

A Developmental Investigation of Processing Costs in Implicature Production

Nausicaa Pouscoulous

*L2C2, Institut des Sciences Cognitives, CNRS & Université de Lyon
Lyon, France*

and

*Institut Jean Nicod
Paris, France*

Ira A. Noveck

*L2C2, Institut des Sciences Cognitives, CNRS & Université de Lyon
Lyon, France*

Guy Politzer

*Institut Jean Nicod, CNRS
Paris, France*

Anne Bastide

*CNRS & Université de Paris 8
Paris, France*

Much developmental work has been devoted to *scalar implicatures*. These are implicitly communicated propositions linked to relatively weak terms (consider how *Some* pragmatically implies *Not all*) that are more likely to be carried out by adults than by children. Children tend to retain the linguistically encoded meaning of these terms (wherein *Some* is compatible with *All*). In three experiments, we gauge children's performance with scalars while investigating four factors that can have an effect on implicature production: (i) the role of (the presence or absence of) distractor items; (ii) the nature of the task (verbal judgments versus action-based judgments); (iii) the choice of scalar expression (the French quantifier *quelques* versus *certains*); and (iv) the type of scale that contextualizes the weak utterance (the affirmative *All* versus the negative *None*). Experiment 1 replicated

earlier findings showing that 9-year-olds are more likely than adults to consider as true statements such as *Some turtles are in the boxes* (uttered when all turtles are in the boxes) while employing the quantifier *certain*s in a truth evaluation task containing multiple distractor items. The task in Experiment 2 increased implicature production across all ages (4-, 5-, and 7-year-olds as well as adults) but maintained the developmental effect while using *quelques* in an action-based task containing no distractor items. Experiment 3 showed that 9-year-olds are more likely to produce implicatures with *quelques* than they are with *certain*s in the action task while adults are not affected by the choice of term. Overall, these results identify seemingly harmless task features that can prevent even older children (9-year-olds) from carrying out implicatures (e.g., through the inclusion of distractors) while also showing how implicature production among even young children (4- to 5-year-olds) can be facilitated by task features (e.g., the use of an action task) and without the introduction of special training.

1. INTRODUCTION

When a professor says “Some students failed the exam,” the hearer is entitled to infer pragmatically that *not all of the students failed*. Theoretical work on the interface of semantics and pragmatics has played a critical role in understanding how this inference works (Horn (1989), Levinson (2000), Carston (1998), Chierchia (2004), Wilson and Sperber (2004)). The general consensus is that the weaker term (e.g., the quantifier *some*), while logically compatible with a stronger term from the same scale (e.g., *all*), prompts the inference because the speaker did not *use* the stronger term. As can be seen, the scalar expression *some* may be interpreted in two different ways: either with an inference-driven, pragmatic reading, which excludes *all*, or with its literal, semantic meaning, which is compatible with *all*. This description of *some* can be extended to a host of *scalar* terms such as the disjunction *or*, which can implicitly exclude the stronger conjunction *and* (as in (1a)); the modal *might*, which excludes *must* (as in (1b)); and the verb *start*, which implicitly excludes the more informative *finish* (as in (1c)):

- (1) a. Molly is going to be picked up by Mommy or Daddy (not *both* of them).
- b. Steven might be home (it is not true that he *must* be there).
- c. Yogi started to paint his house (he did *not finish* it).

Recent experimental work has focused on children’s interpretation of scalar expressions like *some* and *or*. This work has brought to light a surprising developmental trend showing that children are more likely than adults to treat the weak term (e.g., *some*) as compatible with one that is stronger on the scale (*all*). For instance, Noveck (2001), who was the first to conduct systematic

experiments on children's treatment of scalar expressions, showed that 8- to 10-year-olds do not make scalar inferences associated with *some* and *might* to the same degree as adults. In one of Noveck's (2001) experiments, most children accepted sentences such as 'Some giraffes have long necks' while adults tended to reject them as false (on the grounds that all giraffes have long necks). This effect has proven to be robust as reported by recent studies (described below) that specifically aimed to investigate the effect and by virtue of classic studies that inadvertently included equally underinformative cases, including three where *or* was used to describe scenarios in which *and* would have been more appropriate (Paris (1973), Sternberg (1979), Braine and Romain (1981)).

The recent work has focused on younger children, e.g., 5-year-olds, who have been shown to be less likely than adults to generate scalar inferences associated with expressions such as *some* and *start* (Papafragou and Musolino (2003))¹ as well as with *or* (Chierchia, Guasti, Gualmini, Meroni, and Crain (2004); for a review, see Noveck (2004), Pouscoulous and Noveck (2004), and Siegal and Surian (2004)). In this recent literature, many have gone on to argue that children's performance with scalar expressions ought not to be mistaken for an incapacity to draw scalar inferences and provide evidence in support of this view (Chierchia et al. (2004), Papafragou and Musolino (2003), Guasti et al. (2005)). For example (and as we will discuss in detail later), Papafragou and colleagues (Papafragou and Musolino (2003), Papafragou and Tantalou (2004)) showed that children as young as 5 are generally able to produce implicatures when they are given appropriate training and when the experimental circumstances enhance the contrast between the weak, underinformative test sentence and the stronger fact that stems from the scenario.

Although it is highly relevant to demonstrate that one can find pragmatic competence among younger children, these arguments overlook Noveck's (2001) more general point, which is that a scalar implicature comes at some (perhaps even a small) cost in terms of effort and that this cost is less likely to be incurred by those who have fewer resources available. Assuming that children have fewer cognitive resources available to them than adults, it follows that this pragmatic inference is bound to appear less accessible to children. Moreover, as a task gets harder (say, by making it part of a reasoning task, by introducing distractions that conceal the purpose of the experimenter-participant exchange, or by using relatively difficult scalar terms), even fewer resources are bound to be available as the scenario offers children an opportunity to produce implicatures. In the present

¹Papafragou and Musolino's (2003) study also investigated number terms (e.g., a puppet would use "three" to describe a situation where a larger number would have been more appropriate). They showed that 5-year-olds ultimately deal differently with number words than they do with other weak terms in that cardinals are more likely to get an adult treatment (even though the children are not completely adult-like). This finding confirms the prevailing view in recent linguistic theory, according to which number words do not have a scalar semantics (Horn (1992), Carston (1998), Geurts (1998; 2006), among others).

work, we propose that it is hard to find routine employment of scalar inferences among children, not because they lack the required pragmatic competence *per se*, but because their cognitive resources are more limited than those found among adults. We present evidence that shows that implicature production can be readily forestalled among older (9-year-old) children by virtue of task features that render a task more difficult (Experiments 1 and 3) as well as encouraged—without training—among younger (4- to 5-year-old) children through facilitating features (Experiment 2). In order to set the stage for the three experiments to be presented, we begin by reviewing two papers (Papafragou and Musolino (2003), Guasti et al. (2005)) that aimed to show that 5-year-olds and 7-year-olds are more competent than one would think based on Noveck (2001).

Papafragou and Musolino (2003, Experiment 1) actually first confirmed the developmental effect described by Noveck (2001) by showing that 5-year-olds are less likely than adults to produce enrichments with *some*, *start*, and *three*. Afterward, they modified the experimental setup in two ways in order to carry out their second experiment. First, participants received training at enhancing their awareness of pragmatic anomalies before they were tested. Children were told that the puppet would say “silly things” and that the point of the game was to help the puppet say it better (e.g., they would be asked whether a puppet described a dog appropriately by saying “this is a little animal with 4 legs”). In the event that the child did not correct the puppet, the experimenter did. Second, the paradigm puts the focal point on a protagonist’s performance. Unlike in Experiment 1, where participants were asked to evaluate a quantified sentence like *Some horses jumped over the fence* (when in fact all the horses did), the paradigm in their second experiment raises expectations about the stronger case (*all*). Participants would hear a test sentence like “Mickey put some of the hoops around the pole” (after having been shown to succeed with all of the hoops), but they were also told how Mickey claims to be especially good at this talent and that this is why another character challenges him to get all three around the pole. With these changes, 5-year-olds were more likely to produce scalar implicatures than they were in the first experiment. Nevertheless, the 5-year-olds still produced enrichments less often than adults, which indicates that—even with training and with a focus on a stronger contrast—pragmatic enrichments require effortful processing among these children.²

Guasti et al. (2005), who argue that pragmatic enrichments ought to be as common among 5-year-olds as they are among adults, further investigated the findings of Noveck (2001) and Papafragou and Musolino (2003). In their first

²Papafragou and Tantalou (2004) aim to show that 5-year-olds can be encouraged to produce scalar implicatures at adult levels. However, their data are based on a nonstandard paradigm (i.e., it gives participants no justifiable reason to accept the minimal interpretation of a term such as “some”). Moreover, much of the study’s claims are based on children’s self-reports and these lead to the conclusion that at most 56% of Papafragou and Tantalou’s participants derived scalar implicatures.

experiment they replicated the finding from Noveck (2001, Experiment 3) with *some* and used this as a baseline to study independently the role of the two factors manipulated by Papafragou and Musolino (2003). One is the role of training and how it affects children's proficiency at computing implicatures (Experiments 2 and 3) and the other is the role of placing emphasis on the outcome of a scalar implicature (Experiment 4). Experiments 1 through 3 showed that training young participants to give the most specific description of a given situation can indeed have a major effect on performance. While their initial experiment showed that 7-year-olds accept statements such as *Some giraffes have long necks* 88% of the time (compared to 50% for adults), when trained in this manner their acceptance rate drops to 52%, making them adult-like. Nonetheless, this effect is short-lived, that is, it does not persist when the same participants are tested a week later (Experiment 3). In the last experiment, the authors used a truth-value judgment task (Crain and Thornton (1998)) and rendered the *all* alternative more salient in context. This was achieved, for instance, by presenting the participants with a story featuring several characters who were deciding whether the best way to go collect a treasure was to drive a motorbike or ride a horse. After some discussion, all of them chose to ride a horse. In this way it is made clear that the sentence subjects have to judge, "Some of the characters chose to ride a horse," is underinformative and thereby a "bad" description of events. The results indicated that when the outcome of scalar implicatures is highly relevant in context children will compute them in an adult-like manner.

Although extremely instructive, these results leave a number of questions unanswered. Overall, Guasti et al.'s (2005) and Papafragou and Musolino's (2003) experiments demonstrate that, under the right experimental circumstances, children will produce more implicatures, but even then the younger ones do not always produce them as routinely as adults do. As suggested by Noveck (2001) (also, see Guasti et al. (2005)), a plausible explanation for this delay is that inferring scalar implicatures requires *effort* and that children have fewer available resources than adults. Although there is evidence that scalar inferences involve additional processing costs for adults (Noveck and Posada (2003), Bott and Noveck (2004), Breheny, Katsos, and Williams (2006), De Nys and Schaeken (2007)), so far none has been proffered to explain the developmental delay in children.

What are the sources of such costs in the developmental studies? For the purpose of the present article we identify three factors whose inclusion in past tasks might well "weigh down" implicature production. Two of these are related to the nature of the task and one to the complexity of the linguistic material. We turn first to factors stemming from the task.

One factor that appears to vary across tasks is the presence or absence of distractors. Whereas Noveck (2001) used up to seven control sentences for every test statement, others (e.g., Papafragou and Musolino (2003)) employ as few as one control for every test statement. It could well be the case that a higher

number of controls, which typically avoid highlighting the (weak-strong) contrasts that are so critical for scalar inferences, leaves children unprepared to detect such contrasts when implicatures are potentially called for. A second factor concerns task requirements. So far, developmental studies on scalars have used verification tasks (truth value judgment tasks or similar experimental designs). In these experiments, participants are typically asked whether or not sentences uttered by a puppet are in some way correct or true. There are of course variations on this theme: Children might be asked whether the puppet said it in the best possible way (Papafragou and Musolino (2003)) or which of two puppets said it best (Chierchia et al. (2004)). But all these procedures are nontrivial because they require metalinguistic judgments, which even for adults are hard in many cases (see the nonperfect rates of correct responses to control problems among adults in Bott and Noveck (2004)). There is a broad consensus in the acquisition literature that metalinguistic tasks are harder than tasks that obviate the need for a verbal response, e.g., those tasks whose dependent variable is based on an action (Michael Tomasello, personal communication).³ This might well be the case here. We will discuss this further in the introduction to Experiment 2.

Apart from the fact that the experimental task as such may present difficulties for children, the linguistic material used may also play a role in distracting participants from making implicatures. Although paradigms have been designed to simplify implicature tasks and highlight contrasts, there are other potential sources of complexity that have not been investigated so far having to do with the test utterances themselves. Here we directly investigate one factor that has been overlooked—the word-choice for the scalar term—and we introduce another. The overlooked issue concerns the fact that many languages, including languages used in earlier experiments (e.g., French, Greek, and Italian), have several ways of translating English *some*, the scalar term that has received the most attention in developmental studies. French, for example, distinguishes between *quelques* and *certains*, both of which are existential plurals and, as we will discuss later, there are reasons to believe that the latter (which is a partitive) is more complex than the former. One of the issues to be addressed in our experiments is whether this difference in complexity is reflected in children's performance on implicature tasks.⁴

To further investigate the role that linguistic factors can play with respect to scalar inference, we introduce negative sentences because, generally speaking, negated expressions are more complex than their positive counterparts. Negative expressions take longer to process, cause more errors, and are harder to retain

³For further discussion see the *Journal of Child Language* (issue 2, 2004). For a detailed discussion of the limitations of the truth judgment evaluation task see also Musolino (2006) as well as Musolino and Lidz (2006).

⁴We point out straightaway, to avoid misunderstanding, that despite appearances the meaning of the French word “certains” is not equivalent to the English “certain.” For a semantic analysis of the French terms *certains* and *quelques* see Corblin (2001) and Gondret (1976).

than positive ones (see Horn (1989, chapter 2) for a survey of the experimental literature; for more recent work, see Prado and Noveck (2006)). Scalar inferences arise in negative utterances too, although in such cases scales are reversed. For instance, a scalar implicature has the potential to arise out of *Some children are not in the classroom* because it negates the stronger claim that there are no children in the classroom (the result being that one can conclude pragmatically that at least one child is in the classroom). So far, experimental studies on scalar inferences have not taken into account possible differences between negative and affirmative sentences.

Besides the four factors that address issues of complexity, there are two other matters that we highlight about our experiments. The first is that we aim to access *spontaneously drawn* scalar inferences as opposed to those following specific training. So far, performances revealing impressive pragmatic abilities among the youngest children (e.g., 5-year-olds) occur when the task includes training to detect pragmatic anomalies (e.g., Papafragou and Musolino (2003, Experiment 2)) and the recent tasks that employ training (Guasti et al. (2005)) study 7-year-olds. In the present work, we aim to establish the extent to which younger children can make scalar inferences without previous training. Along with most researchers working on the acquisition of scalar terms, we believe that there is no general pragmatic incapacity among children and they should be able to make scalar implicatures spontaneously when a task is easy enough. It would be more convincing if this ability could be demonstrated without special training or cues. Second, we also endeavor to reintroduce a developmental perspective. In an effort to highlight what young children can do, more recent experimental studies have focused on one young age group at a time as they were compared to adults. In the present work, several ages are investigated ranging from 4-year-olds to 9-year-olds to adults in order to sketch a more precise developmental curve.

We now turn to the three experiments that address the issues just raised. Experiment 1 serves as a prelude in that it replicates Noveck's (2001) pragmatic-developmental effect with respect to the weak quantifier *some* across two age groups (9-year-olds and adults). This experiment is relevant in two ways. First, it re-establishes that the developmental effect can be found by testing relatively old and linguistically competent children (as we just indicated, many of the more recent studies focus on younger children). Second, it provides a paradigm that can be readily modified and that can then help test our hypotheses. In Experiment 2, we make three changes to the design of Experiment 1. First, distractors, which were numerous in Experiment 1, are removed. Only key critical and control items are included, which simplifies the participant's task. Second, the nature of the task is changed: The truth evaluation (the dependent variable) is determined by an action and not a metalinguistic evaluation. Third, the French indefinite *certain* which was used in the first experiment is replaced by *quelques*. In order to establish whether the development of implicature production increases

progressively with age, Experiment 2 tests children from three age groups (4-, 5-, and 7-year olds) plus adults. In Experiment 3, we focus on what we considered to be the most intriguing source of complexity from a linguistic point of view, that is, the difference between the French quantifiers *quelques* and *certains*. Finally, as a matter of course, all three experiments included both affirmative and negative sentences, in order to uncover how this distinction affects children's and adults' performance.

2. EXPERIMENT 1

Unlike Noveck (2001; based on Smith (1980)), which presented participants with quantified sentences that drew on encyclopedic knowledge (participants were required to evaluate sentences such as "Some elephants have trunks"), we present participants with questions about four boxes whose contents are directly visible to them. We adopted this sort of presentation for two reasons. First, we avoid differing levels of world knowledge that are likely to vary with age. While 9-year-olds would be expected to know that "giraffes have long necks" or that "elephants have trunks," an artificial scenario leaves less to chance. Second, it could be argued that the quantifiers *all* and *some* work in an exceptional way when used to quantify over a large domain as in *all the animals of a species* (or, in general, over *all the objects belonging to a category*) and that these work in a more standard way when they are used with respect to a clear set that is referenced in context (one that is clearly identifiable by the hearer).⁵ The present scenario steers clear of each of these worries.

The experimental paradigm placed four cardboard boxes in front of participants with different plastic animals placed in and around the boxes. The participants were then asked whether or not they agreed with a puppet that made statements about the scenario. The responses to these statements serve as the dependent variable. The main test item is the French equivalent of "Some turtles are in the boxes," where all of the turtles are in the boxes. If participants make the implicature, they would be expected to disagree with the puppet, whereas if they treat the word *some* in its logical sense they would agree with the statement. As in Noveck (2001), the experiment is conducted in French and uses *certains* as a translation of *some* (see Déret (1998) for another study with children and adults using *certains* as an expression of *Some*). We chose to test 9-year-olds and adults. Our predictions were that we would find roughly the same results described in Noveck (2001). Children ought to appear more logical than adults because adults are expected to make the implicature more readily than children.

⁵For a criticism of experimental designs using sentence judgments in investigating scalar implicatures in children, see also Guasti et al. (2005).

2.1. Method

2.1.1. *Participants.* Twenty-three 9- to 10-year-olds (mean age 9;6) and 19 adults participated. Child participants were recruited from a suburban, middle-class, after-school program in Boulogne (France) and adults were architecture students at the *École Camondo* in Paris. All the participants were French native speakers.

2.1.2. *Materials and procedure.* Participants were presented with a scenario involving four cardboard boxes and a selection of plastic animals. The animals were arranged so that statements could be made using quantifiers such as *all (tous)* and *some (certains)* and a particular type of animal. Statements were made via a puppet handled by the experimenter. For example, the experimenter could say “All the dolphins are in the boxes,” and the participant had to determine whether the statement was true or false (by agreeing or disagreeing). In all the experiments here, the experimenter read each statement in a similar fashion throughout and without undue emphasis on any one word.

Participants were presented with 20 sentences concerning the scene in front of them. One was the main test item (“Some turtles are in the boxes,” when in fact all turtles were in the boxes), another nine were control items (we refer to these 10 collectively as ‘critical sentences’), and 10 were filler sentences designed to disguise the purpose of the study. The critical sentences were created by crossing two factors: the type of quantifier and the type of animal. Three different types of quantifiers were used (*All, Some, None*), as well as the negation of *some (Some are not)*; and three types of animal (dolphins, some of which were inside the boxes and some outside; turtles, each of which was inside a box; and elephants, all of which were outside the boxes). This produces up to 12 possible statements. Outside the main item and the utterance “Some elephants are not in the boxes” discussed below, two straightforward statements of the 12 potential items were randomly removed to reduce the total number of items and better disguise the purpose of the study. See Figure 1 for a graphical representation of the scenario and Table 1 for the critical sentences and their truth values.

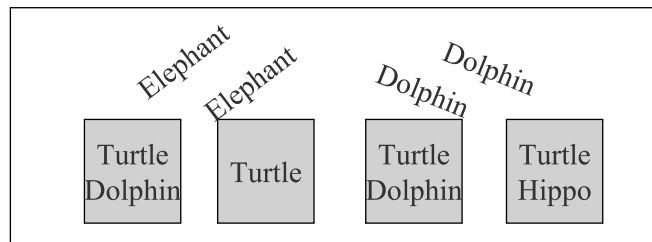


FIGURE 1 An illustration of the scenario presented in Experiment 1.

TABLE 1
 Percentage of Children and Adults Who Responded Logically to the Critical Statements
 Presented in Experiment 1, Including the One of Central Interest (in Bold)

<i>Statements</i>	<i>Truth-Value (Logical Answer)</i>	<i>Children (N = 23)</i>	<i>Adults (N = 19)</i>
All the turtles are in the boxes. (Toutes les tortues sont dans les boîtes.)	True	100%	100%
Some turtles are in the boxes. (Certains tortues sont dans les boîtes.)	True	91%	53%
Some turtles are not in the boxes. (Certains tortues ne sont pas dans les boîtes.)	False	100%	100%
All the dolphins are in the boxes. (Tous les dauphins sont dans les boîtes.)	False	100%	100%
Some dolphins are in the boxes. (Certains dauphins sont dans les boîtes.)	True	100%	100%
No dolphin is in a box. (Aucun dauphin n'est dans une boîte.)	False	91%	100%
Some dolphins are not in the boxes. (Certains dauphins ne sont pas dans les boîtes.)	True	100%	100%
Some elephants are in the boxes. (Certains éléphants sont dans les boîtes.)	False	100%	95%
No elephant is in a box. (Aucun éléphant n'est dans une boîte.)	True	96%	100%
Some elephants are not in the boxes. (Certains éléphants ne sont pas dans les boîtes.)	True	70%	63%

Notes. The statements were uttered by a puppet and participants were required to indicate whether they “agree” or “disagree” with him. “True” in the table corresponds with “agree.”

As mentioned earlier we also included negative sentences with *some* in the design. The utterance “Some elephants are not in the boxes” shares many of the features of our main test item because it is underinformative relative to the scenario (i.e., none of the elephants are in the boxes). A logical interpretation of *some are not* is compatible with a more informative quantifier *none*, making an “Agree” response justifiable. However, if the participant makes a pragmatic interpretation via an implicature (where *Some are not* would imply *Not None*), a “Disagree” response is justified. No prior study has studied underinformative sentences in such negative scenarios. Given that their negative character makes them fairly complex, however, we are not confident that children (or adults, for that matter) would be able to properly process the sentence, let alone produce the implicature to which this utterance could potentially give rise.

The filler questions concerned properties of the animals that were irrelevant to the study, such as “All the turtles are red” and “There is a turtle and an elephant in the same box.” Table 2 presents the filler items and their truth values.

TABLE 2
 Percentage of Children's and Adults' Correct Responses to the Distractor Items
 Presented in Experiment 1

<i>Statements</i>	<i>Truth- Value</i>	<i>Children (N = 23)</i>	<i>Adults (N = 19)</i>
Some animals in the boxes are red. (Certains animaux dans les boîtes sont rouges.)	True	100%	100%
One of the animals out of the boxes lives in trees. (Un des animaux hors des boîtes vit dans les arbres.)	False	96%	95%
None of the animals in the boxes know how to fly. (Aucun des animaux dans les boîtes ne sait voler.)	True	87%	100%
There is a hippopotamus and a dolphin in the same box. (Il y a un hippopotame et un dauphin dans la même boîte.)	False	100%	100%
There is a turtle and a dolphin in the same box. (Il y a une tortue et un dauphin dans la même boîte.)	True	96%	89%
There is an elephant and a turtle in the same box. (Il y a un éléphant et une tortue dans la même boîte.)	False	96%	100%
There is an elephant and a panda in the same box. (Il y a un éléphant et un panda dans la même boîte.)	False	100%	100%
All the animals in the box could live in water. (Tous les animaux dans les boîtes peuvent vivre dans l'eau.)	True	96%	84%
There are an odd number of dolphins. (Il y a un nombre impair de dauphins.)	False	96%	100%
There is a box with three animals. (Il y a une boîte qui contient trois animaux.)	False	100%	100%

Note. See Table 1 notes.

The participants, tested individually, were introduced to a puppet named George. They were informed that the puppet says things that are right sometimes and wrong sometimes. The participants were then told that the puppet would state things about the boxes and the animals displayed in front of them (as described above) and that they were to say if they agreed or disagreed with the puppet. Participants were presented with three very simple familiarization questions: "What kind of animals do you see in and around the boxes?", "How many elephants do you see?", and "How many turtles are there?", in order to make sure they recognized all the types of animals displayed in front of them and that they had properly looked at the scenario before beginning the experiment.⁶ These were not expected to pose any difficulty and they did not. The experimenter then went through the entire set of 20 sentences with each participant. Three random orders of the statements were prepared.

⁶Note that the familiarization technique used here, and in Experiments 2 and 3, merely confirms that the participants appreciate the details of the scenario. It is not comparable to the training given in Papafragou and Musolino (2003, Experiment 2) nor to the one used in Guasti et al. (2005, Experiment 2).

2.2. Results and Discussion

Both children and adults had high rates of correct responses to the collection of 18 unambiguous questions—97% for each group. Table 1 lists participants' performance with the critical statements and Table 2 with the distractors. One can see that the 9-year-olds were adult-like in their responses to the eight unambiguous control items and to the 10 filler items (all $ps > 0.1$ using separate χ^2 tests for each question). As expected, children were more likely than adults to respond logically to the main test item, "Some turtles are in the boxes": 91% versus 53%, respectively; $\chi^2 = 8.05$, $p < .005$. The 91% figure is higher than one would predict based on chance assumptions, $\chi^2 = 15.7$, $p < .001$.⁷ There were no effects due to order.

This finding validates the anticipated pragmatic-developmental effect and provides further support for its robustness. Although this task is novel for an implicature paradigm, the use of arbitrary materials is not new. Noveck (2001, Experiments 1 and 2), Papafragou and Musolino (2003), and Guasti et al. (2005, Experiment 4) used arbitrary materials to investigate scalar implicatures. Our main goal was to verify that the present methodology had its intended effect before making changes that ought to encourage even younger children to make the implicature.

Before turning to Experiment 2, where we describe how we simplified the scenario, we summarize the participants' performance with the utterance "Some elephants are not in the boxes." One can see that the children and adults provided comparable levels of logically correct performance (70% and 63% responded affirmatively, respectively). Neither of these rates defies chance predictions. It is difficult to interpret this result because it could mean that this utterance is in fact difficult to process for both groups or perhaps that the adults are purposefully equivocal while the children have difficulty. We will continue to track this type of utterance in the remaining experiments so that we can establish which of these two possibilities is more likely.

3. EXPERIMENT 2

The goal of the second experiment was to demonstrate that even very young children *are* capable of making the implicature with the box paradigm developed in Experiment 1. As indicated in the Introduction, we made three changes that were aimed at making the task easier and at encouraging pragmatic responses in children. More specifically, our overall aim was to come up with ways that would spare younger children precious cognitive resources that they may in turn use to compute the scalar implicature linked to the weak quantifier.

⁷All the responses in the tasks to be described in this article can be of one of two types; thus references to predictions based on chance are .5 here and elsewhere.

First, we changed the scenario. This experiment concerns only tokens that were in boxes; that is, there are no questions about other objects nor are there questions about tokens outside the boxes, making it relatively free of distractors. This makes the task more direct and potentially more obvious.

Second, we changed the means of responding. Until now, tasks required some sort of metalinguistic judgment (of true or false, agree or disagree, “said well” or not), which is arguably hard to do for a child as young as 4 (in fact, some of our own pilot tests with 4-year-olds showed that they were dumbfounded by such requests). We therefore investigated comprehension of quantified sentences using other means and took, as inspiration, developmental paradigms that rely on some form of action (for key references and reviews of language acquisition experiments using nonverbal responses among young children see, e.g., Chomsky (1969), Pinker (1995), Bloom (2000), and Tomasello (2003)). This led to scoring based on a participant’s gesture. Concretely, consider a scenario in which two out of five boxes contain a token and the puppet says “I would like all the boxes to contain a token”: An appropriate response would be to add a token to each of the empty boxes. The main test item in this experiment arises when the puppet utters “I would like some boxes to contain a token” when the scenario displays each of five boxes already containing a token. If participants believe that *some* is compatible with *all* they should leave the boxes unchanged; otherwise they should remove at least one token.⁸ The two components—the wish utterance and the gesture response—work hand in hand and make the experimenter-participant exchange sensible. Either one of these modifications alone would make the task appear strange. Expressing the wish without the gesture would force the participant to describe what action *would* be taken, which would only reinsert an (even more complicated) verbal aspect. A gesture as a reply to a categorical statement is odd.

Third, instead of using the French quantifier *certain*s we used *quelques*, which we deemed easier to process, especially by younger hearers. This assumption is in line with corpus analyses showing that *certain*s is less frequent than *quelques* in children’s written production (Lambert and Chesnet (2001)), and is less frequent in children’s books (Lété, Sprenger-Charolles, and Colé (2004)). Furthermore, although both *certain*s and *quelques* are high frequency items for adults, the former is lower frequency than the latter (Content, Mousty, and Radeau (1990); see also <http://www.up.univ-mrs.fr/~veronis/data/freqmots-oral.html>). This formal evidence is supported by comments from our interviews

⁸One will notice that the boxes are quantified here, unlike in Experiment 1 where objects in and around the boxes were. Either formulation is technically possible (compare *I would like all tokens to be in a box* vs. *I would like all boxes to contain a token*). We adopted the latter because the one-to-one relation between a token and box is clearer with “boxes containing a token” than it is with “tokens being in a box.” We are assuming that participants’ behavior is unaffected by this modification.

with teachers, who mentioned that younger children understand *quelques* more readily than *certaines*.

From a semantic point of view, analyses point out that, although both *certaines* and *quelques* mean *some* in French and have the same grammatical use in our experimental sentences, *certaines* is more complex because, while *quelques* is a simple existential, *certaines* is a partitive (Corblin (2001), Gondret (1976)). The difference shows clearly, for instance, in the fact that *quelques* can be used in measure phrases, whereas *certaines* cannot: *il y a quelques/*certaines années* (cf. English: ‘some/*some of the years ago’) (for further discussion, see Corblin (2001)).⁹ Assuming that this analysis is on the right track, *certaines* is more complex than *quelques*. Compare “J’ai vu quelques enfants,” which says merely that the speaker saw some children, with “J’ai vu certaines enfants,” which says in addition that the children are part of a contextually salient set. This semantic analysis would suggest that *certaines* carries a meaning that goes beyond *quelques*, indicating that it should be a more difficult quantifier to process.¹⁰

Here, we note two important points. First, the fact that *certaines* is a more complex word should make it more costly for younger participants to process, but this does not imply that they do not master its meaning. Children performed very well on the controls we used in Experiment 1 to ensure that they understood the quantifier properly (as will be seen, similar controls were used in Experiments 2 and 3, yielding the same outcome). Second, were we to find that there is a greater implicature production linked to *quelques* than to *certaines*, this could in no way be accounted for by appealing to a difference in meaning. If anything, French adult native speakers would tend to make more implicatures with *certaines* than with *quelques* precisely because it is a partitive and therefore raises the salience of a larger set.

Turning back to the study, which included 150 children in nursery school, kindergarten, and second grade,¹¹ and 21 adults, we hypothesize that our modifications would encourage even the youngest children to make the “pragmatic” responses, thus suggesting that implicature production is not an ability reserved for adolescents or adults. We also anticipated that, despite our efforts to facilitate pragmatic responses across all ages, we would still find traces of the well-known developmental effect showing that younger children are less likely to produce the implicature than older children and adults.

⁹*Certaines* is a partitive, but it isn’t necessarily a specific indefinite. Besides, as mentioned earlier, it is important to bear in mind that it does not have the same function as the English *certain*.

¹⁰French is by no means the only language to have pairs of indefinite determiners where English only has *some*: Greek, Italian, Dutch (de Jong and Verkuyl (1985), de Hoop (1995)), Turkish (Enç (1991)), and Rumanian, among other languages, display the same pattern.

¹¹We are using the nomenclature of the American system; in French, these would be referred to as *Maternelle moyenne section*, *Maternelle grande section*, and *Cours Élémentaire 1 (CE1)*.

3.1. Method

3.1.1. Participants. There were 66 children from a nursery class (mean age 4;5), 30 who came from a kindergarten class (mean age 5;6), and 54 second graders (mean age 7;5). The children (across the three age groups) came from three different primary schools in the Paris region. There were also 21 adults—undergraduate students at the *Université de Lyon 2*—who took part in the study. All the participants were French native speakers.

3.1.2. Materials. Participants were presented with three scenarios involving five cardboard boxes and plastic tokens. The first, to be referred to as the *Subset* scenario, showed that two of the boxes contained a token. The second, to be referred to as the *All* scenario, showed that all the boxes had a token. The third scenario, to be referred to as the *None* scenario, showed that none of the boxes had a token (see Table 3).

For each of the scenarios the participants heard the same four utterances. The statements listed below were constructed with the quantifiers *all*, *some*, *none*, and *some are not* (which correspond with the four possible quantifiers A, I, E, O in an Aristotelean system, respectively):

- All: I would like all the boxes to contain a token.
Je voudrais que toutes les boîtes contiennent un jeton.
- Some: I would like some boxes to contain a token.
Je voudrais que quelques boîtes contiennent un jeton.
- None: I would like no box to contain a token.
Je voudrais qu'aucune boîte ne contienne de jeton.
- Some are not: I would like some boxes to not contain a token.
Je voudrais que quelques boîtes ne contiennent pas de jetons.

We will refer to these as the *All*, *Some*, *None*, and *Some are not* utterances. The critical affirmative experimental statement that allows us to study implicature is the *Some* utterance, “I would like some boxes to contain a token,” in the *All* scenario. If the participant employs the linguistically encoded meaning of *some* (which is compatible with *all*), the participant ought to change nothing in the scenario. However, if the participant takes up a pragmatic interpretation, the participant ought to remove one or more tokens from the boxes. The other questions and situations were introduced to complete the design and served as controls, mostly to ensure that the participants correctly understood the utterances and were capable of performing the task.

As in Experiment 1, we included a critical negative experimental statement, “I would like some boxes to not contain a token,” stated in the *None* scenario. It has many of the same features as the *Some* utterance in the *All* scenario above, except that it takes place in a negative context (making it *prima facie* a more

TABLE 3
The Four Utterances Presented, the Three Scenarios Presented,
and the Anticipated Responses for Experiment 2

<i>Utterance</i>	Subset Scenario	All Scenario	None Scenario
	□ * □ * □	* * * * *	□ □ □ □
(1) I would like all the boxes to contain a token.	Change (add 3 tokens)	Keep as is	Change (add 5 tokens)
(2) I would like some boxes to contain a token.	Keep as is	Logical: Keep as is Pragmatic: Change (remove at least one)	Change (add at least 1)
(3) I would like no box to contain a token.	Change (remove all)	Change (remove all)	Keep as is
(4) I would like some boxes to not contain a token.	Keep as is	Change (remove at least 1)	*Logical: Keep as is Pragmatic: Change (add at least one)

*This case is technically like the *Some* utterance in the *All* scenario, where *no action* can be taken to be a logical interpretation and an *action* (to add) would be considered pragmatic. However, this utterance involves a negatively quantified statement, making it *prima facie* difficult to process.

difficult task). A minimal interpretation of *some are not* is compatible with a more informative quantifier *none* and thus no action is justifiable. However, if the participant makes a pragmatic interpretation via an implicature (where *Some are not* would imply *Not None*), an action ought to take place (the participant ought to add one or more tokens to the boxes).

3.2. Procedure

We interviewed each participant individually. As in the first experiment, they were first acquainted with a puppet and told that the puppet would make statements concerning the boxes and (this time) the tokens. The procedure differed from Experiment 1 only in that the participants were now asked to make changes to the scenario in order to comply with the puppet's wishes by removing one or more tokens, adding one or more tokens, or by doing nothing. We will consider that a participant fully agreed with the puppet's statement only when she did not change anything at all in the scenario.

Before presenting the experimental sentences we ran a training session in order to determine a participant's minimal competence. The three training questions concerned the tokens and boxes but employed numbers. The three scenarios (A–C below) and their respective training sentences are presented below:

- A) One box contains a token and the puppet says: “I would like two boxes to contain a token” (*Je voudrais que deux boîtes contiennent un jeton*). We expected the participant to add a token to an empty box.
- B) Three boxes contain a token and the puppet says: “I would like three boxes to contain a token” (*Je voudrais que trois boîtes contiennent un jeton*). We expected participants not to change anything.
- C) Again three boxes have a token, but this time the puppet says: “I would like two boxes to contain a token” (*Je voudrais que deux boîtes contiennent un jeton*). We then expected the participant to remove a token.¹²

If a child failed to respond appropriately to the first two of the three statements, the training statements were presented again. The children generally had no difficulty with the training and no participant was eliminated from the study. Once the training session was completed, we presented the participant with the *Subset*, *All*, and *None* scenarios described above. In each scenario the puppet uttered the four statements and always in the same order. Whereas the *Subset* scenario was always presented first, the two others were randomly alternated across participants.

3.3. Results

The percentage of participants who responded logically is shown in Table 4. For the *Some* utterance in the *All* scenario, the logical response is equivalent to the decision to keep the status quo indicating that they took the sentence to be, in effect, “fine the way it is.” This categorization can be viewed strictly or truth-functionally. When it is strict, even harmless modifications qualify as a change. For example, many participants of all ages would add a token after hearing the *Some* utterance in the *Subset* scenario, where two boxes already contained a token. Note that this additional token does not change the truth value of the puppet’s utterance; it only tweaks the notion of *Some*. As a result, Table 4 includes a looser interpretation of “fine the way it is”; however, for the sake of analyses, we consider only proactive responses that result in a change of the truth value as true modifications. In other words, we consider only the stricter of the two categorizations of “change made.”

We first highlight two results concerning the *Some* utterance of central interest, “I would like some boxes to contain a token,” in the *All* scenario. This is followed by underlining three general characteristics of the data and by an analysis of responses to the *Some are not* utterance in the *None* scenario.

¹²An action here would be the result of an implicature because, technically speaking, no action is called for (two boxes *do* contain a token). As shown and argued elsewhere (Papafraçou and Musolino (2003)), however, numbers are considered exceptional and prompt very high rates of implicature production in experimental tasks. Indeed, participants removed a token here without hesitation.

TABLE 4
 Percentage of Participants in Experiment 2 Who Provided Logical Responses
 with Respect to Each of the Three Different Scenarios

	Age	N	Subset Scenario □ * □ * □	All Scenario * * * * *	None Scenario □ □ □ □
<i>All utterances:</i>			(LR: Change)	(LR: Keep as is)	(LR: Change)
I would like all	4	66	95	97	100
the boxes to	5	30	100	97	100
contain a token	7	54	100	100	100
	A	21	100	100	100
<i>Some utterance:</i>			(LR: Keep as is)	(LR: Keep as is)	(LR: Change)
I would like	4		64 (100)*	32	100
some boxes to	5		67 (100)	27	100
contain a token	7		89 (100)	17	100
	A		80 (100)	14	100
<i>No utterance:</i>			(LR: Change)	(LR: Change)	(LR: Keep as is)
I would like no	4		78	69	64
box to contain	5		80	90	97
a token	7		94	98	98
	A		100	100	100
<i>Some are not utterance:</i>			(LR: Keep as is)	(LR: Change)	(LR: Keep as is)
I would like	4		65 (100)*	64	41
some boxes to	5		70 (100)	90	30
not contain a	7		91 (100)	98	41
token	A		95 (100)	95	14

Notes. LR = Logical Response. *The values for the “keep as is” response in the subset condition are interpreted strictly; even harmless changes (e.g., adding a token to a box) are considered violations. However, the values in parentheses show the “keep as is” responses when taken to mean “keep the truth value the same.”

The first observation is that the general modifications made to the task, which were designed to encourage pragmatic responding, had their desired effect. As can be seen in Table 4, only 32% of 4-year-olds responded logically by indicating that the scenario should stay as is. This proportion is significantly lower than one predicted by chance, $\chi^2(1) = 8.7$, $p < .005$. The second is that rates of logical responding decrease monotonically and significantly with age (based on Jonckheere’s test for ordered alternatives), $z = 2.19$, $p = .0143$ (one-tailed).¹³

¹³We also conducted a second analysis to verify that the logical-to-pragmatic age transformation stands with a stricter analysis, i.e., we included *only* those participants (children and adults) who made no harmless changes upon hearing the *Some* utterance in the *Subset* scenario. This analysis also reveals that logical responses decrease significantly with age, $z = 2.12$, $p = .017$ (one-tailed) and in a near-monotonic order (36%, 17%, 11%, and 12% for 4-year-olds, 5-year-olds, 7-year-olds, and adults, respectively).

There are three other general findings that help better characterize the pragmatic-developmental effect reported here. First, one can see that the children, including the 4-year-olds, were generally competent. In all 12 cells, the children responded in a direction that corresponded with adults and, in 11 of the 12, at rates that are significantly above chance levels. The only cell that prompted statistical equivocality among the 4-year-olds is in the *Some are not* utterance in the *None* scenario. This is the complex negative utterance that we are tracking and that can, in principle, prompt an implicature as well. We will consider this utterance later.

Second, all the children—including the 4-year-olds—appear to have a very solid understanding of the affirmative quantifiers *All* and *Some*. We base this on two observations: (i) The children were indistinguishable from the adults when it concerns the *All* utterance, no matter what the scenario, and: (ii) Every participant made a change after hearing the *Some* utterance in the *None* scenario. The utterances with *None* and *Some are not* appear to have posed a greater challenge to the youngest children. For example, with respect to the *None* utterance, 31% of the 4-year-olds did *not* make a change when faced with the *All* scenario and 36% made a change to the *None* scenario.

Third, the responses to the *Some are not* utterance in the *None* scenario (“I would like some boxes not to have a token” uttered when no boxes did) revealed again that it was challenging to our participants. Indeed, the adult participants themselves spontaneously remarked on the relative difficulty of these utterances. We thus summarize the results here while proceeding with caution and, among the children, we focus on the 7-year-olds because they are the only children that demonstrate a general mastery with the negatively quantified utterances (interestingly, 7-year-olds have also been the youngest to show general competence among all control sentences in tasks used in Noveck (2001)). Assuming that the 7-year-olds are competent in processing the *Some are not* utterance in the *None* scenario, it can be argued that they are making fewer changes than adults because they are less able to produce the implicature as the task is made harder, ($\chi^2(1) = 4.5, p < .05$). This would be further evidence showing that children are less apt to produce implicatures than adults. Also, a within-subjects *t*-test showed that the 7-year-olds’ rate of logical performance on the *Some are not* utterance in the *None* scenario (42%) is significantly greater than the rate of logical response on the *Some* utterance in the *All* scenario (17%), ($t(53) = 3.46, p < .005$). This makes sense in that one would expect affirmative utterances to be more facilitative of implicature production than negative ones.

3.4. Discussion

Simplifying the scenario had its desired effect. Overall, the likelihood of finding behavior that reflects implicature production increases dramatically when compared to the task in Experiment 1. Whereas rates of logical responding for

9-year-olds were above 90% in Experiment 1, they drop to around 17% for the oldest children here (the 7-year-olds). The same holds for the adults, who are no longer equivocal and strongly pragmatic. The task conditions here encourage adults too to make the pragmatic enrichment.

Remarkably, when one takes into account the general observations concerning performance on this task—that children are generally competent, especially with the affirmative quantifiers, and that the youngest children prefer making changes—it is impressive that 32% of 4-year-olds choose to remain motionless in the *All* scenario as they hear “I would like some boxes to contain a token.” Thus, even when the task is simplified, we find that a sizable minority of children do not spontaneously produce implicatures. That responses indicating logical interpretations only decrease after that confirms that scalar implicature-making, even in this relatively easy task, increases with age.

Despite yielding percentages of implicature-production among young children in a task at rates that surpass any that have been reported thus far, we remain cautious. A more conservative analysis (see footnote 13), which included only those participants who correctly made no unnecessary changes in response to the *Some* utterance in the *Subset* scenario, revealed that 64% of the 4-year-olds can be properly classified as specifically implicature-producers. Since this percentage is only marginally significant ($p = .06$) when compared to predictions based on chance, we are reticent to conclude that 4-year-olds routinely produce implicatures. On the other hand, using the same conservative criterion, we find that 5-year-olds and 7-year-olds (and adults) are indeed producing implicatures with regularity. Thus, if one makes the task easy enough, one can get younger children to produce more implicatures. Most importantly, a developmental curve remains, regardless of one’s criteria for evaluating implicature-production.

These data also reveal that 7-year-olds are the youngest to resemble adults with positive scalar terms, which are the terms that have been studied most intensively. However, when one takes into consideration the implicature in the more difficult negative context (the *Some are not* utterance in the *None* scenario), one finds that adults are, once again, more proficient at producing them than the 7-year-olds. Taking all the above together, these data show that there is indeed a developmental trend as one goes from 4-year-olds to 7-year-olds to adults.

The main goal of this work has been to show that, while implicature production can be encouraged by facilitating the participants’ task, it is a developmental work in progress. This much has been demonstrated. We would be remiss, however, if we did not set out to at least start determining which factor(s) played a decisive role in facilitating implicature production for children across the two tasks. In the experiment that follows, we return to an investigation of older children (9-year-olds) and adults in order to determine what role, if any, the choice of quantifier term makes.

4. EXPERIMENT 3

As we pointed out earlier, the task in Experiment 2, albeit isomorphic to the one in Experiment 1, was simplified in three ways. We reduced the noise in the context, a verbal truth evaluation was replaced by an action in reply to a request, and the quantifier *certain*s was replaced by *quelques*. With three factors, one can imagine conducting a single experiment that could have up to eight (2^3) separate cells in order to isolate the influence of each factor alone or in combination. This is clearly impractical and leads us to present one experiment here that focuses on word-choice (*certain*s vs. *quelques*), while keeping the remaining facilitative factors from Experiment 2. We ascertain the extent to which implicature production is affected when one group of 9-year-olds receives the word *quelques* and another *certain*s. If the semantic analyses we presented earlier are correct, one ought to find that children, and perhaps adults, are more likely to produce implicatures with the less complex *quelques*.

We investigated 9-year-olds because their linguistic competence is relatively advanced and, at the same time, we know from prior studies that they are still prone to making fewer implicatures than adults. Their performance with these two words would test the hypothesis that the word itself has an effect on implicature production. Moreover, this allows us to ultimately tie together developmentally the data from Experiments 1 and 2. The children in Experiment 1 were 9 years old, while one of the conditions in the next experiment (the one using *quelques*) is essentially an extension of Experiment 2 (which included children who were 4, 5, and 7).

4.1. Method

4.1.1. Participants. Twenty-five children, 9 and 10 years of age (mean age 9;7) and 28 adults participated in this study. The children came from a middle school in a small, middle-class city outside of Lyon, France. The adults were students at the *École Nationale des Arts Décoratifs* in Paris. All participants were native French speakers.

Thirteen children were randomly allocated to the condition using the word *quelques* and 12 to the condition using *certain*s. Similarly, the adults were randomly assigned to one of these two conditions.

4.1.2. Materials, design, and procedure. As in Experiment 2, the participants were asked to fulfill the wishes of a puppet concerning boxes and tokens displayed in front of them. This had to be done either by adding or removing tokens, or by doing nothing. The participants were presented with the same three scenarios (presented in one of two orders) as in Experiment 2 (see Table 3) and the puppet uttered the same four utterances (*All*: “I would like all boxes to contain a token”; *Some*: “I would like some boxes to contain a token”;

None: “I would like no box to contain a token”; and, *Some are not*: “I would like some boxes to not contain a token”) in front of each (*Subset*, *None*, and *All*) scenario. Also, all participants went through the same training session described in Experiment 2 (and with no difficulty).

Unlike in Experiment 2, the participants were assigned to one of two conditions. One condition employed the word *quelques* and the other *certaines*. While the key sentence of the experiment expressed the same utterance as in Experiment 2—i.e., “I would like *some* boxes to contain a token” uttered in the *All* scenario—we were interested in comparing performance when the *Some* utterance contained *quelques* versus *certaines*. As before, sentences with *Some are not* uttered in the *None* scenario were included in both conditions.

4.2. Results

As can be seen in Table 5, the children and adults behaved similarly with respect to the unambiguous control questions and we will not describe them further as we concentrate on performance with the two *Some* utterances in the *All* scenario. Each of the rates indicating logical responding among adults

TABLE 5
Percentage of Participants in Experiment 3 Whose Responses Indicate a Logically Correct Response with Respect to Each of the Three Different Scenarios

	Age	Quantifier	N	Subset Scenario		All Scenario					None Scenario
				* □ * □ □ □	□ □ □ □ □	* □	* □	* □	* □	* □	□ □ □ □ □
<i>All</i> utterance:				(LR: Change)		(LR: Keep as is)					(LR: Change)
I would like all the boxes to contain a token	9–10	<i>Toutes</i>	12	100	100	100	100	100	100	100	100
		<i>Toutes</i>	13	100	100	100	100	100	100	100	100
	Adults	<i>Toutes</i>	14	100	100	100	100	100	100	100	100
		<i>Toutes</i>	14	100	100	100	100	100	100	100	100
<i>Some</i> utterance:				(LR: Keep as is)		(LR: Keep as is)					(LR: Change)
I would like some boxes to contain a token	9–10	<i>Certaines</i>		67 (100)*		42					100
		<i>Quelques</i>		54 (100)		0					100
	Adults	<i>Certaines</i>		79 (100)		21					100
		<i>Quelques</i>		64 (100)		7					100
<i>None</i> utterance:				(LR: Change)		(LR: Change)					(LR: Keep as is)
I would like no box to contain a token	9–10	<i>Aucune</i>		100		100					100
		<i>Aucune</i>		100		100					100
	Adults	<i>Aucune</i>		93		100					100
		<i>Aucune</i>		100		100					100
<i>Some are not</i> utterance:				(LR: Keep as is)		(LR: Change)					(LR: Keep as is)
I would like some boxes to not contain a token	9–10	<i>Certaines</i>	12	83		100					42
		<i>Quelques</i>	13	62		100					8
	Adults	<i>Certaines</i>	14	100		100					21
		<i>Quelques</i>	14	100		100					14

Notes. LR = means Logical Response. *The values in parentheses are “keep as is” responses when responses are interpreted as “essentially, keep as is” because the changes made did not affect truth values.

(7% for *quelques* and 21% for *certaines*) is significantly lower than predictions based on chance, $\chi^2(1) = 11.27$, $p < .01$ for *quelques*, and $\chi^2(1) = 4.4$, $p < .05$ for *certaines*. Thus, adults showed that they are making the implicature (as they were in Experiment 2) and that the quantifier term had no particular effect on performance. However, the children were much more likely to indicate a logical treatment of the quantifier (to refrain from removing tokens) when the existential quantifier is *certaines* than when it is *quelques* (42% vs. 0%, respectively), indicating that *quelques* is linked to higher rates of implicature production among the children (Fisher exact test, $p = .015$). There is thus an interaction between quantifier and age: children are likely to behave like the pragmatic adults, unless the quantifier is *certaines*. When children heard the *Some* utterance expressed as *certaines* in the *All* scenario, the wish-action task prompted a response pattern that is not different from predictions based on chance.

The *Some are not* utterance in the *None* scenario (“I would like some boxes not to have a token” uttered when no boxes did) largely led to the same effect as the one for our main item. As in the analysis of our main item, each of the rates indicating logical responding among adults (14% for *quelques* and 21% for *certaines*) is significantly lower than predictions based on chance, $\chi^2(1) = 7.14$, $p < .01$ for *quelques*, and $\chi^2(1) = 4.4$, $p < .05$ for *certaines*. Thus, adults show that they are making the implicature (as they were in Experiment 2) and that the quantifier had no effect on performance across the two groups. The children, again, are more likely to indicate a logical treatment of the quantifier (to refrain from removing tokens) when the existential quantifier is *certaines* than when it is *quelques* (42% vs. 8%, respectively), indicating that *quelques* is linked to higher rates of implicature production among children here as well (Fisher exact test, $p = .063$). As before, this is indicative of an interaction between quantifier and age: children are likely to behave like the pragmatic adults, unless the quantifier is *certaines*. When children heard the *Some are not* utterance expressed as *Certaines ne contiennent pas* in the *None* scenario, the wish-action task prompted responding that is comparable to predictions based on chance.

4.3. Discussion

The percentages of logical responding indicate that children make more implicatures with *quelques* than with *certaines*. The 9-year-olds’ results with *quelques* (0% “logical”) represent a natural continuation of those reported with younger children in Experiment 2, where 17% of 7-year-old children responded logically. As we described in the introduction to Experiment 2, *certaines* is a more complex quantifier. Although *certaines* is generally understood by children (see performance on the control problems here and in Experiment 1), we argue that it is harder to process, leaving precious few resources to produce the implicature.

An utterance with *quelques*, being easier to process, allows the children to draw the implicature more readily. The adults, on the other hand, have enough resources available to appear unaffected by our choice of word.

It is important to highlight another aspect of the data. Even though this experiment employed *certain*s, participants' performance is more "pragmatic" here than in Experiment 1. In Experiment 1, children and adults gave logical responses 91% and 53% of the time, respectively. In Experiment 3, the percentages for children and adults were 42% and 21%, respectively. This radical reduction indicates that the choice of scalar term is not the only factor facilitating implicature understanding. Although these are comparisons across experiments, these findings indicate that the elimination of distractors or an action-based evaluation (or both together) encourage roughly 50% of a sample of children and roughly 30% of an adult sample to carry out the pragmatic inference.

5. GENERAL DISCUSSION

The foregoing studies were aimed at investigating developmental effects linked to scalar items, with a focus on the role played by complexifying (or distracting) factors. Our main goal was to demonstrate how adventitious linguistic and task-related processing demands may negatively affect the computation of scalar implicatures, especially in young children. In order to achieve this, we manipulated three complexity factors discussed in the Introduction: the presence/absence of distractor items, the nature of the task (truth evaluation vs. manipulation task), and the choice of scalar expression (*quelques* vs. *certain*s). Furthermore, we included in all experiments items testing negative sentences, which had remained uninvestigated in the literature.

Experiment 1 showed that 9-year-olds are more likely to accept the logical meaning of *some* which is compatible with *all*, while adults tend to be equivocal, further demonstrating that this effect is robust (Chierchia et al. (2004), Guasti et al. (2005), Noveck (2001), Papafragou and Musolino (2003)). In the present case, the effect occurs in a standard truth-evaluation task that uses the weak quantifier *certain*s, arbitrary materials, and plenty of distractor items.

Experiment 2 revealed that with a set of minor changes (a nonverbal manipulation task using *quelques* and containing no unnecessary distractors) children as young as 4 years of age can be encouraged to derive implicatures. This result is important for two reasons. First, this pushes down the age for which implicature production has been reported. Second, it shows for the first time that younger children (4- and 5-year-olds) can draw scalar implicatures spontaneously, that is, without previous training (such as that used in Papafragou and Musolino (2003)).

Although this evidence supports the notion that young children can produce implicatures to an even greater extent than has been reported previously, caution

is called for: Despite the dramatic increase of implicatures among younger participants, their performance is still not adult-like. Only 7-year-olds reveal behavior that approaches that of adults among the standard cases (similar results for this age group were obtained by Guasti et al. (2005)) and even among them adult-like implicature performance is less likely when it concerns negative sentences (as we discuss later). None of the extant studies (ours included) supports the unqualified claim that in experimental settings, 4- or 5-year-olds spontaneously produce scalar implicatures the way adults do. *Perhaps* then there is indeed an age at which young children cannot make scalar inferences with ease. (We are *not* claiming that children *cannot*, we are only saying that, thus far, no one can support a claim saying that such young children can produce implicatures like an adult.)

Furthermore, the data show that implicature-production is an ongoing developmental phenomenon (from age 4 to 7 and beyond) that is affected by various factors. While 4-year-olds are responding logically at rates that are not too far removed from chance levels, one can see that pragmatic responses continue to increase with age. This confirms the trend found in previous studies among older children and adults. From the age of 4 onward, children become increasingly adept at making scalar implicatures and the developmental curve is monotonic.

Two further factors we manipulated in our experiments had never been studied before in relation to implicature production—the weak term that plays a role in implicature production and implicature production in negative contexts. First, Experiment 3 shows that the choice of indefinite expression (*quelques* versus *certaines*) is crucial for children. For children—but not for adults—the expression *quelques* is associated with higher rates of implicature production than *certaines*. This shows that lexical complexity in itself plays a role. Children clearly understand the meaning of both expressions, but the added processing cost of *certaines* makes the task harder, thus reducing children’s rate of implicature production. This result is important for the literature. For instance, the fact that the word *certaines* discourages implicature production in children as old as 9, and not in adults, may explain part of the developmental effect reported by Noveck (2001), who employed *certaines* in sentences such as “Some elephants have trunks.” The importance of the exact linguistic expression might also account for small differences in the outcome of experiments using similar designs but run in different languages. Surely, this factor should be taken into account in a field where there is, fortunately, much linguistic diversity: Experiments on scalar terms have been done in at least five languages (Greek, Italian, English, German, and French).

Secondly, although 7-year-olds are able to compute scalar implicatures at rates comparable to adults with a simpler task or a simpler quantifier, they still fail to do so in negative sentences. In Experiment 2, 7-year-olds, who were arguably the youngest to demonstrate adult-like competence with control negative quantifiers, gave more logical responses than adults to the *Some are not* utterance in the *None* scenario. Similarly for Experiment 3, the 9-year-olds were less likely to

produce the implicature to the *Some are not* utterance in the *None* scenario when the quantifier was *certainly*. Thus, there is still an age effect—this time for the difficult negative sentences.

How do our results fit in with the extensive linguistic and philosophical literature on scalar implicatures (Chierchia (2004), Carston (1998), Geurts (1998), Horn (1989), Levinson (2000), Sperber and Wilson (1995), Wilson and Sperber (2004))? These authors generally agree that scalar implicatures are a kind of pragmatic inference and that context plays a crucial role; but otherwise their accounts deviate in important respects. Although strictly speaking, our results may be compatible with most theories currently on the market (Levinson's and Chierchia's being possible exceptions), we feel that the most natural explanation would be along the lines of Relevance Theory (Carston (1998; 2002), Sperber and Wilson (1995), Wilson and Sperber (2004)). On this account, scalar implicatures are ordinary pragmatic inferences drawn by hearers in order to arrive at an interpretation of an utterance that meets their expectations of relevance. A nonenriched interpretation of a scalar term (the one that more closely coincides with the word's lexical meaning) will often suffice as a relevant-enough interpretation of the utterance in which it occurs, and in these cases no inferences are necessary. A scalar inference may be drawn by a hearer in an effort to make an utterance more informative and thus more relevant. How far the hearer goes in construing the utterance's interpretation is governed by considerations of effect and effort; hearers expect the intended interpretation to provide satisfactory effects for minimal effort. While a scalar implicature makes the utterance more informative, it will typically involve extra effort (for more discussion, see Bott and Noveck (2004), Noveck (2004), and Noveck and Sperber (2007)).

This view accounts nicely for our results as well as more generally for the latest findings in the field. The main claim of Relevance Theory regarding the processing of scalar implicatures by younger and older participants would be that implicature production should appear more readily when: i) its effects in contexts are greater, and ii) when the effort required to derive them is not too strenuous (and this should be particularly true for children whose cognitive resources are not at adult levels). This coincides precisely with the overall picture of the state of the art of developmental studies on scalar implicatures. On the one hand, studies such as Papafragou and Musolino (2003, Experiment 2) and Guasti et al. (2005, Experiment 4) demonstrate that children are more likely to draw scalar inferences when the given semantic scale is made highly relevant in context—that is, when the expected effects of the implicature are greater. On the other hand, the work reported here indicates that complexity is an important factor in the understanding of scalar terms—that is to say that effort plays a crucial role.

To summarize, when a task seeking implicatures is simplified along with its linguistic material, we observe increased pragmatic responses for participants of all age groups. We take this result to indicate that increased complexity

negatively affects implicature processing. Thus having sufficient cognitive resources is crucial to making scalar inferences and accounts, at least partly accounts for the striking developmental-pragmatic effect that is generally observed in the literature.

ACKNOWLEDGMENTS

The authors wish to thank Lewis Bott, Bart Geurts, Julien Musolino, Dan Sperber, Michael Tomasello, and two anonymous reviewers with respect to an earlier version of this article as well as to the RD&P team in Lyon for their helpful suggestions on all phases of the work. This work was supported by a grant from the CNRS (*ATIPE*) awarded to the second author (IN) and was conducted while the third author (GP) was affiliated with *Cognition & Usages, Université de Paris-8*.

REFERENCES

- Bloom, P. (2000) *How Children Learn the Meaning of Words*, MIT Press, Cambridge, Massachusetts.
- Bott, L. and I. A. Noveck (2004) "Some Utterances are Underinformative: The Onset and Time Course of Scalar Inferences," *Journal of Memory and Language* 51, 437–457.
- Braine, M. and B. Rumain (1981) "Children's Comprehension of 'or': Evidence for a Sequence of Competencies," *Journal of Experimental Child Psychology* 31, 46–70.
- Breheny, R., N. Katsos, and J. Williams (2006) "Are Generalised Scalar Implicatures Generated by Default? An On-Line Investigation into the Role of Context in Generating Pragmatic Inferences," *Cognition* 100, 434–463.
- Carston, R. (1998) "Informativeness, Relevance and Scalar Implicature," in R. Carston and S. Uschida, eds., *Relevance Theory: Applications and Implications*, John Benjamins, Amsterdam, The Netherlands.
- Carston, R. (2002) *Thoughts and Utterances*, Blackwell, Oxford, UK.
- Chierchia, G. (2004) "Scalar Implicatures, Polarity Phenomena and the Syntax/Pragmatics Interface," in A. Belletti, ed., *Structures and Beyond*, Oxford University Press.
- Chierchia, G., T. Guasti, A. Gualmini, L. Meroni, and S. Crain (2004) "Adults and Children's Semantic and Pragmatic Competence in Interaction," in I. A. Noveck and D. Sperber, eds., *Experimental Pragmatics*, Palgrave Macmillan, Basingstoke.
- Chomsky, C. (1969) *Acquisition of Syntax in Children from 5 to 10*, MIT Press, Cambridge, Massachusetts.
- Content, A., P. Mousty, and M. Radeau (1990) "BRULEX: Une base de données lexicales informatisée pour le français écrit et parlé," *L'Année Psychologique* 90 551–566.
- Corblin, F. (2001) "Où Situer 'certains' dans une Typologie Sémantique des Groupes Nominaux?" in G. Kleiber, B. Laca, and L. Tasmowski, eds., *Typologie des Groupes Nominaux*, Presses Universitaires de Rennes, Rennes.
- Crain, S. and R. Thornton (1998) *Investigations in Universal Grammar. A Guide to Experiments on the Acquisition of Syntax and Semantics*, MIT Press, Cambridge, Massachusetts.
- de Hoop, H. (1995) "On the Characterization of the Weak-Strong Distinction," in E. Bach, E. Jelinek, A. Kratzer, and B. H. Partee, eds., *Quantification in Natural Languages, Vol. 2*, Kluwer, Dordrecht, 421–450.

- de Jong, F. and H. Verkuyl (1985) "Generalized Quantifiers: The Properness of Their Strength," in J. van Benthem and A. ter Meulen, eds., *Generalized Quantifiers in Natural Language*, Foris, Dordrecht, 21–43.
- Déret, D. (1998) "Le syllogisme catégorique: une réexamination du terme 'certains' dans une perspective développementale," *Revue Canadienne de Psychologie Expérimentale* 52, 174–182.
- De Nys, W., and W. Schaeken (in press) "When People are More Logical Under Cognitive Load: Dual Task Impact on Scalar Implicature," *Experimental Psychology*, 54, 128–133.
- Enç, M. (1991) "The Semantics of Specificity," *Linguistic Inquiry* 22, 1–25.
- Geurts, B. (1998) "Scalars," in P. Ludewig and B. Geurts, eds., *Lexikalische Semantik aus kognitiver Sicht*, Gunter Narr, Tübingen, 85–117.
- Geurts, B. (2006) "Take 'Five': The Meaning and Use of a Number Word," in S. Vogeleeer and L. Tasmowski, eds., *Non-Definiteness and Plurality*, Benjamins, Amsterdam/Philadelphia, 311–329.
- Gondret, P. (1976) "'Quelques,' 'plusieurs,' 'certains,' 'divers': Etude sémantique," *Le Français Moderne* 44, 143–152.
- Guasti, M. T., G. Chierchia, S. Crain, F. Foppolo, A. Gualmini, and L. Meroni (2005) "Why Children and Adults Sometimes (But Not Always) Compute Implicatures," *Language and Cognitive Processes* 20, 667–696.
- Horn, L. R. (1989) *A Natural History of Negation*, Chicago University Press, Chicago.
- Horn, L. (1992) "The Said and the Unsaid," in C. Barker and D. P. Dowty, eds., *SALT 2: Proceedings of the Second Conference on Semantic and Linguistic Theory* (Ohio State Working Papers in Linguistics 40), 163–192.
- Lambert, E. and D. Chesnet (2001) "Novlex: une base de données lexicales pour les élèves de primaire," *L'Année Psychologique* 101, 277–288.
- Lété, B., L. Sprenger-Charolles, and P. Colé (2004) "MANULEX: A Grade-Level Lexical Database from French Elementary-School Readers," *Behavior Research Methods, Instruments, & Computers* 36, 156–166.
- Levinson, S. (2000) *Presumptive Meanings*, MIT Press, Cambridge, Massachusetts.
- Musolino, J. (2006) "On the Semantics of the Subset Principle," *Language Learning and Development* 2, 195–218.
- Musolino, J. and J. Lidz (2006) "Why Children Aren't Universally Successful with Quantification," *Linguistics* 44, 817–852.
- Noveck, I. A. (2001) "When Children are More Logical Than Adults: Experimental Investigations of Scalar Implicature," *Cognition* 78, 165–188.
- Noveck, I. A. (2004) "Pragmatic Inferences Related to Logical Terms," in I. A. Noveck and D. Sperber, eds., *Experimental Pragmatics*, Palgrave, Basingstoke, 301–322.
- Noveck, I. A. and A. Posada (2003) "Characterizing the Time Course of an Implicature: An Evoked Potentials Study," *Brain and Language* 85, 203–210.
- Noveck, I. A. and D. Sperber (2007) "The Why and How of Experimental Pragmatics: The Case of 'Scalar Inferences,'" in N. Burton-Roberts, ed., *Pragmatics*, Palgrave, Basingstoke, 184–212.
- Papafragou, A. and J. Musolino (2003) "Scalar Implicatures: Experiments at the Semantics-Pragmatics Interface," *Cognition* 86, 253–282.
- Papafragou, A. and N. Tantalou (2004) "Children's Computation of Implicatures," *Language Acquisition* 12, 71–82.
- Paris, S. G. (1973) "Comprehension of Language Connectives and Propositional Logical Relationships," *Journal of Experimental Child Psychology* 16, 278–291.
- Pinker, S. (1995) "Language Acquisition," in L. R. Gleitman and M. Liberman, eds., *An Invitation to Cognitive Science*, 2nd ed., MIT Press, Cambridge, Massachusetts, 135–182.
- Pouscouλους, N. and I. Noveck (2004) "Implicature et développement," *Psychologie Française* 49, 193–207.
- Prado, J. and I. Noveck (2006) "How Reaction Times Can Elucidate Matching Effects and the Processing of Negation," *Thinking and Reasoning* 12, 309–328.

- Siegal, M. and L. Surian (2004) "Conceptual Development and Conversational Understanding," *Trends in Cognitive Sciences* 8, 534–538.
- Smith, C. L. (1980) "Quantifiers and Question Answering in Young children," *Journal of Experimental Child Psychology* 30, 191–205.
- Sperber, D. and D. Wilson (1995) *Relevance: Communication and Cognition* (2nd ed.), Basil Blackwell, Oxford.
- Sternberg, R. J. (1979) "Developmental Patterns in the Encoding and Combination of Logical Connectives," *Journal of Experimental Child Psychology* 28, 469–498.
- Tomasello, M. (2003) *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press, Boston.
- Wilson, D. and D. Sperber (2004) "Relevance Theory," in G. Ward and L. Horn, eds., *Handbook of Pragmatics*, Blackwell, Cambridge, Massachusetts.

Submitted 29 April 2005

Final version accepted 2 October 2006

Copyright of *Language Acquisition* is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.