an empty bed and a bed with a penguin on it, Figure 4, bottom row) and two potential targets: the pragmatic target (i.e., the bed with only the teddy bear, Figure 4, top-right) and the underinformative target (i.e., the bed with a teddy bear and a penguin, Figure 4, top-left). As for scalar implicatures, if the ad-hoc implicature is computed the children should select the bed with the teddy bear alone, under the reasoning that, if Daniele wanted them to point to the other bed (the one with the teddy bear *and* the penguin) he should have referred to that bed by explicitly mentioning both things. The fact that he didn't should entitle the listener to derive the inference that the referent of the request is the bed with only the teddy bear on it.

After a warm-up item that did not involve implicatures, children were presented with four ad-hoc implicatures and a control sentence in a pseudo-randomized order. Note that in the case of the scalar implicature task, we needed to control that children correctly understood the sentence with *all*, and for this reason two additional control items were present in the scalar task that are not present in this task.

In addition to the two Picture Selection tasks for testing implicatures, we wanted to measure some possible factors that could be correlated with the ability to derive

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

Figure 4. Example of one of the items in the Picture Selection task for ad-hoc implicatures: "Guess which one is my bed, I give you a clue: on my bed there is a teddy bear". Pictures in the display include: the pragmatic target (top-right), the underinformative competitor (top-left) and two distractors (bottom row).

pragmatic inferences of different kinds. To this purpose, children were also tested with a battery of standardized tests: the Raven Coloured Progressive Matrices (CPM, Italian standardization by Belacchi, Scalisi, Cannoni & Cornoldi, 2008) to test for non-verbal IQ; the test for Lexical comprehension and the test for Grammatical comprehension taken from the Batteria di Valutazione del Linguaggio (Marini, Marotta, Bulgheroni & Fabbro, 2015) to test for receptive vocabulary and morphosyntactic abilities; and a series of Theory of Mind (ToM) tasks adapted in Italian from the first four tasks of Wellman and Liu (2004) to test metarepresentational abilities up to 1st order ToM. Our aim was to verify whether cognitive skills (as measured by Raven CPM and by ToM tasks) and/or linguistic (lexical and morphosyntactic) abilities were correlated with none, only one or both types of implicatures, with the goal of identifying possible predictors of pragmatic inferencing on the one hand, and of clarifying which underlying mechanisms are common, and which are different, in the derivation of scalar and ad-hoc implicatures.

The tasks were administered by an experimenter in a quiet room of the kindergarten or of the school, after children were familiarized with the experimenter and with the laptop personal computer. The test session lasted approximately 40 minutes. In the case that children were tired, the tasks were administered in different days. This study is part of a more extensive research on pragmatic comprehension in typically and atypically developing children that has been approved by the University of Trento's Ethical Committee.

Results

As in the previous Picture Selection task study, the responses were coded as 'correct' if the child selected the pragmatic target; in the case of *some*, this corresponded to the picture in which only some of the objects were affected (for example, the bottom right picture in Figure 2); in the case of ad-hoc implicatures, this corresponded to the picture in which only the object mentioned was present (for example, the top right picture in Figure 4). In all the other cases, the answer was coded as 'incorrect'.

Children's accuracy on controls was above 95% in both tasks (99% in the ad-hoc implicature task and 95% in the scalar implicature task; Figure 5, left panel); overall,

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

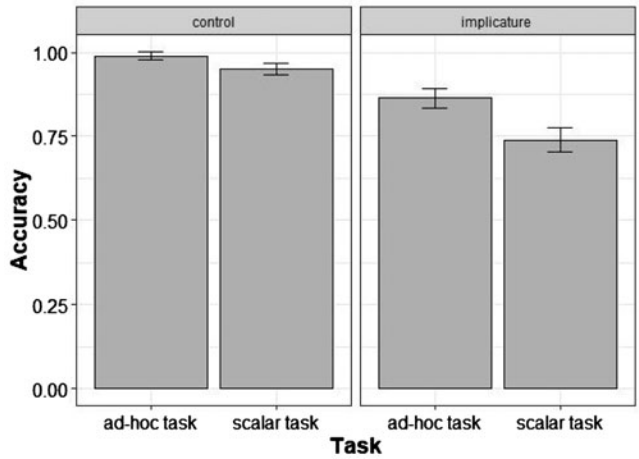

Figure 5. Children's proportion of accuracy on controls (left panel) and implicatures (right panel) in the two tasks (ad-hoc vs. scalar implicature Picture Selection task).

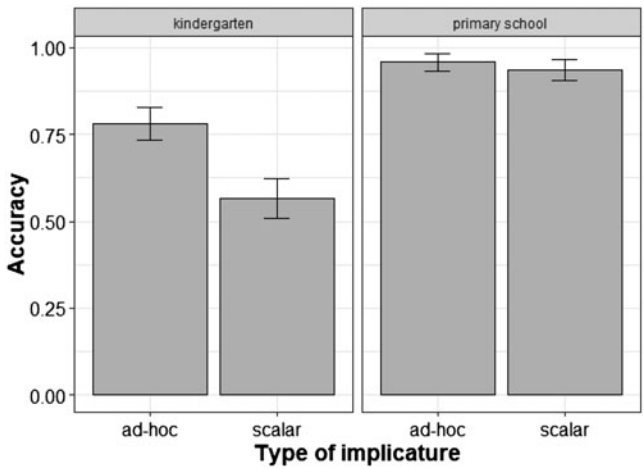

Figure 6. Children's proportion of accuracy on the two types of implicatures (ad-hoc vs. scalar implicatures) across school groups (pre-school children, $N = 75$, left panel, and primary school children, $N = 66$, right panel).

the rate of derivation of ad-hoc implicatures was higher than scalar implicatures (86% vs. 74% respectively; Figure 5, right panel).

The gap between children's performance with ad-hoc and scalar implicature is even more evident if we split children by Schooling, as in Figure 6. As is shown in the graph, the kindergarten children struggle more with scalar than ad-hoc implicatures (57% vs. 78%), while the older children show parallel performance in both (94% for scalar implicatures and 96% for ad-hoc implicatures). No difference is observed for controls across age groups and tasks (in all cases, performance is above 90%). It is worth noting that we fully replicated the results from Experiment 1 for scalar implicature for the younger children.

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

Table 3. Distribution of passers and failers in the scalar implicature task and the ad-hoc Implicature task (Experiment 2).

		ad-hoc implicature task	
		passers	*failers*
scalar implicature task	*passers*	99	4
	failers	21	17

As observed in the previous study, the overall majority of the incorrect responses corresponded to the selection of the underinformative target (around 90% of the times) and the majority of the children provided consistent responses in the underinformative trials (87% in the ad-hoc task and 84% in the scalar implicature tasks), either consistently selecting (or failing to select) the pragmatic target.

As in Experiment 1, we coded children as 'passers' or 'failers' to check for their distribution across tasks (Table 3). As is evident, although the majority of children either pass or fail in both tasks, 21 children (15%) fail in the scalar but succeed in the ad-hoc task, while only 4 children (less than 3%) fail in the ad-hoc but succeed in the scalar task. A McNemar test revealed a significant difference in the proportion of passers and failers in the two tasks ($\chi^2(1) = 10.24$, $p < .01$), showing that the ad-hoc implicatures task was reliably easier than the scalar implicatures task.

We first ran a series of logistic regressions considering accuracy as the dependent variable, subjects and items as random factors, and age, condition (implicatures vs. controls) and task (scalar vs. ad-hoc), as well as their interactions, as fixed effects. For a better fit of the model, all the variables were centered prior to analysis; as before, we followed a backward stepwise model selection-procedure; following this procedure, all the interactions were removed because they did not significantly improve the model fit. The model that best fits the data reveals: a significant effect of age (accuracy is significantly higher for older children; Est = .082, $SE = .010$, $z = 8.170$, $p < .0001$), a significant effect of condition (accuracy on implicatures is significantly lower than accuracy on controls; Est = -1.524, $SE = .506$, $z = -3.011$, $p < .01$) and a significant effect of task (accuracy is significantly lower for the scalar than the ad-hoc task; Est = -1.542, $SE = .506$, $z = -3.0311$, $p < .001$).

Two questions remain for further investigation: first, which factors eventually contribute to modulate children's success or failure with implicatures? Provided that a difference was revealed between types of implicatures, are these modulated by different abilities? To answer these questions, we focused on implicature items, excluding controls.

We ran a series of logistic regression analyses considering accuracy (on the implicature condition only, as we said) as the dependent variable, subjects and items as random factors, and age, type of implicature (ad-hoc vs. scalar), morphosyntax, lexicon, non-verbal IQ and 1st order ToM as fixed effects, as well as their interactions. For a better fit of the model, all the variables were centered prior to analysis; as before, following a backward stepwise elimination procedure, all the factors and interactions that resulted as non-significantly contributing to the best fit of the model were removed. The final model revealed a significant effect of age (Est = .055, $SE = 0.012$, $z = -4.368$, $p < .0001$), type of implicature (accuracy on scalar

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

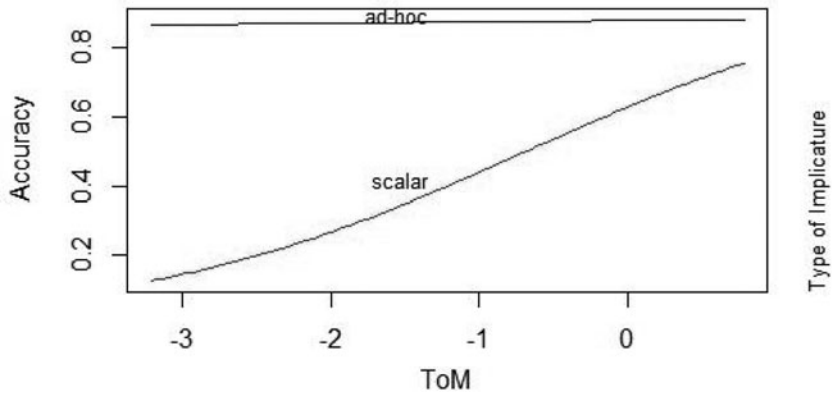

Figure 7. Plot of partial effects of the mixed-effects model fit tested above, considering pre-school children's performance in the two scalar tasks separately and children's levels of ToM as predictor.

implicature was lower than accuracy on ad-hoc implicatures: Est = -1.524, $SE = .685$, $z = -2.223$, $p < .05$), morphosyntax (accuracy is significantly higher for children with higher scores; Est = .085, $SE = .021$, $z = 4.118$, $p < .0001$).

To better understand which factors modulate children's performance at a stage of development in which most of the children do not fully master the ability to derive implicatures – namely, during pre-school age – and what modulates their performance when this ability is almost fully mastered – namely, after age 6 – we further split the groups depending on schooling and performed a series of logistic regression analyses following the same procedure explained above. The final model on the group of primary school children revealed a significant effect of morphosyntax (accuracy is significantly higher for children with higher scores; Est = .245, $SE = .0653$, $z = 3.760$, $p < .001$), and a marginally significant effect of non-verbal IQ (Est = .066, $SE = .037$, $z = 1.769$, $p = .077$) and lexicon (Est = -2.972, $SE = 1.796$, $z = -1.655$, $p = .098$). No significant difference between ad-hoc and scalar implicatures emerged in this case (Est = -1.340, $SE = 1.141$, $z = -1.175$, $p = .240$).

The final model on the group of pre-school children revealed a significant effect of type of implicature (accuracy on scalar implicature was lower than accuracy on ad-hoc implicatures: Est = -1.473, $SE = .637$, $z = -2.313$, $p < .05$), morphosyntax (accuracy is significantly higher for children with higher scores; Est = .070, $SE = .0189$, $z = 3.726$, $p < .001$), and a significant interaction between type of implicature and ToM: despite not being significant as a factor (Est = .036, $SE = .240$, $z = .150$, $p = .88$), ToM resulted in significantly modulating accuracy on scalar, but not ad-hoc implicatures (Est = .733, $SE = .262$, $z = 2.793$, $p < .01$). This is evident in Figure 7: while performance on ad-hoc implicatures does not depend on levels of ToM (even children with low ToM scores have a good performance with ad-hoc implicatures), performance with scalar implicatures does positively correlate with ToM scores, as plotted in Figure 7.

Our data suggest that these two types of implicatures might possibly rely on different abilities, that are not yet fully developed in young children. To further inspect these results, we performed two separate Focused Principal Component Analyses on scalar and ad-hoc implicatures (Figure 8) in pre-school children. These analyses revealed some interesting patterns. First of all, the ToM skills, as measured in our tasks, seem to be independent from linguistic abilities and non-verbal IQ, as they located

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

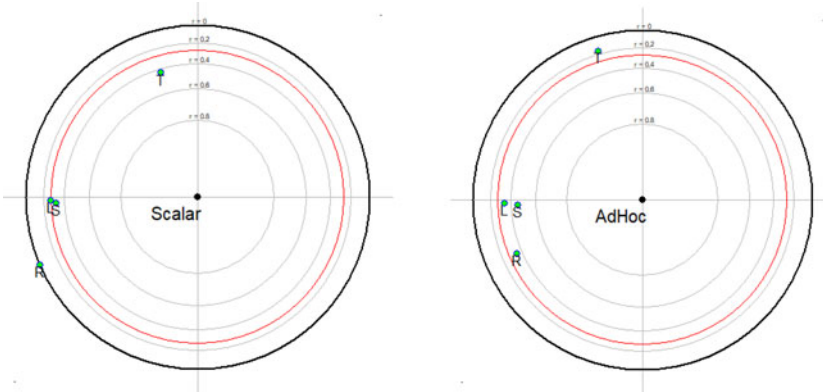

Figure 8. Output of the Focused Principal Component Analysis for scalar implicatures (left panel) and ad-hoc implicatures (right panel) in pre-schoolers considering ToM (T), lexicon (L), syntax (S) and non-verbal IQ (R). Variables that lay within the smaller radius represent the stronger correlations with the variable of interest displayed in the center. The red circle delimits statistical significance at the 5% level.

orthogonally from these factors. Second, the factor that has a higher correlation with scalar implicatures is ToM ($r = .04$), and the factor that correlates less is non-verbal IQ ($r < .02$). Third, this pattern is different from ad-hoc implicatures, given that in this case both non-verbal IQ and linguistic abilities seem to be the most correlated factors ($r = .04$ for non-verbal IQ and Lexicon; $r = .06$ for Morphosyntax). These descriptive analyses reveal an interesting difference between the two types of implicatures, which goes beyond the fact that children performed differently with them. Such analyses also reveal that the underlying factors of these two kinds of implicatures seem to differ.

Discussion

Children showed an overall excellent performance on controls, and an overall good performance on ad-hoc implicatures. In fact, the group of older children showed an optimal performance on both types of implicatures. The younger children, instead, behaved differently in the two tasks. First, they did not encounter difficulties in detecting the pragmatic target in the case of inferences that were built on a contextual basis, as is the case for ad-hoc implicatures. Second, they struggled more to detect the pragmatic target for the scalar quantifier, and almost half of the children consistently failed this task, pointing to the underinformative target for *some*. These results are in line with those found by Horowitz *et al.* (2017) in a similar comparison between scalar and ad-hoc implicatures in pre-school aged children.

Beyond their results, we investigated this difference in a developmental perspective, showing that children up to age 6 have more difficulties with scalar implicatures than ad-hoc implicatures, while this difference disappears after age 6, when their performance becomes optimal in both kinds of implicatures. This result is also in line with previous developmental findings about the *some but not all* implicature (Foppolo *et al.*, 2012: Experiment 1).

The correlation analyses showed a difference between scalar and ad-hoc implicatures in the factors underlying children's success. Children's morphosyntactic competence

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

positively predicted their ability to generate both kinds of pragmatic inferences. On the contrary, the measures of ToM were found to modulate children's success with scalar inference, while no impact of ToM on the ability to generate ad-hoc inferences was found.

General discussion

In this paper we tested pre-school and school-aged children's ability to compute generalized (scalar) implicatures in different tasks (Experiment 1) and compared their performance on particularized (ad-hoc) and scalar implicatures (Experiment 2). We also looked for possible correlations between the ability to derive implicatures and other cognitive and linguistic measures (Experiment 2).

The present results are consistent with some previous findings in showing that pre-school aged children have difficulties in deriving pragmatic inferences, especially those involving a scalar quantifier, and that their performance is better with those inferences in which alternatives are construed ad-hoc, on the basis of the context (Horowitz *et al*., 2017). Extending previous findings, we also characterized the developmental trajectory of the two types of implicatures: we show that preschoolers perform better with ad-hoc than scalar implicatures; we also show that the delay of the latter over the former kind of implicature is confined to a specific stage in development, namely pre-school age, as both ad-hoc and scalar implicatures reach an optimal performance in school-age.

We also made a methodological contribution by showing that the Truth Value Judgment task, traditionally employed to investigate children's pragmatic ability and recently criticized, prompts a rate of pragmatic responses comparable to the (allegedly) less demanding Picture Selection task in which alternatives are presented visually and no metalinguistic judgment is provided. In Experiment 1 we found no difference between the Truth Value Judgment task and the Picture Selection task in the rate of derivation of scalar implicatures: in both tasks, children's access to the strengthened meaning of *some* was around 55%, and children were bimodally distributed (Guasti *et al*., 2005).

We finally analyzed the impact of different linguistic (lexicon and morphosyntax) and cognitive measures (non-verbal IQ and ToM) on the rate of pragmatic responses and showed a different pattern of dependencies in the two types of inferences investigated. In particular, Experiment 2 showed evidence for a significant contribution of morphosyntactic abilities in the derivation of pragmatic inferences in general, and a positive correlation between ToM abilities and scalar, but not ad-hoc, implicatures in the younger group.

As we outlined in the introduction, children's difficulty with pragmatic inferences is a robust finding in the acquisition literature. What is still debated is the source of this difficulty: under some accounts this difficulty is believed to be more grounded in a yet immature linguistic representation of the scale (Barner *et al*., 2011; Foppolo *et al*., 2012); in others, it is associated with yet immature pragmatic or processing systems that make children more tolerant than adults when facing pragmatic anomalies (Katsos & Bishop, 2011), or less prone to detect what is contextually relevant or optimal in a given exchange (Skordos & Papafragou, 2016).

We believe that the comparison between ad-hoc and scalar implicatures in Experiment 2 can shed some light on the disagreements between LEXICALIST, PRAGMATIC and PROCESSING accounts of children's failure with pragmatic inferences. Certainly, it seems hard to reconcile the difference shown between ad-hoc

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

implicatures and scalar implicatures within a PROCESSING account. If the mechanisms involved in the generation of all kinds of implicatures rest on the same basis, they should require, in principle, similar processing costs, and thus children's asymmetrical competence is unexpected. Similarly, a difference between different kinds of implicatures is not readily predicted by a PRAGMATIC (TOLERANCE) account: if children's pragmatic system were yet immature, the consequence would be a failure in all kinds of pragmatic inferences, unless one assumes that the mechanisms involved in the generation of the inference, or in the retrieval of the alternatives, are not equivalent across types of triggers. In all cases, it seems hard to reconcile these alleged differences in the processes or mechanisms underlying pragmatic inference without letting the nature of the scale involved in the computation enter the picture.

Conversely, this difference is straightforwardly accounted for within a LEXICALIST approach to scalar implicatures: according to the latter, some children might have reached the stage in which they CAN compute all the steps involved in the derivation of pragmatic inferences, but they might have problems in accessing the scalar alternatives in some cases. In particular, problems are predicted in those cases in which the scale requires an *a priori* lexicalization, a process that might take some time. This account would explain why younger children show a good performance with ad-hoc implicatures but not with scalar implicatures: only the former do not require a preliminary lexicalization of the scale, as alternatives are directly retrieved from the context. The scale, instead, needs to be in place before enriching the meaning of *some* as *some but not all*. This represents a viable explanation of the reason why scalar alternatives are costlier than contextually retrieved alternatives, and it emphasizes the linguistic nature of scalar, but not ad-hoc, implicatures.

The analyses of the factors involved in the two types of inferences also might reflect a difference between the two types of inferences, thus adding to the debate about the steps required in pragmatic inferencing in general, and about the nature of the differences across different kinds of implicature. Two main findings are worth discussing with respect to these results. One is the correlation between children's linguistic competence and their general pragmatic proficiency reflected in the derivation of both types of implicatures. This link is somehow expected, given that the ability to derive inferences in communicative exchanges is part of the competence of a skilled interlocutor, which develops over time, in tandem with the individual's linguistic proficiency. This finding is also consistent with the fact that "higher-level" linguistic competence, such as morphosyntax, affects the rate of pragmatic inference, and not mere lexical knowledge which might reflect more basic linguistic skills. Of particular interest is the second finding, that reveals that ToM modulates children's performance on scalar implicatures but it does not modulate children's accuracy with ad-hoc implicatures

The role of ToM in pragmatic abilities is at present hotly debated. Some authors consider ToM as a necessary precursor of pragmatic language skills (Happé, 1993; Winner, Brownell, Happé, Blum & Pincus, 1998), while others have argued that a proficient ToM results from exposure to (and experience with) pragmatic language (Peterson and Siegal, 2000). In a recent review, Bosco, Tirassa and Gabbatore (2018) report that ToM is shown to correlate with pragmatic ability in some studies, while in other studies ToM alone seems not to be the only relevant factor to explain the empirical differences of performance across different kinds of pragmatic tasks, thus suggesting the involvement of another, at least partially different, faculty, in pragmatic success. At the same time, interesting findings were revealed from the

https://doi.org/10.1017/S030500092000032X Published online by Cambridge University Press

investigation of scalar implicatures in atypical populations with presumably impaired ToM abilities. In a recent study, Hochstein, Bale and Barner (2018) measured high-functioning adolescents with Autism Spectrum Disorder (ASD) in their ability to compute scalar and ignorance implicatures and to reason about the epistemic states of their interlocutors. They found that ASD did not differ from typical young adults in the rate of derivation of conversational implicatures, although they failed to spontaneously consider speakers' specific epistemic states when actually computing such implicatures. A preliminary conclusion drawn by these data is that epistemic reasoning is not strictly necessary for deriving all scalar implicatures: as the authors concluded, "epistemic reasoning may not be a core, constitutive component of scalar implicature".

In another study by Schaeken, Van Haeren and Bambini (2018), ten-year-old children with ASD were tested with scalar and ad-hoc implicatures and were found to provide more pragmatic answers than the group of children matched for IQ. The authors concluded that verbal intelligence and mental flexibility might have a greater role than theory of mind in pragmatic inferencing. Testing a different population, a negative correlation was found between the number of logical responses in a sentence evaluation task for scalar implicatures and the scores on a *faux-pas* test administered to adults with schizophrenia (Wampers, Schrauwen, De Hert, Gielen & Schaeken, 2018). The authors, however, were skeptical in interpreting this observed correlation as truly revealing a direct causal relation between these abilities. Indeed, they were more prone to interpret it as the result of an underlying core cognitive process that determines both ToM ability and the ability to derive scalar implicatures, and that might be impaired in this population (such as low working memory, weak central coherence, or formal thought disorder).

Along similar lines, the fact that ToM was not found to modulate children's accuracy with ad-hoc implicatures in our study might suggest that ToM is not a relevant factor in the derivation of pragmatic inferences, unless one argues that the pragmatic competence required to derive different kinds of pragmatic inferences relies on different mechanisms. At the same time, the fact that it modulates children's derivation of scalar implicatures might be interpreted as a mere consequence of the fact that scalar implicatures, like ToM abilities, take time to be acquired. We acknowledge the fact that this interpretation is merely speculative. However, we would like to emphasize that a growing literature points out that infants may possess mental state reasoning skills that are revealed in non-verbal spontaneous response tasks (e.g., Surian, Caldi & Sperber, 2007; for a review Baillargeon, Scott & He, 2010), but do not emerge in verbal-elicited responses to ToM tasks, like those used in the present study. Moreover, limitations in executive functions are likely to be the source of this task dissociation (e.g., Liu, Wu, Wu, Li, Cai & Liu, 2018).

Thus, the significant association between explicit ToM tasks and the scalar, but not the ad-hoc, implicature tasks might suggest that the former type of implicature imposes greater demands on executive components than the ad-hoc implicature task, compatibly with what has been previously found, for example, by Marty and Chemla (2013). This is quite likely considering that the alternatives in the case of ad-hoc implicatures are retrieved directly through relevant contextual information, and thus only rely on bottom-up process, whereas scalar implicatures require the integration of bottom-up (context) and top-down information (lexicalized scalar alternatives) for their generation. This might also explain why the rate of scalar implicatures was not higher in the Picture Selection task than the Truth Value Judgment task, as shown

by Experiment 1, despite the fact that scalar alternatives were visually available in the former but not in the latter. Maturational factors that underlie the growth of executive components of elicited responses in ToM tasks may also underlie improvements in scalar implicatures tasks and this would account for the association found in the present study. Consistently with previous accounts, we suggest that ToM scores might simply capture the maturational level that is required to master scalar implicatures, which is higher than that required to master ad-hoc implicatures since only the former kind of implicature requires a step in which alternatives are lexicalized, while an *a priori* knowledge of alternatives is not required for the latter kind. These conclusions, however, are to be taken with caution, since they rest on a unitary account of pragmatic inferencing, which is still a topic of debate (Domaneschi & Bambini, 2020), especially in light of the fact that in some studies the rate of scalar implicatures is found to be higher than ad-hoc implicatures (Schaeken *et al*., 2018).

All in all, although further research is needed for a deeper understanding of the faceted phenomenon of pragmatic inferences and the processes underlying it, we argue that our results are best captured by a LEXICALIST approach to scalar implicatures: children know how to play the game of being cooperative interlocutors, and they can do so effectively, provided that they know the TERMS involved in the game.

Acknowledgements. This work has been partly supported by a grant from the Fondazione ONLUS Marica De Vincenzi to Greta Mazzaggio. We are grateful to the children and families who took part in our research and to the teachers and directors of the schools for their availability.

References

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

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

Belacchi, C., Scalisi, T. G., Cannoni, E., & Cornoldi, C. (2008). *CPM coloured progressive matrices: standardizzazione italiana: manuale*. Firenze, Italy: Giunti OS.

Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018). Why pragmatics and Theory of Mind do not (completely) overlap. *Frontiers in Psychology*, *9*, 1453.

Bredart, S. (1984). Children's interpretation of referential ambiguities and pragmatic inference. *Journal of Child Language*, *11*, 665–672.

Chemla, E., & Singh, R. (2014a). Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass*, *8*(9), 373–386.

Chemla, E., & Singh, R. (2014b). Remarks on the experimental turn in the study of scalar implicature, Part II. *Language and Linguistics Compass*, *8*(9), 387–399.

Chierchia, G. (2013). *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). *The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures*. In Proceedings from the Annual Boston University Conference on Language Development, 25, 157–168.

Domaneschi, F., & Bambini, V. (2020). *Pragmatic Competence*. To appear in Fridland, E. & Pavese, C. (Eds.). Routledge Handbook of Skill and Expertise. Springer.

Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar implicatures in child language: Give children a chance. *Language Learning and Development*, *8*, 365–394.

Foppolo, F., & Marelli, M. (2017). No delay for some inferences. *Journal of Semantics*, *34*(4), 659–681.

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

Guasti, T.M., 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.

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119.

Hochstein, L., Bale, A., & Barner, D. (2018). Scalar implicature in absence of epistemic reasoning? The case of autism spectrum disorder. *Language Learning and Development*, *14*(3), 224–240.

Horn, L. (1972). *On the semantic properties of the logical operators in English*. Ph.D. dissertation, UCLA, Los Angeles, CA.

Horowitz, A. C., Schneider, R. M., & Frank, M. C. (2017). The trouble with quantifiers: Exploring children's deficits in scalar implicature. *Child Development*, *8*(6), e572–e593.

Jackson, S., & Jacobs, S. (1982). Ambiguity and implicature in children's discourse comprehension. *Journal of Child Language*, *9*, 209–216.

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

Liu, M., Wu, L., Wu, W., Li, G., Cai, T., & Liu, J. (2018). The relationships among verbal ability, executive function, and theory of mind in young children with cochlear implants. *International Journal of Audiology*, *57*(12), 881–888.

Marini, A., Marotta, L., Bulgheroni, S., & Fabbro, F. (2015). *Batteria per la valutazione del linguaggio in bambini dai 4 ai 12 anni*. Firenze: Giunti OS.

Marty, P. P., & Chemla, E. (2013). Scalar implicatures: working memory and a comparison with only. *Frontiers in Psychology*, *4*, 403.

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

Peterson, C. C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language*, *15*(1), 123–145.

Pouscoulous, N., Noveck, I., Politzer, G., & Bastide, A. (2007). Processing costs and implicature development. *Language Acquisition*, *14*, 347–375.

Schaeken, W., Van Haeren, M., & Bambini, V. (2018). The understanding of scalar implicatures in children with autism spectrum disorder: dichotomized responses to violations of informativeness. *Frontiers in Psychology*, *9*, 1266.

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

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition* (2nd ed.) Cambridge, MA: Harvard University Press.

Stiller, A., Goodman, N. D., & Frank, M. C. (2015). Ad hoc implicature in preschool children. *Language, Learning and Development*, *11*(2), 176–190.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586.

Surian, L., & Job, R. (1987). Children's use of conversational rules in a referential communication task. *Journal of Psycholinguistic Research 16*, 369–82.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2015). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, 1–30.

Wampers, M., Schrauwen, S., De Hert, M., Gielen, L., & Schaeken, W. (2018). Patients with psychosis struggle with scalar implicatures. *Schizophrenia Research*, *195*, 97–102.

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

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language*, *62*(1), 89–106.

Cite this article: Foppolo F, Mazzaggio G, Panzeri F, Surian L (2021). Scalar and ad-hoc pragmatic inferences in children: guess which one is easier. *Journal of Child Language* **48**, 350–372. https://doi.org/10.1017/S030500092000032X