Language Learning and Development, 8: 365–394, 2012

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475441.2011.626386

Psychology Press
Taylor & Francis Group

Scalar Implicatures in Child Language: Give Children a Chance

Francesca Foppolo and Maria Teresa Guasti
Department of Psychology, University of Milano-Bicocca

Gennaro Chierchia
Department of Linguistics, Harvard University

Children's pragmatic competence in deriving conversational implicatures (and scalar implicatures in particular) offers an intriguing standpoint to explore how developmental, methodological, and purely theoretical perspectives interact and feed each other. In this paper, we focus mainly on developmental and methodological issues, showing that children from age 6 on are adult-like in deriving the scalar implicature related to the scalar quantifier *some* (i.e. they interpret *some* as *some but not all*), while children at age 4 and 5 only sometimes reject underinformative-*some* in a classical Truth Value Judgment Task (Experiment 1). They do so despite their excellent performance in pragmatic tasks that evaluate their competence with the rules of talk exchange, such as the Conversational Violations Test (Experiment 4) and the Felicity Judgment Task (Experiment 5). To give children a better chance to reject underinformative-*some* when *all* is at stake, we manipulated the experimental design and materials in three different ways: 1) in Experiment 2, we tested the partitive *alcuni dei* (*some of*) instead of the bare quantifier *qualche* (*some*); 2) in Experiment 3, we attempted to prime the scale <some, all> by asking children to judge a correct statement with *all* before the critical underinformative statement with *some*; 3) in Experiment 6, we aimed at making children more aware of the ambiguity of *some*, between its basic meaning (*at least some, possibly all*) and its strengthened meaning (*some but not all*). A surprising improvement is recorded in the last experiment, in which the rejection of underinformative-*some* by 5-year-old children rose to 72.5% (it was 42% in Experiment 1). We suggest that the children's low performance with scalar inference might be linked to the interplay of different factors as in the development of other general cognitive abilities, such as the ability to change one's strategy (Shallice, 1982) or to shift one's perspective (Gopnik & Rosati, 2001), the maturation of the lexicon (Barner & Bachrach, 2010), and especially their great sensitivity to the task, methodology and materials used to test their pragmatic abilities.

Correspondence should be addressed to Francesca Foppolo, Department of Psychology, University of Milano-Bicocca, Ed. U6, room 3124, P.za dell'Ateneo Nuovo, 1 – 20126, Milano, Italy. E-mail: francesca.foppolo@unimib.it

INTRODUCTION

The Phenomenon of Scalar Implicature

The topic of scalar inference in children has been a source of great debate in recent years, and this interest is interacting in intriguing ways with semantic/pragmatic theorizing. Our first goal in this paper is to offer some answers to the developmental question of how and when children of different ages derive pragmatic inferences associated with the scalar quantifier *some*. Our second goal is to explore some methodological issues that arise from previous research as well as our own on children's alleged *in*competence with pragmatic inference, by assessing their competence with respect to some basic ingredients that are prerequisites for the derivation of scalar inferences. We will focus on the way children (and adults) react to sentences like (1) and at what age and/or under what circumstances they interpret them as (2):

(1) *Some* children had spaghetti for lunch
(2) *Not all* children had spaghetti for lunch

Following Grice (1957, 1975, 1989) and much literature inspired by him, it is argued that by hearing (1), the conversational (scalar) implicature (2) is generated. Viewing communication as a cooperative exchange, Grice proposed a list of conversational rules (or maxims) that constitute the basis for pragmatic competence. To be cooperative speakers, participants engaged in a conversational exchange should "say no more and no less than is required for the purpose of the (talk) exchange" (Maxims of Quantity I-II), "tell the truth and avoid statements for which there is insufficient evidence" (Maxims of Quality I-II), "be relevant" (Maxim of Relation), and "avoid ambiguity, confusion and obscurity" (Maxims of Manner). The basic idea is the following: uttering a certain sentence involves choosing from a range of reasonable alternatives to it. In particular, sentence (1) evokes (3) as its alternative possibility:

(3) *All* children had spaghetti for lunch.

Moreover, (1) is semantically compatible with (3), given that (1) is entailed by (3); in every situation in which it is true that all children ate spaghetti, for example, it will also be true that some of them ate spaghetti. As a consequence, choosing to utter (1) "*Some* children had spaghetti for lunch" over the stronger alternative (3) "*All* children had spaghetti for lunch" is taken to convey that this alternative does not hold (at least, whenever the number of children having spaghetti is relevant). This is the same as inferring (2) "*Not all* children had spaghetti for lunch." This is what is called a scalar implicature (SI).

The actual steps of this inference remain controversial and will not be discussed in this paper. However, the distinction that ensues from this characterization of the facts between the first level of propositional or "basic" meaning and the second level of pragmatically "enriched" meaning is widely agreed upon. The existence of a contrast between the target sentence (e.g., (1)) and its alternative (e.g., (3)) seems crucial. In fact, quantifiers like *some* can be seen as forming a scale of the following sort: some<many<most<all (Gazdar, 1979; Horn, 1972). Using a generalized notion of entailment, we can view the elements ordered in this scale in a subset/superset relationship: *all* $\subseteq$ *some*. In other words, a sentence that contains the upper elements in a given scale (e.g., *all*) asymmetrically entails a sentence containing the lower elements in that scale

(i.e., *some*). This makes the sentence (3) "*All* children had spaghetti for lunch" more informative than (1) "*Some* children had spaghetti for lunch," where informativeness is defined in terms of the set of circumstances in which the sentence is true. According to this definition, (3) is more informative than (1) because (3) is true in a subset of situations in which (1) is true. In other words, (3) has more chances of being false, thus its informative power is greater. This fact is crucial to understand why implicatures are derived, that is, by exploiting something like first Grice's first Maxim of Quantity according to which speakers should make their contribution as informative as required (for the current purposes of the exchange). Assuming that the speaker is following the general Principle of Cooperation (and to the Maxim of Quantity in particular), the use of *some* in sentence (1) would be *underinformative* if the speaker had evidence that (3) were true; thus, the hearer is entitled to infer that (3) does not apply, namely: he will derive the SI in (2).

One way of thinking about this process is that the scalar alternatives are part of the lexical information associated with the class of items that are members of a scale, so that the relevant alternatives are or might be activated whenever such entries are used, also as a function of the properties of the context in which the scalar term appears (cf. Chierchia, 2006, for a formal implementation of this idea). In other words, we can think of the lexical entry *some* and of other scalar items (e.g., *or*) as consisting of two components: the representation of the basic (or logical) meaning, which in the case of *some* is the representation of an existential quantifier, and the scale (or scalar layer of meaning), which has to be filled out separately for each scalar item, as illustrated in (4):

(4) $||\text{some}|| = \lambda P \lambda Q \exists x[P(x) \wedge Q(x)]$
$||\text{some}||^{\text{S-ALT}} = \{ \text{some} < \text{many} < \text{most} < \text{all} \}$

According to the schema in (4), a sentence such as "*Some* kids are eating spaghetti" can be used in any circumstance in which there is *at least one* individual that is a kid and is eating spaghetti (logical meaning of *some*, first line in (4)). The interplay between the scale associated to *some* (i.e., the class of its scalar alternatives, as shown in the second line in (4)), and the Gricean Maxims reported above (the Maxim of Quantity in particular) results in a SI: the meaning of *some* gets strengthened and *some* is interpreted as *some but not all* (by negating the stronger alternative). To summarize, leaving more subtle theoretical distinctions aside, the process of derivation of a SI goes through the following general steps:

(5)
(i) the computation of the basic meaning of the scalar entry (e.g. the basic meaning of *some* as an existential quantifier is accessed in the lexicon);
(ii) the activation of the scalar alternatives (i.e. *all* is activated upon hearing *some* and the hearer acknowledges the fact that the speaker could have used *all*);
(iii) the selection of the most informative interpretation (if relevant to the purposes of the interchange), in accordance with the Maxim of Quantity.

Steps (i) and (ii) are preliminary (i.e., necessary but not sufficient conditions) to derive the SI, which is only generated at step (iii), when the knowledge of pragmatic norms is required.

The theoretical debate on scalar inference involves three broad families of approaches: Canonical Neo-Gricean approaches such as Horn (1972), Levinson (2000), and Chierchia and McConnell Ginet (2000) i.a.; Relevance Theoretic approaches such as Sperber and Wilson (1986/1995) and Carston (1998) i.a.; Grammatical approaches such as Chierchia (2004, 2006;

Chierchia, Fox, & Spector, in press), Sauerland (2004), and Fox (2007) i.a. There are many interesting theoretical distinctions among these different perspectives and even subtler distinctions within each approach. Also, different theoretical approaches lead to different expectations about children's behavior with scalar inference, and they explain an alleged difference between children's and adults' performance in pragmatic tasks in distinct ways. However, we will not discuss these issues here given that they are not relevant for the specific purposes of this paper and refer the reader to a recent comprehensive review (Geurts, 2011). Instead, we will focus on the acquisition of pragmatic inference and discuss some of the acquisition literature on SI, to which we now turn.

The Acquisition of Scalar Implicatures

In recent years, different experimental studies on the acquisition of SIs have been carried out using different experimental methods with children of different ages, who speak different languages. Starting from the pioneering works by Carol Smith (1980) and Braine and Rumain (1981) that investigated children's logical and pragmatic competence, most of the experimental data collected since then converge on the evidence that children do not derive pragmatic inferences as much as adults. In particular, children seem to consistently interpret scalar items like *some* and *or* logically, that is, as meaning "*some and possibly all*" and "A *or* B and *possibly both.*" Children's early failure with scalar inference has been accounted for in terms of their limited processing capacity that, for instance, would prevent them from handling the cost of Reference-Set computation (Reinhart, 2006) or in terms of an alleged cost of SI computation (e.g., Noveck, 2001, within a Relevance Theory framework). Definitely, though, the picture that emerges from the acquisition literature is far from clear. The age at which children show adult-like pragmatic competence varies greatly across studies and seems to strictly depend on the type of task, the materials used and the scalar item tested. Thus, the question of whether these differences might be accounted for in methodological terms as task requirements imposed on participants or in terms of a maturational account of the pragmatic ability in children is far from being solved. It is exactly at this juncture where we intend to intervene in this paper.

Let us first review some crucial findings. For what concerns the type of scalar item involved in the judgments, Papafragou and Musolino (2003, Experiment 1) investigated the interpretation of three different scalar items (*some, two, start*) in 5-year-old Greek-speaking children and found a low percentage of derivation of the SIs associated with *some* and *start* (12.5% and 10%, respectively) and a better performance on the numeral scale (65% of derivation of the SI). In a similar vein, Papafragou (2003) compared 5-year-olds' rejections of underinformative statements containing the aspectual verb *start* and its synonymous *begin* and *half*, documenting better performance on the latter. Poor performance with another scalar item, *or*, is reported in Chierchia, Crain, Guasti, Gualmini, and Meroni (2001) and Chierchia et al. (2004). Using a Truth Value Judgment Task, Chierchia and colleagues showed that children interpreted *or* exclusively only 50% of the time in sentences such as "Every space-guy took a strawberry *or* an onion ring" in which the SI associated to the exclusive interpretation of *or* (*A or B and not both*) was more appropriate and thus favored by the adults. Finally, Doitchinov (2005) reported that 7-year-old German children failed in deriving the pragmatic inferences associated with the use of epistemic modals such as *können* (may/might) and *vielleicht* (maybe/perhaps) in a picture selection task,

although they were at ceiling in interpreting *some* pragmatically (see also Noveck, 2001, for a similar result on modals in French).

As far as the task is concerned, a clear example of how this affects children's and adults' rate of SI derivation is offered by Noveck (2001) and Guasti et al. (2005). Following Smith (1980), Noveck (2001) investigated the emergence of SIs in children using a Statement Evaluation Task in which the participants had to evaluate a series of sentences containing *some* and *all* that were uttered in isolation, without the support of a visual or prementioned background context. He showed that otherwise logically competent children did not derive the SI associated with *some* up to age 11. For example, only 11% of the 7-year-old French children rejected underinformative statements such as "*Some* elephants have trunks" while French adults did so 67% of the time. This result was replicated with 7-year-old Italian children and adults by Guasti et al. (2005), using the same task used by Noveck and the same materials translated in Italian. These authors, though, also tested the same age group using a different task, the Truth Value Judgment Task (Crain & Thornton, 1998) in which participants were shown a story acted out with props and toys and were asked to judge a puppet's description of it. Surprisingly, almost all the 7-year-old children and all the adults rejected underinformative *some*-statements in this task, showing a significant improvement in the rate of SI derivation compared to the Sentence Evaluation Task (cf. the discussion in Guasti et al., 2005). An effect of task demands was also considered by Papafragou and Musolino (2003) to explain children's poor performance in their first experiment cited above. They conjectured that this result was due to the children's failure to understand that they were not being asked about the truth or falsity of statements (which were in fact true, given that *some* is in a subset relation with *all*) but rather about the pragmatic appropriateness of the statements. In Experiment 2 reported in Papafragou and Musolino, a training session was added in which participants were instructed to correct the puppet's description even when it was truth-conditionally accurate but pragmatically infelicitous. For example, children were trained to reject the description "this is a little animal with four legs" referring to a dog. In this condition, the children's performance improved, and the computation of SIs rose to 52.5% in the case of *some*, 47.5% in the case of *start,* and 90% in the case of numerals. In a similar vein, Papafragou and Tantalou (2004) showed that 77.5% of the 4- to 6-year-old children tested in a task that aimed at approximating naturalistic conversations were capable of assessing informativeness expectations built during a talk exchange. In a typical trial, a puppet was given a job (e.g., to color all of the four paper stars) and then went away to do her job. When she came back, she was asked, "Did you color the stars?", and she answered with the critical sentence, "I colored *some."* The children's task was to reward the puppet if they thought that the puppet completed her job. Most of the children refused to reward the puppet on the basis that she did not color *all* of the stars.[1] Similarly, a difference in the interpretation of *or* was found by Chierchia, Crain, Guasti, and Thornton (1998) contrasting two different tasks: the description mode in which the puppet was asked to describe a situation, as in a classical Truth Value Judgment Task, and the bet mode in which the puppet

[1] Although we think that this manipulation interestingly shows that children are capable of some sort of early understanding of pragmatic implicature in this task, we believe it unclear whether we can interpret children's refusal in this task as a definite indication that they have really derived the SI related to *some* or whether they are simply contrasting alternative descriptions. Given that the puppet is explicitly asked if she colored the stars, a simple "yes" answer would suffice to get the prize. Answers other than "yes" may suggest that the task was not completed, independently from the calculation of the SI related to the scalar term used in the response (see also Papafragou & Tantalou, 2004, p. 76, footnote 6).

made predictions about the possible outcome of a story. Only in the latter case did the children (and adults) overwhelmingly assign *or* the inclusive interpretation, as required by the context of a bet. More recently, Pouscoulous, Noveck, Politzer, and Bastide (2007) claimed that children were more likely to interpret *some* as *some but not all* in simpler tasks. They investigated children from ages 4 to 9 by manipulating the type of task (verbal- vs. action-based judgment tasks) and the complexity of experimental settings (presence vs. absence of distracters). They found that only 32% of the 4-year-olds interpreted *some* logically (i.e., compatible with *all)* in an action-based task while 91% did so in a truth evaluation task. By comparing a binary judgment task and a three-point scale rating task on *some*, Katsos and Bishop (2011) found that only 26% of the 5-year-old English-speaking children tested with a Truth Value Judgment Task straightforwardly rejected underinformative-*some*. Crucially, though, children of the same age were found to be as tolerant as adults with respect to underinformative-*some* in a ternary judgment task (Katsos & Bishop, Experiment 2): 89% assigned a medium-sized strawberry (corresponding to a grade of 2 on a scale of 3) to underinformative descriptions (where a huge strawberry was the optimal reward), while they assigned the lowest grade on the scale of 3 (a small strawberry) in cases of truth-conditionally false descriptions. This result interestingly shows that young children are somewhat sensitive to the pragmatic felicity conditions for the use of *some*, even though this sensitivity does not show up in a classical Truth Value Judgment Task, where they fail to straightforwardly reject underinformative-*some*.

In summary, some general conclusions can be derived from the works on the acquisition of pragmatic inferences discussed so far. In the first place, it is clear that children are very sensitive to training and task manipulations and that their performance varies a lot across tasks, as emerges from a comparison within and across different studies (Noveck, 2001; Chierchia et al., 1998; Papafragou & Musolino, 2003; Guasti et al., 2005; Pouscoulous et al., 2007; Katsos & Bishop, 2011). In the second place, a difference in the derivation of pragmatic inferences is attested across ages and type of scalar term tested. In particular:

(i) some children at 5 years of age still have difficulties in deriving the pragmatic inference related to the use of *some* (e.g., Chierchia et al., 2001, 2004; Papafragou & Musolino, 2003; Papafragou, 2003; Katsos & Bishop, 2011), despite the fact that they show some sensitivity to pragmatic felicity conditions (Papafragou & Tantalou, 2004; Katsos & Bishop, 2011);
(ii) children at age 7 (who speak different languages) behave adult-like in deriving SIs, at least for the scale related to *some* (cf. Doitchinov, 2005, for German; Guasti et al., 2005, for Italian; Pouscoulous et al., 2007, for French);
(iii) children seem to respond differently to different scalar items, as shown by various studies carried out in different languages in which different scales were tested between groups (Papafragou & Musolino, 2003; Papafragou, 2003; Chierchia et al., 2001, 2004; Doitchinov, 2005).

Yet there are still questions that are left unanswered by previous studies. Our aim is that of addressing these issues experimentally. In particular, the developmental question of how and when children of different ages derive the pragmatic inference associated with the scalar quantifier *some* that is, how they react to "Lyn ate *some of* the cookies" when she has eaten them all, has not been fully answered yet. As we have seen, the age at which children show adult-like pragmatic competence varies greatly across studies, and seems to be strictly dependent on the type of

task used to evaluate it. Except for Noveck (2001), that reports an experiment on 8- to 10-year-old French-speaking children, no study thus far has systematically tested children of adjacent age groups on the same task and materials starting with children as young as four. In order to allow a direct comparison with previous studies and to set a baseline for further investigation, we first evaluated children's competence in deriving the implicature associated with the scalar item *some* in a classical Truth Value Judgment Task (Experiment 1) in which participants were shown a story and then were asked to judge a puppet's statement about it. The novelty, in this case, is that we tested children of different but adjacent age groups on the same materials in order to assess the turning point in their development with SIs, that is, at what age children "became" adults in a task that involved judgments about the appropriateness of *some* with respect to a story that would have been more felicitously described by using *all*.

Furthermore, we will explore children's competence with some of the prerequisites that are necessary, though not sufficient, for SI derivation. To this purpose, we will consider some of the factors that might help or prevent children at age 5 to derive the SI associated with *some* from a methodological and a developmental perspective. As we said, the derivation of conversational implicatures is a complex process that involves different abilities and different stages (cf. (5)). While step (i) in the derivation–lexical knowledge of the scalar items has been already attested in different studies even in young children (Barner, Chow, & Yang, 2009; Papafragou & Musolino, 2003), step (ii), which involves the activation of the scalar alternatives, might not be automatized in children yet. For example, the results obtained by Katsos and Bishop (2011) in their second experiment might be interpreted as evidence that children at age 5 are sensitive to scalar alternatives, but no study thus far has tried to manipulate the lexical material in order to favor the activation of the scalar alternatives linked to a certain item. To give children a better chance to reject underinformative *some*, we tried to do this in two ways: in Experiment 2, we tested a synonymous entry for the Italian version of *some* used in the first experiment and tested it in its partitive form (*some of*) to help children to focus on a certain "quantity" in relation to a given set; in Experiment 3, we attempted to prime the *some but not all* interpretation by giving children a description with *all* before the critical description with *some*. As far as step (iii) is concerned, this involves more general pragmatic skills that are a prerequisite for deriving the scalar implicature. In particular, SIs arise only if Gricean conversational maxims, especially the Maxim of Quantity are at play in the interchange. To check for a correlation between the knowledge of pragmatic norms and SI derivation, we compared the children's performance in the Truth Value Judgment Task with their performance in two different tasks that tap into more general pragmatic skills that are a prerequisite for deriving conversational implicatures. Experiment 4 tests the detection of violation of general conversational norms by means of a Conversational Violations Test (Surian, Baron-Cohen, & Van der Lely, 1996; Siegal, Iozzi, & Surian, 2009). Experiment 5 specifically tests children on the Maxim of Quantity by contrasting *some* against *all* in a Felicity Judgment Task (Chierchia et al., 2001). From a purely developmental perspective, it could well be the case that younger children fail to derive scalar inferences simply because they lack general pragmatic abilities that are essential for SI derivation. For example, they might not be sensitive to general pragmatic maxims yet or they might be unaware that *some* and *all* differ along a scale of informativeness related to the quantity of information they convey. Considering some methodological issues that arise from previous research as well as our own on children's alleged *in*competence with pragmatic inference, we further manipulated the experimental design in Experiment 6 by using a within subject design in which strong and weak quantifiers were used

to describe situations in which (a) they truly applied, (b) they truly did not apply, or (c) they were used underinformatively. It can be assumed that recognizing the two layers of meaning associated with a scalar term is a prerequisite to deriving SIs (cf. (4)). Thus, we conjectured that children's performance with SIs might be linked to a general difficulty in shifting their perspectives and/or reversing their strategy, as is attested in tasks that involve executive functions (e.g., the Tower of London test, Shallice, 1982, or the Wisconsin Card Sorting Test, Berg, 1948) and figure reversal in preschoolers (Gopnik & Rosati, 2001). This difficulty might prevent children from being aware of the ambiguity of *some* (cf. (4)); this, in turn, would affect their readiness to shift from the basic to the strengthened meaning of the scalar term, thus blocking the derivation of the pragmatic inference.

EXPERIMENT 1: A DEVELOPMENTAL STUDY

We have seen above that a considerable amount of experimental work has been devoted to the investigation of SIs in children. Although there is general agreement on the fact that children are less prone than adults to generate SIs, it is difficult to draw parallels across the studies mentioned above and to derive consistent generalizations from them, given that they varied in the age and language of the children tested and the task and the methodology used.

In order to better understand this phenomenon, we believe it is important to establish whether there is a developmental effect in the first place and how this effect shows up. For this reason, we carried out an experiment in which children aged 4–7 were tested on their ability to derive the SI associated with the quantifier *some*. Our aim was to determine to what extent children of different ages are able to infer SIs, keeping the method and the experimental materials constant between age groups. This assessment is important to properly evaluate the results of previous studies so as to set the ground for further experimentation. It is also a prerequisite for establishing the reasons why some children behave differently from adults. Our study is different from Noveck's developmental study on *some* cited above (cf. Noveck, 2001, Experiment 3) in that he tested children of an older age range (i.e., 8–10 year olds), speaking a different language (i.e., French), and using a different task (i.e., Sentence Evaluation Task). Moreover, he found that children at age 8 were not adult-like in that task, thus the developmental question about early assessment in children's pragmatic competence with SIs remains open.

Method

Participants. Sixty-three Italian children ranging in age from 4 to 7, as well as 12 adults, participated in the study. Eleven children were discarded from further analysis since they did not finish or understand the task, had language problems, or were not cooperative (five age 4, five age 5, and one age 7). Subjects were divided by age in five different groups: 13 4-year-olds (age range: 4;1–4;10, mean age: 4;5, SD = .31), 12 5-year-olds (age range: 5;0–5;11, mean age: 5;7, SD = .34), 12 6-year-olds (age range: 6;1–6;9, mean age: 6;2, SD = .12), 15 7-year-olds (age range: 7;0–7;9, mean age: 7;5, SD = .28), and 12 adults. Children were recruited from the nursery schools Giustizia (Milan municipality), Milano Due (Segrate, province of Milan), and the primary school in San Pellegrino Terme (province of Bergamo). Adults were undergraduate and graduate students at the University of Milano-Bicocca who volunteered to participate.

Materials and Procedure

We tested participants by means of a Truth Value Judgment Task. We showed children and adults a series of short stories acted out with props and toys; at the end of each story a puppet had to describe what had happened (in one sentence), and participants were asked to judge the accuracy of the puppet's description. The experimental materials consisted of 11 stories, each followed by a target sentence uttered by the puppet. Of these, five were critical statements containing the scalar term *some* that were true but pragmatically inappropriate in the context used. For example, the sentence "*Qualche puffo è andato in barca*" [=Some Smurfs went on a boat] were used in a context in which five out of five Smurfs (i.e., all) went on a boat. To ensure that subjects did not give their answers by chance (rejecting false statements and accepting unambiguously true statements), the stories were interspersed with four fillers, two of which were clearly true and the others clearly false. For example, the sentence "The dinosaur is eating a strawberry" was used to describe a situation in which a dinosaur was eating pizza, thus eliciting a "false" response. Also, to instruct our subjects on the procedure, the test session was preceded by two warm-up stories (one clearly true and one clearly false). If a subject performed poorly in the two warm ups or in two of the fillers she was excluded from the analysis.

A typical critical trial displayed five characters that could choose between two alternatives, for example, five Smurfs on holiday, who could go for a boat trip or a car trip. Differently from Papafragou and Musolino (2003), who used only three characters/objects in the critical trials for *some*, we always used five in our studies. This is because it seems to us somehow infelicitous to use *some* or *all* in the presence of only three characters or objects given that we think the most appropriate way to describe such a situation would be to report the exact number of the characters/objects involved in the action. With small numerosities, in fact, one can easily rely on the automatic process of subitizing and use the numerosity of the set of characters to refer to them. To place emphasis on "quantity", the narrator commented that there were many characters in the story and that it would be very interesting to find out how many of them will choose option A (e.g., the boat) and how many will go for option B (e.g., the car). For example, this is how the narrator introduced the sample story: *"This is a story about a group of Smurfs that are on holiday. Look how many of them we have! They can do a lot of interesting things here. See . . . they have a boat, so they can go for a trip on the river by boat. They also have a car and they can drive their car in the forest. Let's see how many will opt for the boat trip and how many would opt for the car trip. Let's see what happens."* To satisfy the condition of plausible dissent (Crain & Thornton, 1998), the story proper was preceded by each character commenting about the advantages or the disadvantages of each alternative so that both possible outcomes were made plausible. In the end, all the characters opted for the same solution (in the example at hand, they all decided to go for a boat trip). At this point, the narrator asked the puppet to say what was happening in the story in the end. The puppet always provided a description of the story by using a critical (underinformative) statement containing *some*, such as "*Some* Smurfs are going on a boat," which can be judged "true" only if the SI associated with *some* is not derived. Subjects were asked to judge if the puppet said it "well or badly" and if they did not like her statement they were asked to say what was wrong with it and how she could have said it better, so as to improve her language. All the stories were acted out in front of the puppet using props and toys and were recorded with a digital video camera and then transferred onto a CD to be played on a laptop in front of the subject (total recording time approximately 15 minutes). We used the digital version of the stories with the older children and adults, while we decided to use a live

version of the same experiment with the group of 4-year-olds, acting out all the stories in front of the children to involve them in the game and make sure they were paying attention during the narration. Children were tested individually in a quiet room, where they watched the video or the stories together with an experimenter. They were invited to indicate their answers and comments to the experimenter, who filled out a score sheet and took note of their explanations. The test sessions were tape–recorded for further study. Adults were shown the video in small groups and were given a score sheet to write their comments to the puppet's statements.

Results

The distribution of subjects' answers with respect to the mean acceptance rate of critical statements for each age group is plotted in Figure 1. It is immediately evident from Figure 1 that the group of older children (6- and 7-year-olds) patterns like the group of adults but differently from the group of younger children (4- and 5–year-olds), that seemed to pattern at chance. In detail, 4- and 5-year-old children derived the SI associated with *some* (i.e., rejected the critical statements) 43% and 42% of the time, respectively; 6- and 7-year-olds and adults did so 83%, 84%, and 87%, respectively. On control items, however, no difference was found between adults and children of all age groups: adults responded correctly 97% of the time; 5-, 6-, and 7–year-olds gave correct answers 100% of the time; 4-year–olds, 96% of the time. It is important to highlight the fact that when children rejected the puppet's statements, they were always asked to provide the appropriate correction. Interestingly, they all motivated their rejection by consistently invoking the strongest term in the scale, that is, by saying, for example, that "*Some* Smurfs are going on a boat" was a bad description of the story because a*ll* Smurfs were going on the boat. This clearly indicates that children rejected the critical underinformative sentences for the correct reason, recognizing that a scale and a scalar comparison were involved in the evaluation of the sentence.

To better appreciate this result (following Guasti et al., 2005), we considered the distribution of subjects with respect to the number of times they rejected the critical statements (from 0 to 5 times, given that each subject was presented with five different critical trials): only two of the 12 6-year-olds, two of the 15 7-year–olds, and only one adult consistently accepted the critical undrinformative statements. Conversely, the group of younger children split: five of the 13 4–year-olds and seven of the 12 5-year-olds consistently accepted the underinformative

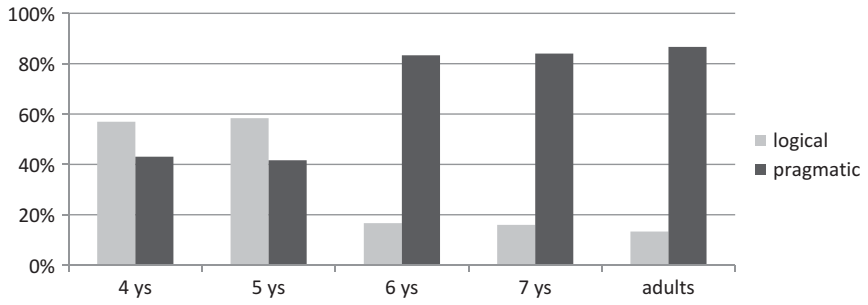

FIGURE 1 Incidence of logical (=acceptance of underinformative-*some*) and pragmatic (=rejection of underinformative-*some*) responses in the developmental study (children from 4 to 7 years of age and adults).

statements (i.e., 5 out of 5 acceptances) while all the others (except for one) consistently rejected them (i.e., 5 out of 5 rejections). Following recent directions about categorical data analysis (Jaeger, 2008), we analyzed our data with SAS using the PROC GENMOD procedure for generalized linear models with repeated measures for binomial distributions. By means of a binomial logistic regression we compared the change of probability of rejecting rather than accepting the critical statements across different critical items and ages. As expected, no statistical difference across different items was found, and thus we did not take this factor into consideration in further analysis. On the contrary, a significant main effect of age was found ($\chi^2(4) = 11.63$, DF = 4, $p = .020$). Contrast estimate results showed that the probability to accept (vs. reject) critical underinformative statements significantly decreased from 4- and 5-year-olds to 6- and 7-year-olds and adults ($p < .05$), while no difference was detected between the age of 4 and 5 ($p = .09$, n.s.) and between the age of 6 and 7 ($p = .96$, n.s) and adults ($p = .8$, n.s.). The details of the contrast estimate analysis for each age group are reported in Table 1.

Discussion

Three main findings in Experiment 1 are worth discussing. In the first place, a clear developmental trend emerged in the ability to derive SIs, in which the age of 6 stood out as the critical turning point in the derivation of the SI related to *some*. From 6 years of age on, children seemed to behave like adults in mostly rejecting the critical statements with *some* when a description with *all* would be more appropriate (and provide suitable justifications). By contrast, children younger than 6 tend to split: only half of them behaved like the older children and adults, mostly rejecting the critical statements, while the other half mostly accepted the critical underinformative-*some* statements. With respect to previous studies that employed a different task (e.g., Noveck, 2001), we showed that, by the age of 6, children could already derive the SI associated with *some* as much as adults when the task was such that the set of evaluation was made directly available to the subjects. This result confirms (and extends) what was previously found by Guasti et al. (2005) with 7-year-old children. As we said, they first replicated Noveck's study in Italian, obtaining very poor performances by 7-year-olds with the Sentence Evaluation Task (Guasti et al., Experiment 1). To show that the context of evaluation matters, they then tested children of the same age with a Truth Value Judgment Task, exactly the same task employed in our study, and

TABLE 1
Contrast Estimate analysis by age

AGE 4	*Estimate*	*S.E.*	χ^2	*p*	*AGE 5*	*Estimate*	*S.E.*	χ^2	*p*
vs. age 5	.94	.79	.01	.942					
vs. age 6	6.61	6.23	4.01	.045*	vs. age 6	7.00	6.80	4.02	.045*
vs. age 7	6.93	5.83	5.30	.021*	vs. age 7	1.99	6.41	5.23	.022*
vs. adults	8.59	7.70	5.76	.016*	vs. adults	9.10	8.41	5.70	.017*
AGE 6	*Estimate*	*S.E.*	χ^2	*p*	*AGE 7*	*Estimate*	*S.E.*	χ^2	*p*
vs. age 7	1.50	1.06	.00	.961					
vs. adults	1.30	1.37	.06	.804	vs. adults	1.24	1.19	.05	.824

found an adult-like performance (Guasti et al., Experiment 4). A further note is worth mentioning at this point: recently, Katsos and Bishop (2011) have challenged the idea that binary tasks like the Truth Value Judgment Task are useful in differentiating between actual implicature derivation and mere sensitivity to informativeness. Their claim is that the link between the rejection of an underinformative utterance containing a certain scalar term and the derivation of the SI related to that scalar term is illicit. In particular, they argue that, if a participant judges a sentence like "The mouse picked up *some of* the carrots" as "false" in a context in which the mouse picked up all the carrots, this cannot be interpreted as evidence that the participant derived the implicature *some but not all*. Their claim is even stronger, though: they argue that, even when an explicit justification is provided, in which for example the participant says that the sentence is false "because the mouse picked up *all* the carrots," one could not conclude that a SI was in fact derived. According to them, the participant's response only shows that the subject is merely sensitive to informativeness and this is sufficient grounds for rejection: "participants could object to underinformative utterances if they recognize that the speaker has given less information than he could, without even considering the implicature arising from the utterance" (Katsos & Bishop, p. 3). Real evidence for the computation of a SI would be obtained, according to them, only if the participants would explicitly say that the sentence is false because "*some means not all*" (Katsos & Bishop, footnote 1). In the first place, we are skeptical with respect to the possibility of ever getting such a justification by participants, because in this case they would say something that is strictly speaking "false" (*some* does not mean "not all"). In the second place, we object to their argument on the grounds of the following observation: given that they do not provide a "formal" definition of informativeness, it is impossible to distinguish what they call "sensitivity" to informativeness (i.e., the rejection of an underinformative statement on the basis of the recognition that a more informative statement could have been made) and the process that leads to a Scalar Implicature in traditional approaches (i.e., the rejection of an underinformative statement on the basis of the recognition that a more informative statement could have been made, where informativeness is measured in terms of "number of circumstances in which the sentence is true"). According to our understanding, saying that an individual is merely sensitive to informativeness means that she would be able to choose the most informative or appropriate option between two alternatives in a certain context. We agree that this would not mean that such an individual is capable of SI computation on the basis of this ability alone. However, we see it as impossible to explain that an underinformative statement is rejected on the grounds that a more informative alternative would have been more appropriate without concluding that a SI was derived in this process. According to us (and, we believe, to all the Gricean and post-Gricean tradition), recognizing that a more informative statement could have been made is the same as generating a SI.

There is a second result in our study worth discussing. Despite the fact that the performance of the 5-year-olds (i.e., 42% of pragmatic responses) was below that of the older children in our study, it was nevertheless higher than the performance reported in the first study by Papafragou and Musolino (2003) discussed in the previous session that employed a similar methodology (i.e., 12.5% of pragmatic responses). A two-sample test for equality of proportions with continuity correction revealed a statistical difference in the pragmatic responses by 5-year-olds between these two experiments ($p < .01$). In consideration of the fact that the procedure and the task were similar in the two experiments, we believe that the improvement obtained in our study might be explained by the difference in the numerosity of the set of characters involved in the actions in the two studies. As we said, we tested *some* in a set of five characters, while Papafragou

and Musolino used only three. It is well known that small numerosities (up to three) can be recognized via the process of subitizing (Kaufman, Lord, Reese, & Volkmann, 1949) and this fact might have interfered with the children's performance in the study by Papafragou and Musolino.

Conversely, we do not think that this difference can be reduced to a mere difference in the two languages tested.[2] In the first place, the critical item tested in all these studies is an existential quantifier, whose semantics generalizes over the specific language in which it is expressed. Given that Papafragou and Musolino (2003) do not provide a formal semantics for *meriki* but they simply treat it as a genuine translation of the scalar quantifier *some*, we have no reason to believe that its semantics differs from the Italian counterpart *qualche*, the French *certain* used by Noveck (2001) or the English *some of* used by Katsos and Bishop (2011), for example. This does not mean that we do not acknowledge the possibility of differences in the use of different forms of the same quantifier within the same language (to cite only some examples: *some* vs. *some of* (Grodner, Klein, Carbary, & Tanenhaus, 2010); *certain* vs. *quelque*s in French (Pouscoulous et al., 2007); and *qualche* vs. *alcuni dei*, cf. Experiment 2 in this paper). Nonetheless, we believe that a generalization about certain features of language is not only possible across different languages but also desirable. Moreover, there are several examples in the literature that confirm that different languages do not differ in the interpretation of different items that share the same semantic features. For example, Guasti et al. (2005) report a replication of Noveck's study on Italian children and adults by means of the same task employed by Noveck (2001) in which the only manipulation was translating the French sentences containing *certains* into Italian sentences introduced by *qualche*. The performance attested for Italian children and adults in Guasti et al.'s study were perfectly comparable to that found by Noveck for French participants in the Sentence Evaluation Task. Similarly, Katsos and Bishop (2011) and Katsos, Roqueta, Estevan, and Cummins (2011) report analogous performance by English and Spanish TD children in a Truth Value Judgment Task employing similar materials to test the scalar quantifier *some of* and *algunos de los,* respectively.

Finally, the younger children's performance deserves a closer look: as we said, children age 4 and 5, as a group, did not consistently derive the SI associated with *some.* By analyzing the distribution of individual subjects with respect to the number of times they provided a given answer, however, we observed that younger children did not answer randomly but were in fact very consistent: subjects persevered in the answers they offered the first time as if they had adopted some sort of strategy to which they stuck during the whole experimental session. This result is consistent with the general difficulty observed in preschoolers performing tasks that involve executive functions and that require participants to shift their perspectives and/or reverse their strategies, (the Tower of London test, Shallice, 1982). It also replicates a finding that went unnoticed by other authors (and already discussed in Guasti et al., 2005), that is, the bimodal distribution held by young children, which is to be handled differently from chance performance. In fact, the 4- and 5-year-olds seem to split into two groups: one group of children that has already caught up with the ability to derive SIs and one group that has not. It is conceivable, though, that in between there is a subgroup of children that might not be able to automatically derive the pragmatic inference yet, but would be able to do so if some additional cue were added to the experimental setting and/or materials. We consider this possibility in the next and last section of this paper.

[2] This possibility was suggested by one of the reviewers.

Giving Children a Chance: Manipulating Lexical Cues

Although the findings of Experiment 1 provide an answer to the question about the trend of development of the pragmatic ability related to the scalar quantifier *some*, they do not help us in understanding what prevents children at ages 4 and 5 to derive pragmatic inferences—if it is a matter of inability at all. At this point, it's worth thinking of a possible set of factors that can influence children's performance. As we said, SI might be thought of, in a nutshell, as the result of at least three basic steps (cf. (5)): (i) the computation of the basic meaning of the scalar entry; (ii) the activation of the scalar alternatives (e.g., activating *all* upon hearing *some* and acknowledging the fact that the speaker could have used *all*); and (iii) the selection of the most informative interpretation (if relevant to the purposes of the interchange and in accordance with the Gricean Maxims). While the first two steps depend on the "lexical" knowledge of the scalar term, which needs to be acquired and consolidated during development, the last step requires the mastery of the maxims that rule our talk exchanges and, more specifically, the mastery of something like the Gricean Maxim of Quantity about the quantity of information conveyed by a message. We know from previous studies that children before age 5 master quantifiers like *some* and *all* properly, that is, do not have problems with step (i), at least at the age considered here. At the level of "lexical" knowledge, however, there are two other possibilities that have not been specifically addressed experimentally in children yet and that tap abilities that are involved in the first and second step of the derivation. On the one hand, it is the comparison between the partitive and the plain use of the scalar item (*some of* vs. *some*; see Grodner et al. (2010) for a similar comparison in English adults). On the other hand, it is the possibility to prime the strengthened meaning of a weak scalar like *some* by using it underinformatively soon after a felicitous use of its stronger counterpart *all*, to test if this prompts the activation of the relevant scalar alternative (step (ii)). We explore these possibilities in Experiments 2 and 3 in the next section.

EXPERIMENT 2: THE PARTITIVE CONSTRUCTION

In Italian, there are at least two words that correspond to English *some*: *qualche* and *alcuni.* While *qualche* is morphologically singular, even though it can refer to a plural set, *alcuni* is morphologically plural. Crucially, only the latter can be readily used in its partitive form, by simply adding the possessive "*dei*" (*of*) to it, as happens for English *some*/*some of.*[3] In this experiment, we replaced the expression *qualche* (translated as *some*) used in the critical sentences in Experiment 1 with the partitive form *alcuni dei* (*some of*). As far as the SI derivation is concerned, it is conceivable that the use of the partitive might help children to focus on the relevance of the set of characters under consideration, thus enhancing the probability that they will reject an underinformative description about it. An advantage of the partitive over the bare quantifier in the derivation of the scalar inference related to *some* has already been documented by Grodner et al. (2010) in a survey on English speaking adults.

[3] *Qualche* cannot be used in its partitive form; one possibility is to use *qualcuno dei*, but its use is controversial.

Method

Participants. A new group of 24 Italian-speaking 5-year-olds (Age range: 5;2–6;2, MA: 5;8, SD. = 3.1) and twelve adult native speakers of Italian participated in this study. The children were recruited from the I Circolo nursery school in Cernusco sul Naviglio, province of Milan. Adults were undergraduate and graduate students at the University of Milano-Bicocca.

Materials and Procedure. The test materials were the same as those used in the first experiment, with the only difference that the critical statements at the end of each story contained the partitive *alcuni dei* (*some of*) instead of the bare quantifier *qualche* (*some*). Also the procedure was identical to the one previously adopted.

Results and Discussion

The rate of acceptance of the critical statements containing *some of* was 62.5%, thus children derived the SI only 38.5% of the time (42% in Experiment 1), while adults did so more than 90% of the time. As in the first experiment, children were split into two groups: out of the 24 5-year-old participants, 15 always accepted the critical statements and nine never did. With the exception of one individual, all the adults always rejected the critical statements instead. With respect to fillers, both groups performed correctly 100% of the time. By the absence of any improvement in this second experiment we can conclude that the use of the partitive construction did not help children to reject the underinformative description of the set.

EXPERIMENT 3: PRIMING THE SCALE

As we have seen in (4), the lexical representation of scalar terms involves two layers/components: the basic meaning and the scale, from which the scalar or strengthened meaning is derived. We are abstracting here from the different ways in which these components might be implemented/stored/activated in the lexicon. The idea at the basis of this experiment is the following: if the difficulty that children encounter in the derivation of the SI is rooted purely at the lexical level (e.g., it resides in a difficulty in the activation of the scalar component corresponding to the second step in the derivation described in (5)), then providing children with a scalar alternative might prime the scale associated with the scalar term, facilitating step (ii) in the derivation of the SI.

Method

Participants. A new group of 12 5-year-olds (age range: 5;4–5;10, mean age: 5;7, SD = 1.83) participated in this experiment. The children were recruited from the Giustizia nursery school (Milano Greco).

Materials and Procedure. The test material was the same as that used in Experiment 1, with the exception that a new set of characters was added to each critical story. So, for example,

in the story of the Smurfs on holiday described before, a bunch of dwarfs and some candy were added to the same set. Differently from previous experiments, in this case we acted out the stories live in front of the child because there were too many toys used in these new stories to be properly captured in the video. In this "live" version, two experimenters were involved, one narrating the stories and the other manipulating the puppet. Each story comprised two parts: the first part focused on the added group of characters (e.g., the dwarfs) and their objects (e.g., the candy) while the second was basically the same story used in the previous experiment. On a typical trial, first the experimenter described the situation and introduced the characters to the puppet and the child. For example, she introduced the stories by saying:

> "*This is story about a group of Smurfs and a group of dwarfs that are on holiday. Look how many of them we have! They can do a lot of interesting things here. For example, they have a boat, so they can go for a trip on the river by boat. They also have a car and they can drive their car in the forest. Oh, look! There is also a lot of candy here! If they are lazy, they can even relax and eat candy. But let's see what happens!"*

Then the story itself started. At first, the attention of the child was drawn to the group of the added characters, the dwarfs in our example, who were discussing eating candy. In the end every dwarf took a piece of candy and sat down to eat it. At this point the experimenter stopped narrating the story and addressed the puppet, asking her what had happened in the story so far. The puppet always described the first part of the story by using a true and appropriate statement containing *all*, for example, *All the dwarfs are eating a piece of candy*. The child was immediately asked to evaluate if the puppet said it "well or badly" with respect to the situation. Then the story went on focusing on the second set of characters. Crucially this part was identical to the one provided in the previous experiment (in our example, this part ended with all the Smurfs going on a boat). Again, the puppet was asked to tell what was happening and this time she described the story with a true but underinformative statement containing *some* (that was identical to the statement used in Experiment 1). As before, the child was asked to evaluate the puppet's statement. If the participant rejected the puppet's description, she was asked to tell the puppet the correct or better way to say what had happened.

Results and Discussion

Children's performance with *all*-statements and fillers was 100% correct. No children hesitated in accepting the true and appropriate descriptions with *all.* On the contrary, their performance with SIs was not adult-like: only 42% of the times children rejected the underinformative *some*-statements. Surprisingly, their performance in this study was identical to the 5-year-old children's performance in Experiment 1. As in the previous study, children split: six of the 12 children consistently accepted the underinformative statements (i.e., 5 out of 5 acceptances); four consistently rejected them (i.e., 5 out of 5 rejections), and the remaining two children rejected the underinformative statements only once and twice, respectively.

As we said, the manipulation made in this experiment aimed at favoring the activation of the scale in virtue of the fact that hearing an *all*-statement before a *some*-statement could prime the scale some<all. This should make the child more aware of the more informative alternatives to *some*, with the effect of having more subjects rejecting the critical underinformative statement

compared to Experiment 1. If the difficulty that emerged in the previous experiments was effectively due to a difficulty in the spontaneous retrieval of the scale, then mentioning the stronger scalar alternative should result in an improved performance by children. However, we found no such improvement. Therefore, we can conclude that the manipulation introduced to prime the scale was not a strong enough cue for children to reject underinformative *some*.

Measuring Children's Chances: Testing Their Pragmatic Knowledge

On a more general level, it is possible that the children who did not derive the SI associated with *some* in the Truth Value Judgment Task (Experiments 1, 2 and 3) lacked the knowledge of conversational maxims in general, and/or the Maxim of Quantity related to the ordering in the scale <some, all> in particular. The third step in the derivation of a SI is the selection of the most informative interpretation, whenever this enriched interpretation is relevant to the purposes of the interchange (cf. (5)). To perform this last step, children need to master the maxims that rule our talk exchanges in general and, more specifically, to master something like the Gricean Maxim of Quantity about the informativeness conveyed by a message. To assess children's mastery of these principles, we conducted two follow-ups on some of the same children that were tested with a classical Truth Value Judgment Task. These are reported as Experiments 4 and 5.

EXPERIMENT 4: MEASURING CHILDREN'S DETECTION OF VIOLATIONS OF CONVERSATIONAL MAXIMS

Experiments 1, 2, and 3 showed that children at 5 years of age do not uniformly derive SIs in a task in which adults' performances are at ceiling. One possibility is that children accept underinformative statements simply because their pragmatic ability is still immature at age 5: they might violate conversational maxims (e.g., the Maxim of Quantity when accepting underinformative-*some*) because they do not know yet when and how these maxims apply. We address this possibility in our next experiment, in which a test for detecting the violations of conversational maxims was performed on the same subjects that were tested with the Truth Value Judgment Task.

Methods

Participants. The participants were the same 24 5-year-old children that participated in Experiment 2. They were tested in this new task two weeks after the first experiment.

Materials and Procedure. The children took a Conversational Violations Test, a test already used with monolingual and bilingual children, children with autism, and adults with right hemisphere lesions (Surian et al., 1996; Surian & Siegal, 2001; Siegal, 2008; Siegal et al., 2009). The purpose was to investigate the subjects' ability to detect violations of Gricean conversational maxims (cf. Surian et al. and Siegal et al. for a comprehensive description of the materials). Using a laptop, participants were shown 25 short conversational exchanges among three doll speakers, one male and two female. For each exchange, one of the two female speakers asked a question to the other two, who in turn gave a short answer: one violated a conversational

maxim while the other did not. The children were asked to "point to the doll that said something silly or rude." Overall, the test comprised five utterances for each maxim (the first and second Maxim of Quantity, the Maxim of Quality, the Maxim of Relation, and the more general Maxim of Politeness), for a total of 25 items (cf. e.g., Siegal et al., 2009, for a detailed description of this task). For example, a typical trial for testing the First Maxim of Quantity is in the following exchange, in which the target utterances were designed to fall short of providing an informative enough answer: Doll 1: "What did you have for lunch?"; Doll 2: "Some food" (underinformative–target); Doll 3: "A pizza" (informative answer).

Results and Discussion

Children split in the Truth Value Judgment Task and only 38.5% rejected the underinformative utterances containing *some of* (cf. Experiment 2). At the same time, though, these same children scored at least 17/25 correct in the Conversational Violations Test, pointing to the puppet that violated the conversational maxims at least 17 times overall. In fact, 18 of the 24 children (75%) performed at ceiling in this task, scoring at least 21 out of 24 (above 95% correct, Binomial test, $p < .05$). The mean score of our group was 22/25. This result on the Conversational Violations Test is in line with what was previously found by Surian et al. (1996) for the control group of typically developing children (age: 6;7) that scored 17/25 in the same task and by Siegal et al. (2009) for the Italian monolinguals (age: 5;4) and for the Slovenian-Italian bilinguals (age: 5;6) that scored 18.6/25 and 21/25 correct, respectively.

We also performed a logistic regression analysis to test if the pragmatic score obtained in the TVJT (Experiment 2) could be explained by the pragmatic score obtained in the CVT overall or by the pragmatic score obtained in the sub-set of items related to the Maxims of Quantity. To this end, we introduced the number of correct responses obtained by each subject in the CVT as an additional predictor/covariate in our statistical model. Interestingly, this did not increase the goodness of fit of the model neither when we considered the total number of correct responses given in the Conversational Violations Test ($\chi^2 = 1.93$, DF = 1, $p = .1646$), nor when we took into consideration the number of correct responses given in the ten trials related only to the Maxim of Quantity ($\chi^2 = .31$, DF = 1, $p = .57$).

Considering the results so far, we have seen that children's performance with pragmatic inference is not adult-like at age 5, as attested by the low incidence of pragmatic responses in the Truth Value Judgment Task (Experiments 1, 2, and 3). At the same time, though, children's ability to detect the violations of the conversational norms in ordinary exchanges seems already in place, as attested by their good performance in the Conversational Violations Test (Experiment 4). Moreover, the fact that a correlation is not found between performances in the two tasks makes it very unlikely that one can explain children's failure with SIs by appealing to a deficit or a delay in conforming to conversational maxims in general or to the Maxim of Quantity in particular.

EXPERIMENT 5: THE FELICITY JUDGMENT TASK

Despite their knowledge of Conversational Maxims in general, and despite their ability to use the scalar terms in situations in which they truly apply, children might not be aware of the fact that some specific scalar items are tied and ordered in a scale of informativeness. Recall that children

need not only to know the basic meaning of the scalar terms to generate SIs, but they also have to know the scale associated with them and to automatically retrieve it. In terms of the lexical representation given in (4), this means that they ought to have acquired both the basic and scalar components of the lexical entry associated with each scalar term. Moreover, the link/association between the two layers of the lexical representation presumably needs to get automatized so that, by accessing one, the other is also activated. In some sense, this mechanism resembles that of the acquisition of inflectional paradigms of verbs. Let us push this parallelism further by means of an example. In the process of acquiring the past tense of the verb "sit," an English baby would probably go through a stage in which she would produce "sitted" instead of the correct irregular form "sat." This is known as the process of overregularization (e.g., Pinker, 1991, 1999; Clahsen, Hadler, & Weyerts, 2004), according to which the first guess the child makes is that all words belonging to the same category (e.g., verbs) are modified by the same range of possible suffixes (e.g., -ed to make them past and –s to make them third person singular). After some (variable) amount of exposure, the English child gets to know that some verbs behave differently from the mainstream, and only when she is aware of this possibility does she start learning irregular forms, eventually producing "sat." In this phase of maturation, we conjecture that it would be easy to get a child to advocate that "sat" is better than "sitted" when explicitly presented with the two options, even if the same child might sometimes use "sitted" instead of "sat" in spontaneous production. Along these lines, we might conjecture that children at a certain stage in development lack the scalar component tout court, or are still in the process of building it and connecting it to the basic layer of meaning. In these terms, it is conceivable that children might pass through a stage in which they know that *all* is better than *some* to describe a certain situation in which *all* truly applies, but they cannot autonomously come up with the scalar representation of *some* in which it is contrasted with *all* within the same scale of informativeness. This prerequisite was investigated by Chierchia et al. (2001, 2004) in 5-year-old children with respect to the scale <or, and>. They first investigated children's derivation of the SI associated with *or* by means of a Truth Value Judgment Task and found poor performance. Then, they tested the same children with a Felicity Judgment Task, in which they were given two alternative descriptions of the same situation and were asked to judge which one was the best. Following Chierchia et al., we took the same children who failed to derive the Scalar Implicature in the Truth Value Judgment Task and gave them a Felicity Judgment Task to test if they could adjudicate between *some* and *all* when the two options were given, despite the fact that they had interpreted *some* as compatible with *all* in the Truth Value Judgment Task.

Methods

Participants. The participants were 17 of the children that had taken a Truth Value Judgment Task in the preceding weeks and failed to derive the SI associated with *some* (specifically, Experiment 1 and another experiment analogous to that which is not reported in this paper; age range: 5;1–5;11, mean age: 5;7, SD = .29). They were tested in this new task about two weeks after the first experiment.

Materials and Procedure. We tested children with a version of the Felicity Judgment Task in which we compared statements with *qualche (some)* and *tutti* (*all)* as alternative descriptions of pictures in which the statement with *all* was always the most appropriate. The children were

told that these puppets were always quarrelling about which of them could say things better and so they decided to test their competence in a competition in which the child was asked to be the judge. The competition consisted in describing some pictures, which were displayed on the table in front of the child and the puppets. For each picture the two puppets in turn gave the best description they could and the child had to judge each time which puppet said it better. The testing materials consisted of a series of 11 pictures, five of which were critical trials. For example, one of the pictures displayed five chipmunks taking a shower: one of the puppets described it as "*Some* chipmunks are taking a shower," which is true but informationally weak in the context, the other puppet described it with the more appropriate (informationally stronger) statement "*All* the chipmunks are taking a shower." Correct and appropriate responses were counterbalanced through "puppets" and sessions so that the final score of the winner was always 6 out of 11 and each puppet in turn turned out to be the winner of the competition.

Results and Discussion

Despite their poor performance in the Truth Value Judgment Task, the children's performance in the Felicity Judgment Task was above 95% correct overall. We believe that this result convincingly shows that children are able to choose which element of the scale <some, all> is more appropriate in a given context; that is, they are sensitive to the Maxim of Quantity. It is worth noting, however, that this result alone does not provide evidence that children can derive SIs. They merely prove that *some* piece of lexical/semantic knowledge that is necessary but not sufficient to derive SIs (i.e. understanding the ordering of informational strength) is available to them (step (ii) in (5)). As we know, the process of deriving SIs is far more complex than choosing the most informative between two given alternatives. It requires the hearer to understand that the utterance with *some* conveys the meaning that *not all* holds for the speaker. In turn, this is possible if the hearer (i) knows the scale, (ii) can recover it by herself, (iii) is able to generate the alternative description containing the stronger item in the scale, and (iv) thus negates it because the weaker term has been used (following Gricean Maxims). Knowing that two items are different in informational strength when asked to compare them does not imply knowing that these items are ordered in a scale, nor that one is automatically activated upon hearing the other. Nor does it prove that children can construct and/or recover the relevant alternative by themselves, given that in the Felicity Judgment Task the two alternatives are given and the task is to choose between them (cf. Reinhart, 2006).

EXPERIMENT 6: TRIGGERING CHILDREN TOWARDS THE DOUBLE LAYERS OF *SOME*

Thus far, we have shown that children at age 5 failed in the Truth Value Judgment Task (Experiments 1, 2, and 3), although they show knowledge of general pragmatic norms (Experiment 4) and sensitivity to informativeness (Experiment 5). We believe that a study by Gopnik and Rosati (2001) on ambiguous representations might be relevant to explain children's failure with SIs (cf. also Rock, Gopnik, & Hall, 1994). In their study, they investigated 3- to 5-year-old children's ability to reverse ambiguous figures (like the classical duck/rabbit or

man/mouse illusions) and the relation between this ability and a general ability to understand multiple representations. What they found is that children at age 5 only reversed (i.e., were able to alternate between the duck and the rabbit) when they were informed of the alternatives, that is to say only when they were told that two possible figures could be seen in the same picture. Quite interestingly, even when they were informed, most of the 4-year-olds and all the 3-year-olds failed to experience reversal. From these results, Gopnik and Rosati draw the conclusion that the ability to reverse depends on a more abstract understanding of ambiguous representations. We believe that this suggestion might be relevant to explain why children at age 4 and 5 experience difficulty with the computation of SIs; as we said, deriving a SI involves at least three steps. In this respect, one could view a statement containing a scalar term like "John ate *some* of the apples" as "ambiguous" between a plain meaning ("There is *at least* some apple that John ate") and a strengthened meaning ("John ate *some but not all* of the apples"). Children seem not to be able to automatically access both interpretations, or to easily shift from one to the other, just like they showed difficulties in reversing their perspective to see a duck or a rabbit within the same image without explicit instructions. At the same time, though, children at age 5 showed no problem in choosing the correct description of the picture when the two alternatives were readily given, just as they easily could point to the duck or rabbit part of the same image when they were informed of the two perspectives. Borrowing from Gopnik and Rosati's suggestion, we tested a new group of 5 year olds with a Truth Value Judgment Task in which appropriate and inappropriate descriptions involving *all* and *some* were presented within subjects to enhance the probability that they could see the ambiguity between the logical meaning of *some* (*at least some*) and its enriched meaning (*some but not all*).

Methods

Participants. A new group of 47 5-year-olds and 40 adult native Italian speakers participated in the experiment. The children were recruited from different nursery schools in the Milan metropolitan area: the Cesalpino and Soffredini nursery schools (Milan municipality) and the nursery school in Cornate D'Adda (province of Milan). The adults were undergraduate and graduate students at the University of Milano-Bicocca who volunteered to participate. As in previous experiments, subjects who, for any reason, did not finish or understand the task were excluded from the analysis. This was the case for seven children, so that final data comprise a total of 40 5-year-olds (age range: 4;11–5;11; mean age: 5;4; SD: 0.15).

Design, Materials and Procedure. The procedure adopted was the Truth Value Judgment Task described in the previous experiments. In this experiment, we changed the experimental design so as to consider within and between group comparisons of critical statements. The whole battery consisted of 48 story-sentence pairs divided in four lists of 12 stories each. Of the whole battery, eight were underinformative critical statements containing *alcuni dei (some of)*, varying between the subject and object position. The rest were control trials on different scalar items (we will not address the difference between scalars in this paper). Lists and stories were treated as a between-subject factor for the statistical analysis: the same story was created in different versions that were displayed to different participants across lists, described by different sentences. For example, the story of a dwarf in his vegetable garden in which five carrots were ready to be

picked up was created in two different variants: one in which he picked up all five carrots and one in which he picked up three of the five carrots. The variants of the same story were used to test two different statements across lists (i.e., "The dwarf picked up *some of*/*all* the carrots"), which varied with respect to their truth/falsity and appropriateness. For example, the variant of the story in which the dwarf picked up five out of five carrots was used to test the critical underinformative sentence with *some of* in the object position in one list and as a control for *all* (true) in another list. The variant of the story in which the dwarf picked up only three of the five carrots was used to control for *some* (true) in one list and for *all* (false) in another list. All sentence-types were seen within lists, associated to different story-types. Subjects were randomly assigned to one of the four lists. Items were presented semirandomly, to keep two consecutive trials from containing the same type of scalar term. Differently from Experiments 1 and 2, we tested a control condition for *some* and *all* as a within subject variable. In particular, the children in all lists encountered the following in order:

(i) a situation in which *all* was used correctly: for example, the sentence "*All* the Smurfs went on a boat" was used to describe a situation in which five out of five Smurfs went on a boat;
(ii) a situation in which *all* was used incorrectly to describe a situation in which *some* would have been appropriate: for example, the sentence "The dwarf picked up *all* the carrots" to describe the situation in which the dwarf picked up three out of five carrots;
(iii) a situation in which *some* was used underinformatively to describe a situation in which *all* would have been appropriate: for example, the sentence "The owl lit *some of* the candles" to describe the situation in which an owl lit five out of five candles.

As before, children were asked to evaluate the description provided by the puppet and correct it if it was wrong. Some considerations about the materials we used are in order here. First, we used the same materials (stories and sentences) as in Experiment 2 to test *some of* in the subject position. As we said, we used five objects for the scale *some*<*all*, to avoid the concerns discussed earlier about the study by Papafragou and Musolino (2003). Second, we tested *some* both in the subject and object position to ensure that the syntactic position itself did not affect the computation of SIs. In fact, Papafragou and Musolino tested *some* in the subject position in their baseline experiment while they tested it in the object position in their second experiment with training. Given that they did not consider this manipulation explicitly when analyzing their data, we wanted to control for this factor. Third, we laid emphasis on the quantity of the objects or characters present on the scene, as we did in Experiments 1 and 2 ("Look! This dwarf has a lot of carrots! Let's see how many he will pick up in the end).

Results

On control items, 100% of adults gave correct responses for all items, except for one case in which one adult made a mistake on one of the control trials. Analogously, children performed more than 90% correct on most controls, with the exception of *un pezzo* (*a piece*).[4] On critical

[4] Statistical analysis, by means of logistic regression, revealed that children performed differently from adults in the control condition in which *a piece of* was used ($\chi^2(1) = 6.89$, $p < .0087$): contrast estimate results show that errors are about 9.75 times more common in children than in adults in this condition. Taking a closer look at how children responded,

sentences containing the scalar item *some*, all 40 adults (100%) correctly rejected the underinformative statements containing *some* in the object position and 39 out of 40 adults (97.5%) rejected underinformative statements with *some* in the subject position. As for children, 75% rejected the underinformative statements in cases of *some* in the object position and 70% rejected it in the subject position; that is, 30 and 28 of the 40 children derived the SI related to *some of* in object and subject position, respectively. Recall that, differently from previous experiments (in particular, differently from Experiments 1 and 2 reported here), each subject in this case had to judge each type of construct only once, while simultaneously having to judge all types of sentences within the same experimental session. As in the previous experiments, children that rejected the underinformative statements were always asked to correct the puppet's statement with a more appropriate description (sometimes they provided it spontaneously). In any case, they always motivated their rejection by invoking the stronger term in the scale. For example, they claimed that "The owl lit *some of* the candles" was not correct because in fact the owl lit *all* the candles he had.

As before, we analyzed our data using a generalized linear model for repeated measures by means of PROC GENMOD in SAS 9.1. First of all, remember that subjects were randomly assigned to one of four lists: as expected, no main effect of List was found in children ($\chi^2(3) = 4.23$, DF = 3, p = .24, n.s.), thus we will not take this factor into consideration in further analysis. A general main effect of age was found ($\chi^2(1) = 20.18$, DF = 1, p = < .0001) due to the fact that children's "errors" (i.e., rejection of true statements, acceptance of false ones, or acceptance of underinformative critical statements) were 6.5 times more common than adults' errors. However, further comparisons showed that children significantly differed from adults only in the rate of pragmatic responses in critical ($\chi^2(1) = 21.60$, p < .0001) but not in control trials ($\chi^2(1) = 2.62$, p = .10, n.s.). Contrast estimate results showed that the acceptance of underinformative critical items was overall 24.6 times more common in children than in adults. Also, a main effect of Trial was found in children (Controls vs. Critical trials, $\chi^2(1) = 13.63$, DF = 1, p < .001): contrast estimate results showed that the odds ratio from critical to control trials significantly decreased at a 0.16 rate in children; that is, non-adult-like behavior was 6.3 times more common in underinformative critical items than in controls (S.E. = 0.055, $\chi^2 = 28.04$, p < .0001). In particular, children performed differently from adults in cases of *some-subject position* (p < .001) and in cases of *some-object position* (p < .0001), while no effect of position (subject vs. object) was found in the rate of acceptance of the underinformative-*some of* by children (p = .53, n.s.). Contrast estimate results showed that children accepted underinformative *some* about 16 times more than adults (S.E. = 17.88, $\chi^2 = 6.93$, p < .01). Nonetheless, an increase in the incidence of children's pragmatic responses was observed across experiments, as shown in Figure 2.

---

we see that in fact some children rejected true statements containing *un pezzo (a piece of)* because they interpreted this item in a "numerical" sense. For example, in a situation in which a character had built a *piece of* of the puzzle (and, crucially, not the *whole* puzzle), some children said that the puppet was wrong because the character had used *three pezzi (pieces)* of the puzzle and not just one. In Italian, in fact, there's no phonological distinction between *un pezzo* (lit. a part of) in which *un* is used as an indefinite article, and *un pezzo* (lit. one piece of) in which *un* is used as the numeral "one". The ambiguity between *un pezzo (part)* and *un pezzo (piece)* might have led some children to take the expression *un pezzo* in its numerical sense, thus causing them to reject the otherwise true proposition. For this reason, leaving aside the data on *a piece of,* no difference is observed between children and adults in the incidence of correct responses in control trials.

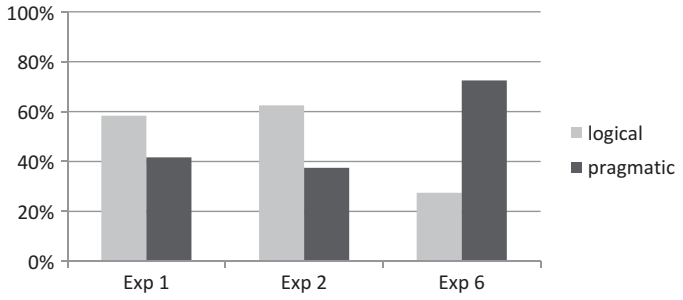

FIGURE 2 5-year-old children's performance across experiments: logical responses = acceptance of underinformative-*some*; pragmatic responses = rejection of underinformative-*some*.

As is evident, the children were more prone to derive the SI related to *some of* in the last experiment reported. This observation is confirmed by statistical analysis: compared to Experiments 1 and 2, in which the probability of finding a "pragmatic" responder among the 5 year old children was not different from chance (e.g., $p = .77$ in Experiment 1, Exact Binomial Test), the probability of finding one in this last experiment is significantly higher than chance (Exact Binomial Test, $p < .01$).

Discussion

The main finding of this last experiment is the improvement recorded on *some* compared to previous studies including our own. As we observed, the percentage of SI computation rose from about 40% (Experiments 1, 2 and 3) to 72.5%. This result is all the more surprising if we consider the fact that the type of materials (stories and sentences) and the experimental procedure used in our experiments were basically the same. The only difference, as we said, was the change in the experimental design, namely the fact that a correct and incorrect use of the stronger element *all* was shown to subjects before they had to evaluate the critical underinformative statement with the weaker scalar *some.* Interestingly, we know from our Experiment 4 that the simple fact of hearing *all* (correct) before *some* (underinformative) did not enhance the probability of the children's rejection of the latter. Thus, the mere presence of *all* within the same trial did not seem to play a crucial role in favoring the SI related to *some*. Furthermore, a similar task was employed by Katsos & Bishop (2011, Experiment 1) in English and by Katsos et al. (2011) in Spanish, in which subjects were shown underinformative sentences with *some* interspersed with true and false control sentences containing *all* and *some* in a within subject design. While the children's performance on controls was adult-like, as in our study, their performance with critical statements was well beyond that recorded in our study (26% and 36% vs. 72.5%). Interestingly, there was a crucial difference between their studies and ours, which might be responsible for the difference in the children's performance between their studies and ours. First of all, in Katsos and Bishop's Experiment 1, the false controls for *all* and *some* were tested in a situation in which the character did something to none of the five objects present (i.e., arrangement 0/5). Conversely,

in our study, each child was exposed to the following elements in this order and in rapid succession:

(i) the upper element in a situation in which it truly applied (arrangement: 5/5);
(ii) the upper element in a situation in which the lower element applied, but was not used (arrangement: 3/5);
(iii) the lower element in a critical condition in which the upper element would have been more appropriate, but was not used (arrangement: 5/5).

Similar conditions were used in Katsos et al.'s study in Spanish, but in fact the order in which these configurations were presented to children was different from our Experiment 6; in their study, children were never shown configuration (ii) before the critical (iii) in any of the orders tested (Napoleon Katsos, personal communication June 22, 2011).

To account for the increase in pragmatic responses observed in our Experiment 6 we can only speculate at this point. The explanation we propose refers back to the study by Gopnik and Rosati (2001, cf. also Rock et al., 1994) on visual illusions discussed above. This study showed that children at age 5 were able to reverse their perspective only when they were made aware of the ambiguity of the picture. Namely, the children were presented with an ambiguous picture in which a duck or a rabbit could be seen by shifting perspective. Without explicit instructions, the children were not aware of the ambiguity, but if the experimenter pointed out that two possible figures might be seen within the same image, the children succeeded in reversing. We believe that a similar facilitation might have applied to Experiment 6. As we said, we can think of an underinformative statement as ambiguous between a basic and a strengthened meaning; we know that children are not automatically aware of ambiguities, but they are able to shift their perspectives when informed of the alternatives. We think that it is exactly this information that they got in our Experiment 6, which helped them to recognize the two layers of the meaning of *some* and retrieve the scale (provided that this is already available in the lexicon) so that they could choose the most appropriate, strengthened meaning and reject the underinformative sentence. We believe the interplay between arrangements 3/5 (described by *all*) and 5/5 (described by *some*) to be the crucial factor in this respect. In the former situation, *some* would have been correct (but *all* was used instead); in the latter, *all* would have been felicitous (but *some* was used instead). This shift in configuration was only given in our Experiment 6.

GENERAL DISCUSSION

A series of experiments investigated children's ability to generate one kind of pragmatic inference, namely scalar implicature, in sentence evaluation tasks. In particular, we first explored the emergence of this ability in a developmental study (Experiment 1) using the same methodology (the Truth Value Judgment Task) and the same materials across different age groups of children. For the group of younger children, we replicated previous results: not all children at age 4 and 5 derived the pragmatic inference related to underinformative *some*, at least in a classical Truth Value Judgment Task. With respect to this finding, it is important to highlight, on the one hand, the bimodal distribution by children, already found in Guasti et al. (2005) for the 7-year-olds and adults in the Statement Evaluation Task. In case of our study, the fact that children are split

into two groups with respect to SI derivation might suggest a maturational process that is not yet completed for some children at age 5. On the other, our developmental study also extended previous results (e.g., Noveck, 2001) in showing that age 6 is a turning point in the development of the SI related to *some*. From age 6 on, children are adult-like and reject *some* when it is used underinformatively, consistently deriving the SI *not all*. We believe that this result raises a developmental issue: the ability to derive Scalar Implicatures seems to be linked to some general cognitive factor that is undergoing development around age 5. Except for a recent study by Katsos and Bishop (2011) that considered "tolerance" the key to explain children's behavior with pragmatic inference, no other study thus far has systematically investigated some of the preliminary factors that might contribute to children's failure or success with the computation of the SI related to *some*.

One possibility that we have explored in our paper is that children at age 5 lack a general pragmatic ability. For example, they might not be sensitive to the violation of conversational norms, or to the Maxim of Quantity in particular, which is a prerequisite to derive scalar implicatures. We tested this hypothesis by means of two different studies that involved some of the same children that were tested with a Truth Value Judgment Task and that failed, as a group, to reject underinformative *some*. We first administered the Conversational Violations Test (Surian et al., 1996) to score the children's ability to detect violations of conversational norms. Then, we administered the Felicity Judgment Task (Chierchia et al., 2001) to the nonpragmatic responders in the Truth Value Judgment Task reported as Experiment 1 (and a similar study that is not described in this paper). In the Felicity Judgment Task, the children had to choose the best description between an underinformative-*some* and an appropriate-*all* description. Surprisingly, most children performed at ceiling in both tasks, showing that they are sensitive to Gricean Maxims, particularly to the Maxim of Quantity (Experiment 3), and they are able to choose the most informative description between *some* and *all* (Experiment 4). Nonetheless, this pragmatic awareness is not sufficient to evoke pragmatic responses in a classical Truth Value Judgment Task, as also discussed in Katsos and Bishop (2011) comparing ternary and binary judgment tasks.

A second possibility is that the 5-year-olds are split into three groups: one group of children that has already caught up with the ability to derive SI; one group that has not; and a third group that lies in an undefined area in between. It is conceivable that the children in this third group are able to derive the pragmatic inference but need some additional cue to do so. In this respect, the large variability in the children's performance reported across different studies suggests that children are extremely vulnerable to manipulations of the experimental materials, design and setting. To enhance the probability of retrieving pragmatic responses by the 5-year-old children who showed no sensitivity to pragmatic infelicity, we manipulated our baseline study in two ways: in Experiment 2 we replaced the form *qualche* with the partitive (and plural) quantifier *alcuni dei*, and in Experiment 3 we attempted to prime the scale by providing the children with a correct description containing the stronger element in the scale (*all*) before asking them to judge the underinformative description with *some*. Contrary to our expectations, in none of these cases did we observe an increase in the derivation of the SI. The absence of such an improvement, however, might suggest that children's difficulties are not to be found at the level of pure lexicon or lexical access. In the last study reported (Experiment 6), we gave the children another chance to show their pragmatic competence. Inspired by a study on figure reversal by Gopnik and Rosati (2001), we manipulated the experimental design so as to "inform" the children about the ambiguity of *some* by showing them a situation in which it would be felicitously used (and was erroneously

described by *all*) before one in which it was used underinformatively (in which *all* was more appropriate). Interestingly, we found a significant improvement in this last experiment, and a rate of rejection of underinformative *some* (i.e., 72.5%) that had never before been attested in the literature for 5–year-olds in a Truth Value Judgment Task. This high rate of SI derivation is striking if we compare it, for example, to the significantly lower rate attested in Katsos & Bishop's Experiment 1 (less than 30% of SI computation) that employed the same task.

To conclude, we believe that a similar explanation might apply to the fact that many children failed with scalar inference in our first experiment and improved in our last experiment. Our suggestion is that the children's failure in deriving the SI related to *some* might be linked to other "cognitive" difficulties manifested by the children in other tasks (cf. Gopnik & Rosati, 2001; Rock et al., 1994). For example, it has been shown that 5 year old children have difficulty in detecting the ambiguity of ambiguous figures and reversing them without being informed of the ambiguity. Moreover, 3- and 4-year-old children still show difficulties in reversal even when informed. Gopnik and Rosati suggested that the immediate experience of perceptual reversals may rely on a broader understanding of multiple representations, a prerequisite for the awareness of ambiguity. We believe that this observation might be relevant in explaining the improvement on *some* in our last experiment, although it is speculation and needs to be addressed in future research.

Along these same lines, one can interpret the results obtained with bilingual children in different linguistic and nonlinguistic tasks. For instance, Siegal, Matsuo, and Pond (2007) compared bilingual and monolingual children on their ability to derive SIs and found that bilinguals outperformed monolinguals. One possible explanation of the bilinguals' facility in this task is their greater experience in shifting language/attitude depending on the interlocutor/situation. This enhances their conversational understanding (Siegal & Surian, 2004). An advantage of bilinguals over monolinguals was also attested in the Conversation Violation Test by Siegal et al. (2009). In the study discussed above, the authors suggested that bilingualism might enhance children's ability to appreciate effective communicative responses, conferring an advantage on their conversational understanding in general. Most interestingly, other studies outside the traditional world of linguistics and psycholinguistics have already shown a positive effect of bilingualism on children's ability to "inhibit" a preponderant response in tasks of evaluation of executive functions (Bialystok & Martin, 2004; Bialystok & Senman, 2004) and a better performance in reversing ambiguous figures and reassigning a different interpretation to them (Bialystok & Shapero, 2005).

To sum, we suggest that at least three factors might interplay to explain children's behavior with underinformative-*some*. One is the maturation of the lexicon that, in the case of scalar items, involves two layers of lexical representation, the basic meaning and the scale: the link between these two needs to be acquired and automatized. This process might be completed before age 5 but might vary across different scalar items (cf. Barner & Bachrach, 2010, for an explanation along these lines). The second factor is the ability to shift one's strategy and perspective. This ability is not yet mature at age 5 (as shown by tests for executive functions) but might be improved with "exercise" and with explicit/implicit instructions that make children aware of the ambiguity. This improvement has been shown for bilinguals, who have more experience in shifting perspective (Siegal et al., 2009) and it has been attested in cases of ambiguous figures when children are informed of the alternative representations within the same image (Gopnik & Rosati, 2001). The third factor (which might be intertwined with the second) is the influence of the task, the experimental materials and design used to test children (and adults). As we have seen, 5-year-old children's rejection of underinformative-*some* varies significantly across studies that

employ the same methodology, that is, the Truth Value Judgment Task. The children were successful at around 10% in Papafragou & Musolino (2003, Experiment 1), around 30% in Katsos and Bishop (2011, Experiment 1) and Katsos et al. (2011), around 40–50% in Experiments 1, 2, and 3 reported here, and reaching almost 75% in Experiment 6. Compared to previous tasks that involved a binary judgment on the use of underinformative-*some*, we believe that the following features contributed to the success obtained in this last experiment. First, the children were exposed to short stories that were acted out in full with props, instead of being shown pictures or semistatic images on a PC monitor (as happened, e.g., in the studies by Katsos & Bishop, 2011, and Katsos et al., 2011). Second, we tried to place some emphasis on the quantity of characters and objects involved in the action, by adding comments about quantity during the narration. We also tried to render the use of *some* all the more felicitous by not using it to refer to small numerosities (as was done by, e.g., Papafragou & Musolino, 2003). Last but not least, we tried to inhibit the logical response by rendering the ambiguity of *some* all the more salient and relevant in the task, by "informing" children that there could be many different ways to talk about the "quantity" of objects. While the first three manipulations were common to Experiments 1, 2 and 3 reported in this paper, the last manipulation pertained to the last experiment only.

In any case, the large variability observed across different studies that employ a similar methodology reaffirms once again that children are very sensitive to subtle changes in the way the task is conceived and administered and extremely vulnerable to flaws in the experimental design, as was shown for other fields of study as well (cf. also Crain & Thornton, 1998). We believe that the recent proposal made by Katsos and Bishop (2011) that children's over-acceptance of underinformative-*some* might be explained in terms of "pragmatic tolerance" can be subsumed under this more general consideration. According to our evaluation, the claim that children are generally more tolerant than adults and thus refrain from deriving SIs cannot be disentangled from this more general observation: children do not derive Sis; that is, they are tolerant of pragmatic infelicities, whenever the task designed to address their pragmatic ability is not sensitive enough to capture their competence. This happens when they are not instructed well enough to be able to render the derivation of the SI relevant for the purpose of solving the task. More generally, we believe that "tolerance" is simply a by-product of the task demands, which affect adult responses as well. In the end, we believe that children are *not* more logical than adults, *nor* more tolerant if they are given the best chance to be pragmatic.

ACKNOWLEDGMENTS

We are grateful to Luca Surian for allowing us to use the Conversational Violations Test. We also thank Joke De Lange, Chiara Iogha and Marco Bernardi for their generous help in testing children and Claudia Caprin for her help with the statistical analysis on preliminary data. Participation in the Bicocca RRReading-Group was also fruitful (as well as enjoyable); special thanks go to Marco Marelli and Davide Crepaldi for sharing their expertise. We are also indebted to the three anonymous reviewers and the editors of this journal, for their sharp and fruitful suggestions that have contributed greatly to the final shape of this paper. Lastly, we cannot forget all the children, parents, directors and staff of the schools that participated in these studies; we send them our utmost gratitude for their kind cooperation.

REFERENCES

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, *60*(1), 40–62.

Barner, D., Chow, K., Yang, S. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, *58*(2), 195–219.

Berg, E. A. (1948). A simple, objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15–22.

Braine, M., & Roumain, B. (1981). Children's comprehension of "or": Evidence for a sequence of competencies. *Journal of Experimental Child Psychology*, *31*, 46–70.

Bialystok, E., & Martin, M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, *7*, 325–339.

Bialystok, E., & Senman, L. (2004) Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development*, *75*(2), 562–579.

Bialystok, E., & Shapero, D. (2005). Ambiguous benefits: The effect of bilingualism on reversing ambiguous figures. *Developmental Science*, *8*(6), 595–604.

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance theory: Applications and implications* (pp. 179–236). Amsterdam, the Netherlands: John Benjamins.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and Beyond*. Oxford, UK: Oxford University Press.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590.

Chierchia, G., Crain, S., Guasti, M. T., & Thornton, R. (1998). Some and or: A study on the emergence of logical form. *Proceedings of the Boston University Conference on Language Development* (Vol. 22, pp. 97–108). Somerville, MA: Cascadilla Press.

Chierchia, G., & McConnell-Ginet, S. (2000). *Meaning and grammar*. Cambridge, MA: The MIT Press.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures. *Proceedings of the 25th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Chierchia, G., Guasti, M. T., Gualmini, A., Meroni, L., Crain, S., & Foppolo, F. (2004). Semantic and pragmatic competence in children's and adults' comprehension of or. In I. A. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 283–300). New York, NY: Palgrave Macmillan.

Chierchia, G., Fox, D., & Spector, D. (in press). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, K. von Heusinger (Eds.), *Handbook of semantics*. Berlin Mouton de Gruyter.

Clahsen, H., Hadler, M., & Weyerts, H. (2004). Speeded production of inflected words in children and adults. *Journal of Child Language*, *31*, 683–712.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar. A guide to experiments on the acquisition of syntax*. Cambridge, MA: MIT Press.

Doitchinov, S. (2005). *Naturalistic and experimental data in language acquisition: The case of epistemic terms*. In S. Kesper & M. Reis (Eds.), *Linguistic evidence: Empirical, theoretical, and computational perspectives* (pp. 123–144).

Fox, D. (2007). Free choice disjunction and the theory of scalar implicatures. In U. Sauerland & P. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). New York, NY: Palgrave MacMillan.

Gazdar, G. (1979). *Pragmatics. Implicature, presuppositions and logical form*. New York, NY: Academic Press.

Geurts, B. (2011). *Quantity implicatures*. New York, NY: Cambridge University Press.

Gopnik, A., & Rosati, A. (2001). Duck or rabbit? Reversing ambiguous figures and understanding ambiguous reference. *Developmental Science*, *4*(2), 174–182.

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*, 377–388.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics*, Vol. 3. New York, NY: Academic Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116, 42–55.

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.

Horn, L. R. (1972). On the semantic properties of logical operators in English. (Unpublished doctoral dissertation). UCLA.

Jaeger, T. F. (2008). Categorial data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language (Special Issue on Emerging Data Analysis)*, *59*, 434–446.

Katsos, N., & Bishop, D. (in press). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*.

Katsos, N., Roqueta, C. A., Estevan R. A. C., & Cummins, C. (in press). Are children with specific language impairment competent with the pragmatics and logic of quantification? *Cognition*.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology 62*(4), 498–525.

Levinson, S. C. (2000). *Presumptive meanings—the theory of generalized conversational implicatures*. Cambridge, MA: MIT Press.

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

Papafragou, A. (2003). From scalar semantics to implicature: Children's interpretation of aspectuals. *Journal of Child Language*, *33*(4), 721–757.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantic-pragmatics interface. *Cognition*, *80*, 253–282.

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

Pinker, S. (1991). Rules of language. *Science*, *253* (5019), 530–535.

Pinker, S. (1999). *Words and rules. The ingredients of language*. New York, NY: Basic Books.

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

R Development Core Team. (2005). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reinhart, T. (2006). *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.

Rock, I., Gopnik, A., & Hall, S. (1994). Do young children reverse ambiguous figures? *Perception*, *23*(6), 635–644.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy*, *27*(3), 367–391.

Sauerland, U. (2005). On embedded implicatures. *Journal of Cognitive Science*, 5.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transcripts of the Royal Society of London B*, *298*, 199–209.

Siegal, M. (2008). *Marvelous minds: The discovery of what children know*. Oxford, UK: Oxford University Press.

Siegal, M., & Surian, L. (2004). Conceptual development and conversational understanding. *Trends in Cognitive Sciences*, *8*(12), 534–538.

Siegal, M., Matsuo, A., & Pond, C. (2007). Bilingualism and cognitive development: Evidence from scalar implicatures. In Y. Otsu (Ed.), *Proceedings of the Eighth Tokyo Conference on Psycholinguistics* (pp. 265–280). Tokyo, Japan: Hituzi Syobo.

Siegal, M., Iozzi, L., & Surian, L. (2009). Bilingualism and conversational understanding in young children. *Cognition*, *1*(10), 115–122.

Smith, C. L. (1980). Quantifier and question answering in young children. *Journal of Experimental Child Psychology*, *30*, 191–205.

Sperber, D., & Wilson, D. (1986/1995). *Relevance: Communication and cognition*. Oxford, UK: Blackwell.

Surian, L., Baron-Cohen, S., & Van der Lely, H. (1996). Are children with Autism deaf to Gricean maxims? *Cognitive Neuropsychiatry*, *1*(1), 55–71.

Surian, L., & Siegal, M. (2001). Sources of performance on theory of mind tasks in right hemisphere-damaged patients. *Brain and Language*, *78*, 224–232.