Journal of Semantics, 38, 2021, 601–637
https://doi.org/10.1093/jos/ffab016
Advance Access Publication Date: 23 November 2021
Article

OXFORD

Children's Interpretation of Sentences Containing Multiple Scalar Terms

Cory Bill
Leibniz-Zentrum Allgemeine Sprachwissenschaft

Elena Pagliarini
Università degli Studi di Padova

Jacopo Romoli
University of Bergen

Lyn Tieu
Western Sydney University

Stephen Crain
Macquarie University

First version received 18 August 2020; Second version received 16 September 2021; Accepted 13 September 2021

Abstract

Sentences containing the scalar term "some", such as "The pig carried some of his rocks", are usually interpreted as conveying the scalar inference that *the pig did not carry all of his rocks*. Previous research has reported that when interpreting such sentences, children tend to derive fewer of these scalar inferences than adults (Noveck (2001); Papafragou & Musolino (2003); Guasti *et al.* (2005), among others). One approach to explaining these results contends that children have difficulties accessing the alternative sentences involved in the derivation of such scalar inferences. This 'Alternatives-based' approach raises the possibility that children's performance may improve if certain scalar terms are presented together in the same sentence, for example, if a sentence contains both an existential quantifier and a universal quantifier, as in "Every pig carried some of his rocks". Such 'EverySome' sentences have been associated with the inference that *not every pig carried all of his rocks*, as well as the stronger inference that *none of the pigs carried all of his rocks* (see Chemla & Spector (2011), among others). We present two experiments that explore the possibility that children might more readily derive scalar inferences from sentences containing such a combination of scalar terms. Experiment 1 investigates children's interpretation of sentences containing only the quantifier *some* and replicates the previously established finding of fewer inference-based interpretations by children compared to adults. Experiment 2 explores children's interpretation of sentences in which "some" is embedded under "every", and reveals that adults and children access inference-based interpretations of such sentences at similar rates. Moreover, adults and children appear to differ with regards to which of the two possible inferences their interpretations are based on. We discuss the

implications of the experimental results for our understanding of children's acquisition of scalar inferences and for proposals that attempt to capture differences between adults' and children's interpretive preferences.

1 INTRODUCTION

The sentence in (1) is often interpreted as conveying an element of meaning akin to the sentence in (2). The meaning in (2) is not a part of the literal meaning of (1), but rather is an inference licensed by the sentence in (1). Its status as a defeasible inference rather than an entailment is clear from the fact that (2) can be explicitly negated without contradiction, as illustrated in (3). The present paper investigates children's knowledge of inferences like (2).

(1) The pig carried some of his rocks.
(2) The pig didn't carry all of his rocks.
(3) The pig carried some of his rocks... in fact, he carried all of them.

The inference from (1) to (2) is commonly referred to as a 'scalar inference' or a 'scalar implicature'. According to traditional accounts (Horn (1972); Grice (1975; 1978)), conversational implicatures, of which scalar inferences are one type, are derived by reasoning over what the speaker said and what she could have said instead. Simplifying, upon hearing a sentence like (1), the hearer will implicitly reason about why the speaker uttered (1) and not other relevant and more informative sentences such as (4).

(4) The pig carried all of his rocks

The fact that the speaker did not choose to utter the more informative sentence in (4) invites the hearer to infer that the speaker likely believes that the alternative sentence is false, and in turn leads the hearer to believe herself that (4) is false. In other words, the utterance of a sentence like (1) (henceforth referred to as a '*Some*' sentence) tends to cause the hearer to compute the inference in (2) (henceforth referred to as the '*OnlySome*' inference). It is also possible for a single assertion to be associated with more than one scalar inference. For example, it has been suggested in the theoretical literature that sentences like (5) (henceforth referred to as '*EverySome*' sentences) can be associated with *NotEvery* inferences like (6) and *None* inferences like (7) (Fox (2007); Chemla (2009); Chemla & Spector (2011); Chierchia *et al.* (2011)).

(5) Every pig carried some of his rocks.
(6) Not every pig carried all of his rocks.
(7) None of the pigs carried all of his rocks.

A series of studies have investigated adults' interpretations of *EverySome* sentences like (5) (Geurts & Pouscoulous (2009); Clifton & Dube (2010); Chemla & Spector (2011); Potts *et al.* (2016); Franke *et al.* (2016); Gotzner & Romoli (2018)). While early studies, such as Geurts & Pouscoulous (2009), found clear evidence of adults deriving *NotEvery* inferences from such sentences, results relating to the derivation of *None* inferences were mixed.[1] Later

1 Specifically, while Geurts & Pouscoulous (2009) report that adults in their inferential task experiment derived *None* inferences around 50% of the time, the authors argued that certain aspects of the task may have artificially inflated the rate of such interpretations. This led them to conduct a second experiment using a verification task that would eliminate the relevant issues. Geurts & Pouscoulous (2009) ultimately conclude that their study does not provide evidence that adults derive *None* inferences from *EverySome* sentences.

studies, however, have consistently reported that adults do derive *None* inferences when interpreting *EverySome* sentences (Clifton & Dube (2010); Chemla & Spector (2011); Potts *et al.* (2016); Franke *et al.* (2016); Gotzner & Romoli (2018); Benz & Gotzner (2020)). To our knowledge, no previous work has examined children's interpretations of *EverySome* sentences.[2]

In contrast, a great deal of research has investigated children's derivation of scalar inferences from a variety of other sentences, including *Some* sentences. Initially, such work consistently found that children were less likely than adults to derive scalar inferences from such sentences (Noveck (2001); Chierchia *et al.* (2001); Papafragou & Musolino (2003); Guasti *et al.* (2005); Foppolo *et al.* (2012)). However, this early work largely investigated classic cases of scalar inferences arising from the use of lexical items like "some", "or", and "might". More recently, as research has expanded to include a wider range of implicature-type meanings, it has been reported that there are a handful of scalar inferences that children are indeed capable of computing (Papafragou & Musolino (2003); Barner *et al.* (2011); Stiller *et al.* (2015); Tieu *et al.* (2016); Singh *et al.* (2016); Hochstein *et al.* (2016); Pagliarini *et al.* (2018); Cremers *et al.* (2018)). The findings of these more recent studies would suggest that children's ability to derive scalar inferences is perhaps more variable than was suggested in earlier work. Theories that attempt to explain children's behavior on scalar implicatures must be able to explain this variability.

One recent approach, which we will refer to as the 'Alternatives-based approach', has proven relatively effective at capturing children's variable behavior on scalar inferences (Chierchia *et al.* (2001); Gualmini *et al.* (2001); Barner & Bachrach (2010); Barner *et al.* (2011); Tieu *et al.* (2016); Singh *et al.* (2016); Skordos & Papafragou (2016)). This approach proposes that children's difficulties with scalar inferences are a result of certain limitations affecting their ability to access the relevant alternative sentences, from which scalar inferences are derived. This proposal captures children's previously reported behavior because many of the cases where their performance improved involved situations in which they were plausibly assisted in accessing these alternatives. For reasons we will return to, it can be argued that *EverySome* sentences provide precisely this kind of assistance to children. And this raises the interesting possibility that children may more readily access inference-based interpretations of *EverySome* sentences, compared to their 'simpler' *Some* counterparts.

Through two experiments, we will explore this latter possibility by investigating children's interpretations of *Some* and *EverySome* sentences. In Experiment 1, we find that children derive fewer inference-based interpretations of *Some* sentences than adults, consistent with previous studies. In Experiment 2, we find that children derive inference-based interpretations of *EverySome* sentences at adult-like rates. However, when we analyse

2 There has nevertheless been a considerable amount of work investigating children's interpretations of universally quantified sentences, such as *Every pig carried a rock* (see Philip (2011) for a review). While there is clearly a degree of similarity between our target *EverySome* sentences and these sentences, they are also different in important respects, especially when it comes to the scalar inferences with which they are associated: while *EverySome* sentences are associated with the noted *NotEvery* and *None* scalar inferences, no such inferences are expected to be derived from universally quantified sentences without "some". For this reason, we will not present this research in any detail here, although we will return to this work briefly in exploring a possible alternative explanation for our results (see Section 4.4).

the results of the two experiments together, we observe only a main effect of group, with adults accessing more inference-based interpretations than children. Thus taken together, the experiments present no evidence that *EverySome* sentences have a facilitatory effect on children's access to the relevant implicatures. The overall results are therefore not in line with the Alternatives-based approach's hypothesis that children should more readily derive inference-based interpretations of *EverySome* sentences than of *Some* sentences.

A second finding of our study is that, while adults and children give similar rates of inference-based responses to *EverySome* sentences, these responses are based on distinct inferences: while adults derive a *NotEvery* inference (e.g., (6)), children's responses tend to be based on the derivation of a *None* inference (e.g., (7)). We explore how a developmental preference for stronger interpretations (Crain *et al.* (1994)) might explain these results.

The rest of the paper is structured as follows. First, we review the previous developmental literature on scalar inference derivation, including some recent proposals that attempt to capture children's behavior. Next, we outline how the Alternatives-based approach raises the possibility that children will more easily access scalar inferences from *EverySome* sentences than *Some* sentences. We then report two experiments, one investigating children's interpretations of *Some* sentences and the second investigating *EverySome* sentences. We conclude by discussing the results of the two experiments and their implications for our understanding of children's acquisition of scalar inferences.

1.1 *Children and scalar inferences*

Over the past few decades, a great deal of research has focused on children's ability to compute scalar inferences. Starting in the early 2000s, a number of studies consistently found that children derived fewer scalar inferences than adults (Noveck (2001); Papafragou & Musolino (2003); Huang & Snedeker (2009); Foppolo *et al.* (2012), among others). For example, Noveck (2001) used a 'reasoning scenario' to investigate children's behavior with a number of different scalar inferences. For one of these inferences, participants were presented with sentences like (8) and asked whether they agreed with them or not. Based on world knowledge, the associated *OnlySome* inference in (9) is false. Therefore, if a participant derived the inference in (9), they were expected to reject the test sentence in (8).

(8) Some giraffes have long necks.
(9) Not all giraffes have long necks.

Noveck (2001) ran the study with 31 8-year-olds, 30 10-year-olds, and 15 adults. All participants were native French speakers and the test sentences were presented in French. Noveck's adult group derived *OnlySome* inferences like (9) 69% of the time, while the two child groups derived them 11% (8 y/o) and 15% (10 y/o) of the time. This result, in conjunction with similar results from two other experiments presented in the same paper, led Noveck to conclude that children are less likely than adults to derive scalar inferences. A number of subsequent experimental studies over the following decade reported similar results (Chierchia *et al.* (2001); Gualmini *et al.* (2001); Papafragou & Musolino (2003); Guasti *et al.* (2005); Foppolo *et al.* (2012)).

While the studies conducted by Noveck (2001) and others provided convincing evidence that children struggled to derive the target inferences, the studies largely focused on a small group of scalar inferences (primarily those associated with the scales: "some"/"all", "or"/"and", and "might"/"must"). As investigations into children's derivation of scalar inferences continued, and particularly as studies started to include different methods and a wider range of scalar inferences, a different behavioral pattern began to emerge. Namely, it was found that there were a handful of inferences that children seemed to derive readily (Papafragou & Musolino (2003); Barner & Bachrach (2010); Tieu *et al.* (2016); Hochstein *et al.* (2016); Pagliarini *et al.* (2018)). For example, a study by Tieu *et al.* (2016) investigated children's understanding of sentences like (10), which triggers the so-called free choice inference in (11); importantly, free choice inferences have received a scalar inference analysis in the theoretical literature (Kratzer & Shimoyama (2002); Alonso-Ovalle (2005); Fox (2007); Klinedinst (2007); Chemla (2009)).

(10) Kung Fu Panda may push the green car or the orange car.

(11) Kung Fu Panda may push the green car and Kung Fu Panda may push the orange car.

Tieu et al. conducted a Truth Value Judgment Task (Crain & Thornton (1998)) experiment with 22 Mandarin-speaking 3- to 4-year-old children. The authors presented participants with the Mandarin versions of target sentences like (10) as descriptions of contexts in which the relevant free choice inference was false (e.g., where Kung Fu Panda was only allowed to push the orange car). Given this context, a rejection of the test sentence by a participant was interpreted as evidence that the participant had derived a free choice inference. Tieu et al. found that while children derived the standard 'not both' exclusivity implicature of disjunction at a typically low rate (18%), they derived free choice inferences like (11) at a much higher rate (91%).

In a similar vein, Hochstein *et al.* (2016) investigated children's interpretations of sentences like (12), which typically trigger ignorance inferences like that in (13). They found that 5-year-old children derived such ignorance inferences at a much higher rate (≈76%) compared to exclusivity inferences (≈30%).

(12) The bear took a cup or a plate.

(13) The speaker is ignorant as to whether the bear took a cup and as to whether the bear took a plate.

Stiller *et al.* (2015) investigated the derivation of so-called ad-hoc implicatures from sentences like (14) with a group of 2- to-4-year old children. Children were directed to identify which of three faces the test sentence was describing. The characteristics of the three faces corresponded to the following ad-hoc scale: <face with no glasses and no hat, face with glasses but no hat, face with glasses and hat>. It was expected that if children derived the target ad-hoc inference in (15), they would select the face with glasses but no hat. Stiller et al. reported that 3- and 4-year-old children derived such ad-hoc inferences at a rate of approximately 75%.[3]

(14) My friend has glasses.

(15) My friend does not have a hat.

A study by Pagliarini *et al.* (2018) investigated whether children would access an inference-based interpretation of sentences like (16), that is, whether their interpretations would include the associated *distributive* inference in (17) or *conjunctive* inference in (18).[4]

3 In contrast, the 2-year-olds in this study appeared to respond randomly.

4 See Singh *et al.* (2016) and Bowler (2014) for proposals regarding the derivation of conjunctive inferences from such sentences when a language lacks a conjunctive alternative. As we will discuss, children (across languages) are proposed to lack access to conjunction as a scalar alternative to disjunction, leading them to also derive conjunctive inferences (Singh *et al.* (2016)).

(16) Every elephant caught a big butterfly or a small butterfly.
(17) At least one elephant caught a big butterfly and at least one elephant caught a small butterfly.
(18) Every elephant caught a big butterfly and a small butterfly.

Pagliarini et al. found that children accessed inference-based interpretations of the target sentences at the same rate as adults did (approximately 55% of the time).

A series of studies have also reported evidence that children readily access an *exactly-n* interpretation of numerals (i.e. "one" as meaning *exactly one*) (Papafragou & Musolino (2003); Barner & Bachrach (2010); Huang *et al.* (2013)), a meaning that has also been proposed in the theoretical literature to correspond to a scalar inference (Sauerland *et al.* (2005); Spector (2007)).

In another example of successful inference derivation by children, Katsos & Bishop (2011) employed a ternary judgment task to investigate children's derivation of *OnlySome* inferences. Specifically, a cartoon character (Mr. Caveman) presented participants with *Some* sentences as descriptions of contexts that were inconsistent with the associated *OnlySome* inference. Participants were instructed to judge Mr. Caveman's description and reward him accordingly using a 3-point scale comprised of different-sized strawberries. Katsos and Bishop found that children gave responses associated with having derived the relevant scalar inference (i.e. they selected the intermediate, medium-sized strawberry) at the same rate as adults.

Finally, a study by Barner *et al.* (2011) found evidence of 4-year-old children accessing *exhaustive* interpretations of sentences containing "only"; importantly, the relevant exhaustive interpretation of such sentences is thought to be derived through a similar process as scalar inferences.

In sum, while earlier studies found that children struggled to derive scalar inferences, a growing number of more recent studies have found that children can in fact readily derive certain scalar inferences.

1.2 *Explaining children's variable success on scalar inferences*

A number of explanations have been proposed to account for children's difficulty with the derivation of scalar inferences. Some early proposals attributed children's performance to general processing difficulties (Chierchia *et al.* (2001); Reinhart (2006)). More recently, some accounts have suggested that children's difficulties with scalar inferences are due to children being more tolerant than adults are of pragmatic infelicity (Katsos & Bishop (2011)).

While these explanations can account for the results of studies that reported low rates of scalar inference derivation by children, they are less able to handle those cases where children readily derive scalar inferences. That is, the limitations attributed to children by these approaches would be expected to affect children's derivation of scalar inferences uniformly. As a result, these approaches fail to explain why, for example, Tieu *et al.* (2016) found children deriving free choice inferences significantly more than exclusivity inferences, despite the experimental contexts being equivalent in the relevant respects.

One way to maintain these accounts in the face of the apparent variability in children's performance would be to adopt an alternative, non-scalar inference analysis for the inferences that children readily access. This is how Papafragou & Musolino (2003) proposed to explain the high rate at which children accessed *exactly-n* interpretations of numerals. Consistent with this approach, recent research has revealed some additional differences between standard scalar implicatures on the one hand and *exactly-n* and free choice inferences on the other. Specifically, differences have been observed with regard to the ease with which these inferences are derived from an embedded position, as well as the speed with which they are derived (see Chemla & Singh (2014) for a review of the relevant literature).

While positing such distinctions between inferences might be empirically justifiable in some cases, it is not cost-free. Adopting such a strategy means abandoning the gains in parsimony achieved by explaining the derivation of so many, seemingly disparate, inferences through a single interpretive process. Moreover, one would have to posit some other mechanism to explain why these meanings behave like scalar inferences in certain contexts, for example, why they disappear in downward-entailing environments such as the scope of negation.

1.2.1 *The Alternatives-based approach* An approach that is able to explain children's variable behavior while retaining a unified scalar implicature analysis of the relevant phenomena is the Alternatives-based approach, which posits that children have certain limitations that affect their ability to access the alternative sentences through which scalar inferences are derived (Barner & Bachrach (2010); Barner *et al.* (2011); Tieu *et al.* (2016); Singh *et al.* (2016); Skordos & Papafragou (2016)). The Alternatives-based approach can be broken down into at least two distinct proposals, based on the specific limitation attributed to children.

One variant of the Alternatives-based approach contends that children's behavior is the result of limitations in their knowledge of the abstract lexical scales (Horn (1972)) involved in the generation of the alternative sentences from which scalar inferences are derived (Barner & Bachrach (2010); Barner *et al.* (2011)). Another variant attributes the behavior to limitations in children's ability to recognise which alternatives are relevant in a given context (Skordos & Papafragou (2016)).[5]

Regardless of the specific limitation proposed, the Alternatives-based approach leads us to expect that children will experience difficulties deriving a number of scalar inferences. However, it also suggests that if children's access to the relevant alternatives is facilitated in certain ways, then they will more readily derive the associated scalar inference.

The developmental results outlined in Section 1.1 appear to be largely consistent with the predictions of the Alternatives-based approach. The studies that found evidence of children successfully deriving scalar inferences generally involved sentences and/or contexts where children were arguably given some form of support to access the relevant alternatives. For example, sentences like (19), from which free choice inferences like (20) are derived, plausibly facilitate children's access to the required alternatives by presenting these alternatives as sub-constituents of the original sentence. Specifically, the free choice inference in (20) is derived from the alternative sentences in (21) and (22), which can both be formulated by deleting elements of the original sentence in (19). Indeed, as we have seen, children derive free choice inferences more readily than exclusivity implicatures (Tieu *et al.* (2016)), for which no such facilitation is provided by the original sentence. Ignorance inferences and *distributive* inferences, both of which children reportedly also derive

5 Note that these proposals are not mutually exclusive. It is entirely possible that children's computation of scalar inferences is influenced by both kinds of limitations.

(Hochstein *et al.* (2016); Pagliarini *et al.* (2018)), can similarly be argued to be derived from sentences that facilitate access to alternatives.

(19) Kung Fu Panda may push the green car or the orange car.
(20) Kung Fu Panda may push the green car and Kung Fu Panda may push the orange car.
(21) Kung Fu Panda may push the green car.
(22) Kung Fu Panda may push the orange car.

Another way to facilitate children's access to the relevant alternatives seems to involve making these alternatives highly salient in the context. This was shown in a study by Skordos & Papafragou (2016), in which children were found to more readily derive *OnlySome* inferences from *Some* sentences when the relevance of the "all" alternative was made highly salient in the context. Similarly, children's successful derivation of ad-hoc inferences in Stiller *et al.* (2015) is consistent with this suggestion, given that the alternatives in the Stiller et al. study were clearly visible and contrasted side by side in the pictured contexts.[6]

These results from the child language acquisition literature would appear to be corroborated by work on adult sentence processing. A number of studies have found that scalar inferences that are derived through lexical alternatives (e.g., exclusivity implicatures) are associated with a processing cost when compared to their corresponding literal meanings. In contrast, inferences derived through sub-constituent alternatives (e.g., free choice inferences) are not associated with such a cost (Chemla & Bott (2014); van Tiel & Schaeken (2017)). In short, the pattern seems to be that scalar inferences that involve lexical substitution are associated with both a lower rate of derivation by children and a processing cost for adults.

One finding that, on the face of it, goes against the predictions of the Alternatives-based approach is children's ready derivation of *exactly-n* inferences of numerals. One fairly straightforward way to capture this result, as proposed by Barner & Bachrach (2010), is to appeal to the manner in which children learn numerals in the first place. In particular, children tend to learn numbers from an early age through exposure to numbers as members of an ordered list. Given this, it is perhaps unsurprising that children's ability to generate alternatives based on the numeral scale would surpass that of any other scale.

A finding that is less easily captured under the Alternatives-based approach is that reported in Katsos & Bishop (2011): children's performance on *OnlySome* inferences improved when they were presented with an intermediate response option, rather than binary yes/no options. It's not clear how access to alternatives could explain children's performance in the task. That is, it is not obvious how offering an intermediate response option would affect children's access to the required alternatives. One possibility, as Katsos & Bishop (2011) themselves note, is that children's adult-like responses in this experiment were motivated by a sensitivity to the under-informativity of the relevant sentences, rather

6 In fact, the idea that children might have difficulties generating certain alternatives aligns remarkably well with work in the theoretical literature by Katzir (2007), Fox & Katzir (2011) and Breheny *et al.* (2017), which propose distinct sources of alternatives for implicatures. This work identifies and distinguishes between different sources of alternative sentence generation. Specifically, some alternative sentences are generated by accessing the lexicon, whereas others are generated from sub-constituents of the assertion or from the context. Framed in this way, the Alternatives-based approach contends that children only experience difficulties generating alternatives from one of these sources, namely the lexicon.

than from having derived a genuine scalar inference. A version of this explanation that would be consistent with the Alternatives-based approach would be that the children in this study were aware of the existence of an alternative expression that would be a better way to describe the context, without being able to explicitly identify exactly what that alternative was.[7] This could explain why they might accept the relevant sentences in a binary judgment task, but choose the intermediate response when given the option. In any case, while the Alternatives-based approach straightforwardly accounts for much of the noted variation in children's behavior with scalar inferences, there are evidently some findings that present more of a challenge.

In sum, the Alternatives-based approach can successfully capture both children's reported difficulties with scalar inference computation, as well as many of the cases in which they appear to succeed. Relevant for our purposes, the Alternatives-based approach makes a prediction regarding the influence of the linguistic or experimental context on children's ability to derive scalar inferences. In particular, if children's access to the relevant alternatives is facilitated by certain properties of the sentential or experimental context, then they will more readily derive the associated scalar inference. This prediction raises some interesting possibilities regarding the relative ease with which children access inference-based interpretations of *Some* and *EverySome* sentences, which we turn to next.

1.3 *Some and EverySome sentences*

To summarize thus far, previous studies such as Noveck (2001) have revealed that, when presented with *Some* sentences like (23), children tend not to derive the associated *OnlySome* inference in (24). According to the Alternatives-based approach, this is because children have limitations affecting their ability to generate the relevant alternative sentence in (25).

(23) The pig carried some of his rocks.
(24) The pig didn't carry all of his rocks.
(25) The pig carried all of his rocks.

EverySome sentences like (26) have been associated with two scalar inferences — the *NotEvery* inference in (27) and the *None* inference in (28).

(26) Every pig carried some of his rocks.
(27) Not every pig carried all of his rocks.
(28) None of the pigs carried all of his rocks.

A series of studies have investigated adults' interpretations of *EverySome* sentences (Geurts & Pouscoulous (2009); Clifton & Dube (2010); Chemla & Spector (2011); Potts *et al.* (2016); Franke *et al.* (2016); Benz & Gotzner (2020)). This work has reported that adults access interpretations that include each of these inferences to some extent.

Given that children have difficulty deriving *OnlySome* inferences from sentences like (23), it is not surprising that no previous work has investigated children's ability to derive the inferences associated with *EverySome* sentences. However, under certain assumptions, the Alternatives-based approach raises the possibility that the inferences in (27) and (28)

7 This could be similar in nature to the so-called 'tip-of-the-tongue' phenomenon, whereby speakers are aware of but are not able to fully retrieve a given word or expression Brown & McNeill (1966).

may in fact be easier for children to derive than the *OnlySome* inference. Before we can explore this possibility in more depth, we will first present the details of the process through which these inferences are thought to be derived in the first place.

1.3.1 *Deriving OnlySome, NotEvery, and None inferences* There is ongoing debate regarding the exact mechanism underlying the derivation of scalar inferences. The traditional Gricean account presented at the beginning of this paper is but one among several proposals. Our investigation does not rely on assuming any particular account, but for ease of exposition, we will adopt the so-called 'Grammatical account' of the inferences under (Chierchia (2006); Fox (2007); Chierchia *et al.* (2011)). According to this account, scalar inferences are derived as a result of the application of a covert exhaustivity operator 'exh', which is akin to a silent "only". This operator exh takes a proposition, affirms it, and negates certain 'excludable' alternatives to it while avoiding contradiction (for example, alternatives that are not entailed by the assertion).

Let us consider the sentence in (29), schematised as 'SOME', which gives rise to the *OnlySome* inference in (30).

(29) The pig carried some of his rocks.
(30) The pig didn't carry all of his rocks.

To derive (30), exh is applied to SOME as in (31). The set of excludable alternatives to SOME include (32), schematised as 'ALL' in (33).[8]

(31) exh[SOME]
(32) The pig carried all of his rocks.
(33) Alt $= \{\text{SOME}, \text{ALL}\}$

The result of the exhaustification process is that the ALL alternative is negated, yielding the meaning in (30), which includes the *OnlySome* inference ¬ALL, i.e. (36).

(34) $\text{SOME} \wedge \neg\text{ALL}$

As we have seen, sentences like (35), schematised as 'EVERY(SOME)', have been associated with *NotEvery* inferences like (36).

(35) Every pig carried some of his rocks.
(36) Not every pig carried all of his rocks.

The Grammatical approach can account for this inference using the same mechanism as above, with exh applied to the whole sentence, as shown in (37). In this case, the relevant alternatives include (38), schematised as 'EVERY(ALL)' in (39):

(37) exh[EVERY(SOME)]
(38) Every pig carried all of his rocks.
(39) Alt $= \left\{\begin{array}{l}\text{EVERY(SOME)},\\ \text{EVERY(ALL)}\end{array}\right\}$

8 A controversial topic in the literature is how exactly one determines the alternatives that exh quantifies over. Addressing this question would take us beyond the scope of this paper. We will simply assume the alternatives indicated above; see Breheny *et al.* (2018) and references therein for discussion.

It is possible to negate the 'EVERY(ALL)' alternative without generating a contradiction and so the resulting meaning in (40) includes the *NotEvery* inference '¬EVERY(ALL)', i.e. (36).

(40) EVERY(SOME) ∧ ¬EVERY(ALL)

As for the *None* inference in (41), within the Grammatical account, there are two main ways this inference can be derived from the *EverySome* sentence in (35).

(41) None of the pigs carried all of his rocks.

The first way assumes a larger set of alternatives over which exh quantifies. Specifically, the set also includes the alternatives created by replacing the quantifier *every* with its scale-mate *some*, generating the set of alternatives in (42) (Chemla & Spector (2011); Magri (2011); Romoli (2012); Gotzner & Romoli (2018)).

(42) $\text{Alt} = \left\{ \begin{array}{l} \text{EVERY(SOME)}, \\ \text{EVERY(ALL)}, \\ \text{SOME(SOME)}, \\ \text{SOME(ALL)} \end{array} \right\}$

With this set of alternatives, if EXH is applied at the whole-sentence level, as in (43), the *None* inference (i.e. (41)) is included in the final meaning in (44).

(43) exh[EVERY(SOME)]

(44) EVERY(SOME) ∧ ¬SOME(ALL)

Crucially, it is possible to negate the alternative SOME(ALL) in (42) without contradiction, resulting in the meaning ¬SOME(ALL), which is equivalent to the *None* inference in (41).

The second way to derive (41) appeals to the fact that exh, by virtue of being a grammatical operator, is able to appear in an embedded position within a sentence. For example, *EverySome* sentences like (45) have two main sites at which exh can appear: at the whole-sentence level, as in (46) and at the embedded level under the universal quantifier, as in (47).

(45) Every pig carried some of his rocks.

(46) exh[Every pig carried some of his rocks.]

(47) Every pig λx[exh[x carried some of his rocks.]]

Without assuming additional alternatives, applying exh at the whole-sentence level results in the derivation of the *NotEvery* inference in (36), as just outlined. Embedding exh under the universal quantifier, as in (47), on the other hand, results in the derivation of the *None* inference in (41).

To illustrate, consider the alternatives over which the embedded exh operates:

(48) $\text{Alt} = \left\{ \begin{array}{l} x \text{ SOME}, \\ x \text{ ALL} \end{array} \right\}$

After abstracting over the variable, the predicate that combines with the topmost quantifier *every* is that in (49), resulting in the interpretation entailing the *None* inference in (50), i.e. *every pig carried some and not all of his rocks* entails that *no pig carried all of his rocks*.

(49) $\lambda x[x \text{ SOME} \wedge \neg x \text{ ALL}]$

(50) $\text{EVERY}(\lambda x[x \text{ SOME} \wedge \neg x \text{ ALL}])$

In sum, there are two ways to derive the *None* inference within the Grammatical account, one involving an embedded exh and the other involving a larger set of alternatives.[9] As mentioned earlier, while the inferences derived from *Some* and *EverySome* sentences are thought to be derived through the same mechanism of exhaustification, the alternatives that are involved vary, and when considered in the light of the Alternatives-based approach, this variation creates the possibility that children might more readily derive inferences from *EverySome* sentences than from *Some* sentences. We discuss this next.

1.3.2 *Children's inference-based interpretations of Some vs. EverySome sentences* As just outlined, the alternative sentences involved in the derivation of *OnlySome*, *NotEvery*, and *None* inferences are all generated by replacing the relevant scalar terms in the original sentence with the existential quantifier "'some" or the universal quantifier "all". In the case of *OnlySome* inferences, the alternative requires the retrieval of the universal quantifier "all" from the lexicon. There is no sense in which the associated *Some* sentence facilitates the generation of the relevant alternative. In contrast, for the *NotEvery* and *None* inferences, the associated *EverySome* sentence can be argued to facilitate the generation of the relevant alternatives. Specifically, an *EverySome* sentence like (51) explicitly presents the scalar terms involved in the generation of the relevant alternative sentences.

(51) Every pig carried some of his rocks.

That is, *EverySome* sentences present an existential quantifier and a universal quantifier explicitly, which could conceivably facilitate the generation of the relevant alternative sentences (e.g., (52)), as the scalar term required to generate the relevant alternative is also a universal quantifier (i.e. "all"). Therefore, in some sense a key part of the relevant alternative is explicitly presented in the original sentence.[10]

(52) Every pig carried all of his rocks.

According to the Alternatives-based approach, when children's access to the relevant alternatives is facilitated through certain aspects of the sentence or experimental context, they will more readily derive the associated scalar inferences. If this presentation of the relevant alternatives within the original sentence counts as providing such facilitation, then we might expect children to derive more scalar inferences from *EverySome* sentences than

9 Both of these options touch upon a variety of controversial issues regarding embedded vs. matrix computation of inferences and their associated set of alternatives, a full discussion of which would take us beyond the scope of this paper.

10 Note that we are assuming that at the appropriate level of representation for constructing scalar alternatives, *every* and *all* are interchangeable. While this is not the traditional way of thinking about alternatives (Horn (1972)), recent work by Buccola *et al.* (2018) proposes and provides experimental evidence for the idea that scalar alternatives are conceptual rather than lexical in nature. Adopting such a perspective, *every* and *all* could be interchangeable in the proposed manner.

from *Some* sentences.[11] Note that the Alternatives-based approach does not give rise to specific predictions regarding which of the two associated scalar inferences (i.e. (27) and (28)) children will prefer. Rather, this approach merely raises the possibility that children's derivation of inference-based interpretations will be facilitated by the properties of *EverySome* sentences that we have highlighted.

Note also that the version of the Alternatives-based approach presented in Skordos & Papafragou (2016) would not necessarily predict the highlighted differences between *Some* and *EverySome* sentences to influence children's derivation of the associated inferences. Skordos & Papafragou (2016) predict that children will perform better when the context makes the alternative sentences more 'relevant'. It is not clear that *EverySome* sentences differ from *Some* sentences in this regard. The hypothesis that children may generate more inferences from *EverySome* sentences than from *Some* sentences is therefore more aligned with versions of the Alternatives-based approach that attribute children's behavior to difficulties generating alternatives involving lexical replacement (i.e. Barner & Bachrach (2010); Tieu *et al.* (2020)).

We turn now to our experiments investigating the inferences that adults and children draw from *Some* sentences (Experiment 1) compared to *EverySome* sentences (Experiment 2).

2 EXPERIMENT 1

The aim of Experiment 1 was to investigate adults' and children's derivation of scalar inferences from *Some* sentences.

2.1 *Method*

2.1.1 *Participants* Twenty-four monolingual English-speaking adults and 20 English-speaking children (4;00-5;11, M = 5;04) participated in the experiment. The adults were recruited through Amazon Mechanical Turk and were paid 4USD for completing the experiment. The child participants were recruited from daycare centres in and around Macquarie University, Sydney, as well as from a participant database.[12] Informed consent was obtained from adult participants and from a parent/guardian of the child participants. The parents/guardians of the child participants tested in the lab were compensated 20 AUD for travel expenses.

11 As noted by an anonymous reviewer, there are a number of differences between *Some* and *EverySome* sentences (e.g., the number of quantifiers), meaning that the comparison does not involve a minimal pair. We agree with the reviewer on this point; however, other than the presence of the stronger alternative, the other respects in which these sentences differ would seem most plausibly to either have no effect on scalar inference derivation, or would lead to fewer scalar inferences being derived from *EverySome* sentences, due to its greater complexity. We therefore view our investigation as a useful starting point in terms of exploring the possibility that this particular factor alone (i.e. the presence of the stronger alternative quantifier) might be able to facilitate scalar inference derivation.

12 The participant database was created and managed by the Australian Research Council Centre of Excellence in Cognition and its Disorders.

2.1.2 *Procedure* The experiment took the form of a Truth Value Judgment Task (Crain & Thornton (1998)). This task involves two experimenters. One experimenter acts out a series of stories, and the other operates a puppet who watches the stories along with the participant. After each story, the first experimenter asks the puppet to describe some aspect of the story they have just been told and the puppet responds with a target sentence. The first experimenter then asks the participant whether what the puppet said was right or wrong. The participant responds with a *yes*- or *no*-judgment. If the participant provides a *no*-judgment, then the experimenter typically asks the participant to provide a justification (i.e. "Why do you think <Puppet's name> is wrong?"/"What really happened?"). Justifications are less typically requested for *yes*-responses, as such requests can be infelicitous and may potentially confuse child participants (Crain & Thornton (1998)).

Adults saw the same materials; rather than being tested live by an experimenter, however, the adult participants completed a web-based version of the experiment, which was created and hosted on the Qualtrics platform. The stories and puppet's sentences were pre-recorded and embedded within a Qualtrics survey. After viewing the puppet's sentence, adults were asked *Was Piggy [the puppet's name] right?*, which they responded to by clicking on either *yes* or *no* response buttons. Next they were asked *Why was Piggy right/wrong?* and had to type in a justification for their judgment. Note that, unlike children, adults were asked to provide justifications for both *yes*- and *no*-judgments. This was done to reduce the possibility that adult participants would be biased towards '*yes*-judgments' in order to avoid having to type out a justification.[13]

Children were tested individually, either in the lab or in a quiet room at their daycare. The experimental session lasted approximately 20 minutes.

2.1.3 *Materials* There were four items in the test condition. Each of these items consisted of a story involving one character (e.g., a pig). This character had a set of four objects (e.g., rocks) placed on an orange square in front of them. The character could decide how many (if any) of their objects they would act upon (e.g., carry). The experimenter took on the role of the character and for each of the character's objects, the character went through a process of considering and deciding whether or not they wanted to act upon that object, and then enacting their decision. In the end, the character acted on all four of their objects (e.g., the pig carried all four of their rocks). At this point the experimenter asked the puppet what had happened in the story, to which the puppet responded with the test sentence in (53).

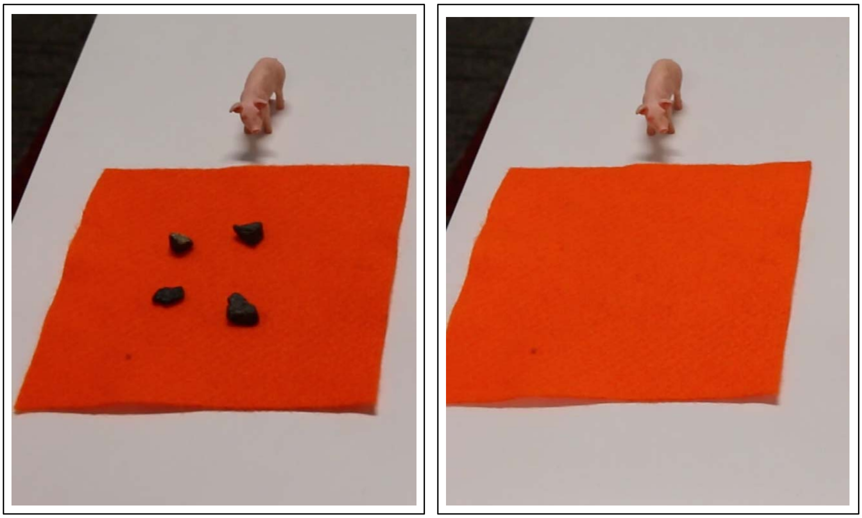
(53) The pig carried some of his rocks.

There were two possible interpretations of our test sentences that participants might access. In the case of a test sentence like (53), the *Literal* interpretation would be (54), which corresponds to the basic truth conditions of the sentence, without any inferences. The interpretation including the *OnlySome* inference would be (55), and corresponds to the *Literal* interpretation enriched with the *OnlySome* inference.

(54) The pig carried at least one of his rocks.

(55) The pig carried at least one of his rocks & <u>The pig did not carry all of his rocks</u>.

13 We were less concerned that children would be influenced in the same way, on the assumption that communicating justifications verbally is less effortful than typing them out.



a: First scene of test item.

b: Final scene of test item. The rocks have all been carried (and placed down outside of the shot).

Figure 1 Scenes accompanying test item in (56).

To illustrate, one of the test items is presented in (56). Photos of the first and final scenes of this story are provided in Figure 1.

(56) Example test condition item

a. *This is a story about this pig [see Figure 1a]. This pig has rocks that he can carry if he wants to. Let's see what he does.*
__Pig:__ "Let me see, I'll carry this rock [carries rock (1/4)], and this one [carries rock (2/4)]. Should I stop? Hmm...I'm feeling really strong today, so I'll also carry this rock [carries rock (3/4)], and this rock too [carries rock (4/4)]." [see Figure 1b]
Experimenter: Okay <Puppet's name>, what happened in that story?

b. **Puppet:** Hmm, the pig carried some of his rocks.

As can be seen, the critical test sentence in (56b) was only consistent with the *Literal* interpretation, i.e., (54), as the character acted upon *all* of their objects. Therefore, it was expected that if a participant settled on the *Literal* interpretation, they would accept the test sentence. In contrast, if a participant's interpretation included the *OnlySome* inference, they were expected to reject the test sentence.

In addition to the critical test items, participants were presented with two practice items at the beginning of the experiment, which were designed to introduce participants to the paradigm. The first of these items was designed to elicit a *yes*-response and the second was designed to elicit a *no*-response. The latter of these items is presented in (57).

(57) Example of practice item

a. **Context:** Snoopy wants to play with a yo-yo.

b. **Outcome:** The yo-yo is too heavy for Snoopy to pick up, so he plays with a slinky instead.
c. **Target sentence:** Snoopy played with the yo-yo.

Participants were also presented with two filler items, which, as outlined in (58), were comprised of simple scenarios paired with basic declarative sentences. The fillers were designed so that they could be paired with either a *yes*-target or a *no*-target. The target sentence was chosen based on a participant's response to the preceding test trial, in order to avoid participants producing more than two *yes*-, or more than two *no*-judgments in a row.

(58) Example of filler item
a. **Context:** Tigger and Buzz are having a competition to see who can throw their hoop the furthest.
b. **Outcome:** Tigger throws his hoop the furthest.
c. **Target sentences:**
i. *Yes*-target: Tigger won the throwing competition.
ii. *No*-target: Buzz won the throwing competition.

In sum, Experiment 1 presented participants with a total of 8 items: 2 practice items, 4 test items, and 2 filler items. The order of presentation of the test and filler items was pseudo-randomised by first creating a random order, and then slightly modifying it to ensure that participants would not accept or reject more than two target sentences in a row. A second version of the experiment was created, with the order of the trials reversed. Presentation order was counterbalanced across participants.

2.2 *Results*

We will consider in turn the truth value judgment responses that participants gave and their justifications for these judgments.

2.2.1 *Judgments* The judgment results are presented in Figure 2. To analyse this data we fitted a mixed effects logistic regression model to the data using the LME4 package in R (Bates *et al.* (2015); R Core Team (2020)). Following the recommendations of Barr *et al.* (2013), we started with a maximal model in which we had Group (Adult vs. Child) as a fixed effect, random by-participant and by-item intercepts, and random by-item slopes. Again following Barr *et al.* (2013), in order to achieve model convergence we: (i) re-coded the fixed effect of Group using deviation coding (also known as sum-coding), (ii) increased the maximum number of iterations in the estimation procedure, and (iii) removed the by-item slopes, based on low variance. We then used a likelihood-ratio test to compare models with and without the fixed effect of Group, revealing a significant effect ($\chi^2(1)=16$, $p<0.001$).

The results suggest that children derived fewer interpretations involving the *OnlySome* inference than adults. While children were quite evenly split between accessing *Literal* interpretations (e.g., (59)) of our test sentences and interpretations including the *OnlySome* inference (e.g., (60)), adults clearly preferred interpretations including the *OnlySome* inference.

(59) ***Literal* interpetation:** The pig carried at least one of his rocks.
(60) **OnlySome interpretation:** The pig carried at least one of his rocks & The pig did not carry all of his rocks.

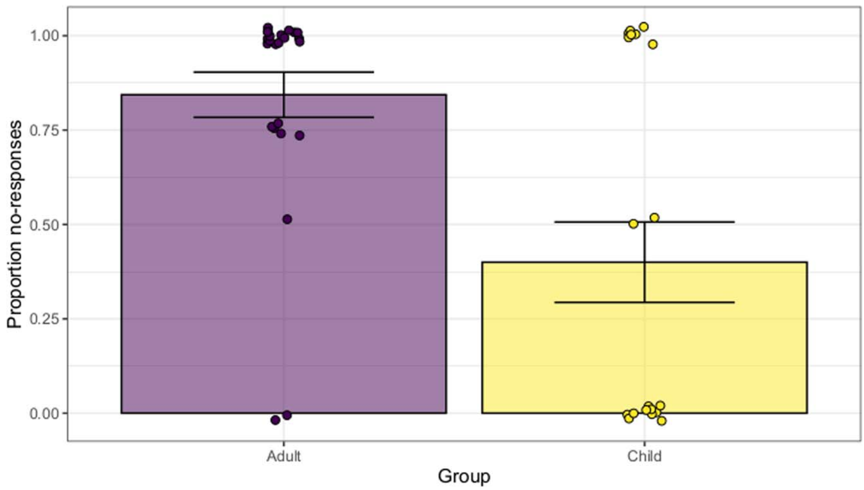


Figure 2 Mean proportion of test sentence rejections across groups. The vertical bars represent the standard error. Dots correspond to individual participants' mean rejection rates. A horizontal jitter of.1 and vertical jitter of.025 were applied for better visualisation.

Additionally, as can be seen in Figure 2, the distribution of the child participants' mean responses was bi-modal, suggesting each child was generally consistent in their preferred interpretation across the different items.

2.2.2 *Justifications* We also elicited justifications for all *no*-responses (recall that *no*-responses were associated with the interpretation involving the *OnlySome* inference).

We recorded 81 *no*-response justifications from adults and 32 from children. All of these justifications (100%) provided an explanation for the *no*-judgment by referring to the fact that the relevant character had acted on *all* of their objects (e.g., *No, he ate every single one; No, she used all of her stars; No, the chicken sold all of his shells*). These responses are consistent with the participants' *no*-judgments being based on having derived *OnlySome* inferences as part of their interpretations of the test sentences.

2.3 *Discussion*

The goal of Experiment 1 was to investigate adults' and children's interpretations of *Some* sentences. Our results are consistent with previous literature (e.g., Noveck (2001); Papafragou & Musolino (2003)) reporting that children derive inference-based interpretations of *Some* sentences at significantly lower rates than adults.

We turn next to Experiment 2, which investigated adults' and children's interpretations of *EverySome* sentences.

3 EXPERIMENT 2

The aim of Experiment 2 was to investigate children's derivation of scalar inferences associated with *EverySome* sentences, namely the *NotEvery* and *None* inferences. The inclusion of this experiment allowed us to explore the possibility that children would more readily access inference-based interpretations of *EverySome* sentences than of *Some* sentences.

3.1 *Method*

3.1.1 *Participants* Eighteen monolingual English-speaking undergraduate students and 31 monolingual English-speaking children (4;00-5;10, M = 4;05) participated in the experiment;[14] none had participated in Experiment 1. The adults took part in the experiment for course credit, or for a payment of 15 AUD. All participants were recruited and tested in Sydney, Australia. Informed consent was obtained from the adult participants, and from the parent/guardian of the child participants. The parents/guardians of the child participants tested in the lab were compensated 20 AUD for travel expenses.

3.1.2 *Procedure* The same Truth Value Judgment Task procedure was used as in Experiment 1.

Children were tested individually, either in the lab or in a quiet room at their daycare. Adults were tested in small groups of up to three participants. The items were split across two sessions, and the sessions were conducted 7-14 days apart. Each session lasted approximately 20 minutes.

3.1.3 *Materials* Experiment 2 investigated the interpretations assigned by participants to *EverySome* sentences in four kinds of contexts. Each context was designed to be consistent with a different set of interpretations.

Critical conditions: Each of the critical conditions contained four items. Each item consisted of a story involving three characters (e.g., pigs). Each of these three characters had a set of four objects (e.g., rocks) placed on an orange square in front of them. Each character could decide how many (if any) of their objects they would act upon (e.g., carry). The experimenter acted out the story, taking on the role of each of the characters in turn. For each of a character's objects, the character went through a process of considering and deciding whether or not they wanted to act upon the object, and then enacting that decision. At the end of the story, the experimenter asked the puppet what had happened in the story, to which the puppet responded with the relevant *EverySome* test sentence.

We will refer to our four critical conditions as 3NONE, 3ALL, 2SOME-1ALL, and 3SOME. In the 3NONE condition, none of the characters acted on any of their objects. In contrast, in the 3NONE condition, each character acted on some but not all of their objects. In the 3ALL condition, every character acted on all of his objects, and finally, in the 2SOME-1ALL condition, two characters acted on two of their four objects, and one character acted on all four of their objects. See Table 1 for a summary of these conditions.

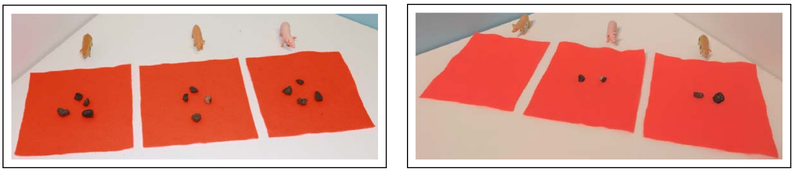
To illustrate, one of the 2SOME-1ALL items is presented in (61). The story in (61a) would be associated with the sentence in (61b). A photo of the final scene of this story is provided in Figure 3b.

(61) Example 2SOME-1ALL item
 a. *This is a story about three pigs [see Figure 3a]. These pigs each have rocks that they can carry if they want to. Let's see what they do:*

14 As we outline in Section 3.1.4, the experiment took place across two sessions; the ages reported correspond to the age of the child participants at the first of the two sessions.

Table 1 Schematic of the contexts for the critical test conditions. The 'context' column displays the number of objects (out of four) that each character acted upon in that condition.

Condition	Context
3Some	2/4, 2/4, 2/4
2Some-1All	2/4, 2/4, 4/4
3All	4/4, 4/4, 4/4
3None	0/4, 0/4, 0/4



a: First scene of the story presented in (61).

b: Final scene of the story presented in (61). The rocks remaining on the orange squares have not been carried.

Figure 3 Scenes from the 2Some-1All item presented in (61).

Pig 1: "Let me see, I'll carry this rock [carries rock (1/4)], and this one [carries rock (2/4)]. Should I stop? Hmm...I'm feeling really strong today, so I'll also carry this rock [carries rock (3/4)], and this rock too [carries rock (4/4)]."
Pig 2: "Let me see, I'll carry this rock [carries rock (1/4)], and this one [carries rock (2/4)]. Should I stop? Yes I will, as I am tired."
Pig 3: "Let me see, I'll carry this rock [carries rock (1/4)], and this one too [carries rock (2/4)]. Should I stop? Hmm...I'm feeling really strong today, so I'll also carry this rock [carries rock (3/4)], and this rock too [carries rock (4/4)]." [see Figure 3b]
Experimenter: Okay <Puppet's name>, what happened in that story?

b. **Puppet:** Hmm, every pig carried some of his rocks.

Control conditions and filler items: Our experiment also included two control conditions, each with two items. These control conditions were designed to ensure that participants understood the basic meaning of the universal quantifier "every". On these trials, participants heard a sentence like (62) either in a context that made the sentence clearly true (i.e. paired with a 2Some-1All context), or in a context that made the sentence clearly false (i.e. paired with a 2Some-1None context).

(62) Every pig carried rocks.

In addition to the test and control items, participants also received five filler items. Each filler item was designed so that it could be paired with one of two possible sentences, one designed to elicit a *no*-response and one designed to elicit a *yes*-response. The experimenter chose the filler sentence based on a participant's responses to previous trials, so as to avoid having more than two *yes*- or more than two *no*-judgments in a row. For example, if a participant had rejected the target sentences of the two items preceding a filler item, then the filler sentence associated with a *yes*-response would be used.[15]

3.1.4 *Test sessions* Participants were presented with all of these items over the course of two sessions presented 7-14 days apart. The conditions were split up between these sessions, in the manner outlined in (63).

(63) a. **Session A:** 3SOME, 3ALL EVERY_TRUE, EVERY_FALSE, FILLERS
b. **Session B:** 3SOME-1ALL 3NONE, FILLERS

The ordering of the items within each session was pseudo-randomised by first creating a random order, and then slightly modifying it such that participants were not expected to accept or reject more than two target sentences in a row (assuming they gave expected responses to filler items). A second version of each session was then created, with the order of the trials reversed. The version of the sessions as well as the order in which the sessions were presented was counterbalanced across participants.

3.1.5 *Identifying interpretations* There were three possible interpretations of our *EverySome* test sentences (i.e. (64)) that participants might access. One possible interpretation contains only the *Literal* meaning in (65), which corresponds to the basic truth conditions of the sentence, without any inferences. Another possible interpretation includes the *NotEvery* inference, paraphrased as in (66); this interpretation corresponds to the *Literal* interpretation enriched with the *NotEvery* inference. A third possible interpretation includes the *None* inference, and can be paraphrased as in (67); this interpretation corresponds to the *Literal* interpretation enriched with the *None* inference. Note that an interpretation containing the *None* inference entails an interpretation containing the *NotEvery* inference. Table 2 outlines the relationship between our different test conditions and the interpretations in (65)-(67).

(64) Every pig carried some of his rocks.
(65) Every pig carried at least one of his rocks (*Literal*)
(66) Every pig carried at least one of his rocks (*Literal*) & *Not every pig carried all of his rocks (NotEvery)*
(67) Every pig carried at least one of his rocks (*Literal*) & *None of the pigs carried all of his rocks (None)*

By comparing participants' responses across the four critical conditions, we could determine which interpretations were being accessed in the following way. A difference in rejection rates between the 3NONE and the 3ALL conditions was evidence that participants accessed to some extent interpretations involving only the *Literal* meaning. A difference in rejection rates between the 3ALL and 2SOME-1ALL conditions was evidence of an interpretation that included (in addition to the *Literal* meaning) the *NotEvery* inference. Finally, a difference

15 Such dynamic fillers have been used in a number of previous acquisition studies (e.g., Musolino (2004); Lidz & Musolino (2006); Tieu *et al.* (2016)). However, as an anonymous reviewer notes, the use of such dynamic fillers means that participants may in fact be presented with different experiments, to the extent that they are presented with different versions of the filler items. Given this potential weakness, we hope that future work can explore alternative solutions to the concern of participants being biased by giving too many consecutive *yes/no*-responses.

Table 2 The relationship between the test conditions and the interpretations of *EverySome* sentences that they are consistent with.

Condition	Consistent interpretations
3Some	*Literal* & *NotEvery* & *None*, e.g., (65)-(67)
2Some-1All	*Literal* & *NotEvery*, e.g., (65)-(66)
3All	*Literal*, e.g., (65)
3None	None of the relevant interpretations

Table 3 Results of Wilcoxon signed-rank tests comparing differences between conditions within each group. *Significant at α =.05, based on the Holm-Bonferroni correction procedure.

Group	Comparison	Test statistic	Effect size	*p*-value
Child	3None vs. 3All.	Z = -4.22	r = -.54	p =.000*
	3All vs. 2Some-1All	Z = -2.1	r = -.27	p =.036
	2Some-1All vs. 3Some	Z = -3.84	r = -.49	p =.000*
Adult	3None vs. 3All	Z = -2.82	r = -.47	p =.005*
	3All vs. 2Some-1All	Z = -2.92	r = -.49	p =.004*
	2Some-1All vs. 3Some	Z = -1.66	r = -.28	p =.098

in rejection rates between the 2Some-1All and 3Some conditions was evidence of an interpretation that included (in addition to the *Literal* meaning) the *None* inference.

3.2 *Results*

We will consider in turn the truth value judgment responses that participants gave and their justifications for these judgments.

3.2.1 *Judgments* The binary truth value judgment results are presented in Figure 4. To investigate which of our target interpretations were accessed by our participant groups, we ran a series of Wilcoxon signed-rank tests to determine whether, for each participant group, there were any differences between conditions. To investigate whether there were any differences between adults' and children's responses within each condition, we also ran a series of Wilcoxon rank-sum tests. Once we had generated the *p*-values for these tests, we used the Holm-Bonferroni procedure (Holm (1979)) to determine which contrasts were significant at an alpha value of '.05'. Tables 3 and 4 present the results of the Wilcoxon signed-rank and rank-sum tests, respectively. We opted to use this statistical analysis in addition to the mixed effects logistic regression analysis outlined below because: (i) it allowed us to include all of the test conditions[16], and (ii) the same analysis was used in Chemla & Spector (2011), which had a very similar experimental design.

As mentioned, we also fitted a maximal mixed-effects logistic regression model using the lme4 package in R (Bates *et al.* (2015); R Core Team (2020)), with Group (Adults vs. Children), Condition (2Some-1All vs. 3All), and their interaction as fixed effects, random

16 We could not include the 3Some or 3None conditions in the logistic regression analysis because the adult responses in these conditions did not vary at all, resulting in non-convergence if they were included in the model.

Table 4 Results of Wilcoxon rank-sum tests, comparing differences between groups within each condition. *Significant at α =.05, based on the Holm-Bonferroni correction procedure.

Condition	Comparison	Test statistic	Effect size	p-value
3None	Child vs. Adult	$W = 252$	$r = -.19$	$p = .186$
3All		$W = 220$	$r = -.19$	$p = .189$
2Some-1All		$W = 424$	$r = -.47$	$p = .001$*
3Some		$W = 333$	$r = -.28$	$p = .051$

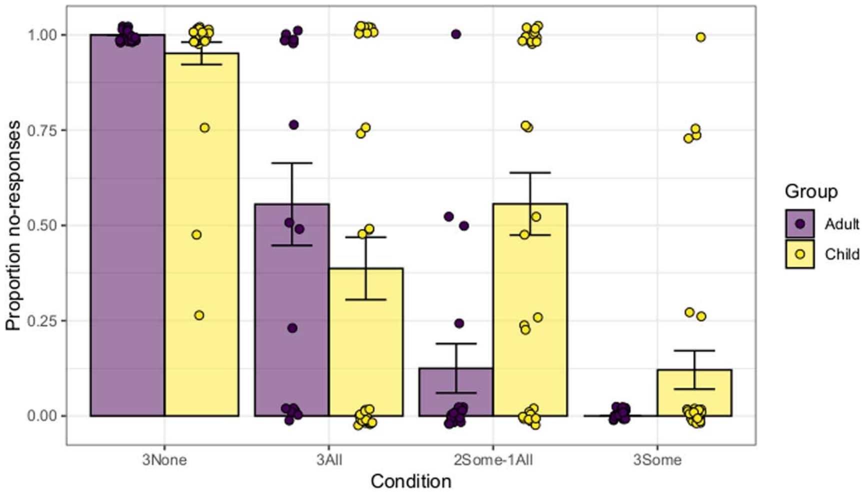


Figure 4 Mean proportion of test sentence rejections across our test conditions. The vertical bars represent the standard error. Dots correspond to individual participants' mean rejection rates. A horizontal jitter of.1 and vertical jitter of.025 were applied for better visualisation.

by-item and by-participant intercepts, random by-item slopes for Group, and random by-participant slopes for Condition. Following Barr *et al.* (2013), in order to achieve model convergence and avoid a singular fit, we: (i) re-coded the factors of Group and Condition using deviation coding (also known as sum-coding), and (ii) removed the by-item slopes for Group, and removed the item intercept, based on low variance. We then used a likelihood-ratio test to compare models with and without the fixed effects of Group and Condition, as well as their interaction. These comparisons revealed no effect of Group ($\chi^2(1)$=0.804, p=0.37) or Condition ($\chi^2(1)$=2.16, p=0.142), but a significant interaction between Group and Condition ($\chi^2(1)$=7.157, p<0.01). Overall the results of this analysis are consistent with the results of the previous analysis outlined in Tables 3 and 4.

These results lead us to the following conclusions regarding how participants engaged with the target interpretations. We observed that both children and adults displayed a statistically significant difference in target sentence rejections between the 3None and 3All conditions, providing evidence that both groups accessed, to some extent, interpretations involving only the *Literal* meaning in (68). We also observed a significant difference between adults' responses in the 3All and 2Some-1All conditions, suggesting they accessed interpretations involving the *NotEvery* inference (i.e. (69)); children's responses in these conditions did not differ significantly, and thus we have no evidence that they accessed such interpretations. Children did, however, display a difference in their responses to the 2Some-1All and 3Some conditions, providing evidence that they accessed interpretations involving the *None* inference (i.e. (70)). Adults, on the other hand, did not display this difference; thus we do not have any evidence that adults accessed interpretations including the *None* inference.

(68) Every pig carried at least one of his rocks.
(69) Every pig carried at least one of his rocks & *Not every pig carried all of his rocks*.
(70) Every pig carried at least one of his rocks & *None of the pigs carried all of his rocks*.

Comparing target sentence rejections between adults and children across each condition (Table 4), we only found a difference between adults and children in the 2Some-1All condition.

These results reveal both similarities and differences between our participant groups. Both groups accessed *Literal* interpretations as well as inference-based interpretations (i.e. interpretations involving either the *NotEvery* inference or the *None* inference) at similar rates. However, the two groups accessed different inference-based interpretations. While children accessed interpretations involving the *None* inference, adults accessed interpretations involving the *NotEvery* inference. This conclusion is corroborated by the different acceptance rates between participant groups in the 2Some-1All condition, as this was the only condition where these different inference-based interpretations would have resulted in different judgments.

3.2.2 *Justifications* We also elicited justifications whenever a participant gave a '*no*-response' to a test sentence. We will focus here only on the justifications provided in the 3All and 2Some-1All conditions (see Appendix A for a full description of the justifications across conditions).

3All: Adults produced 56% (40/72) *no*-responses in the 3All condition and child participants produced 39% (48/124) *no*-responses in this condition. As this condition was inconsistent with both the *None* inference and the *NotEvery* inference, *no*-responses were interpreted as evidence that participants had derived one of these inferences. The bulk of the justifications provided by both our adult and child groups in this condition (100% and 92%, respectively) pointed out that all of the characters had acted on all of their objects (e.g. *All of the cats throwed all of their glowsticks; They all burned all their sticks*). Notably, these justifications are consistent with the associated *no*-responses being motivated by the participants having derived a *NotEvery* or *None* inference. That is, this is the kind of justification one might expect if a participant was interpreting *EverySome* sentences as conveying that *not every/none* of the relevant characters acted on *all* of their objects. It is nevertheless also possible that the participants were merely repeating what had happened in the story. At the very least, these justifications help to rule out the possibility that participants' *no*-responses in this condition were motivated by irrelevant considerations (e.g., not liking the puppet).

2Some-1All: Adult participants produced 12.5% (9/72) *no*-responses in the 2Some-1All condition and child participants produced 56% (69/124) *no*-responses. This condition was only inconsistent with interpretations that included a *None* inference. The bulk of the adults' and children's justifications in this condition (67% and 59%, respectively) focused on the fact that *one* of the characters had acted on *all* of their objects (e.g. *Because this one*

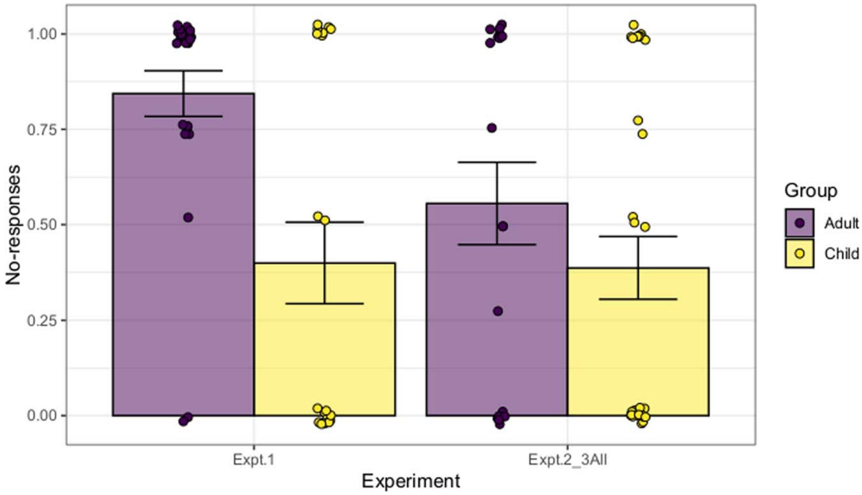


Figure 5 Mean proportion of *no*-responses across the two experiments. The vertical bars represent the standard error. Dots correspond to individual participants' mean rejection rates. A horizontal jitter of.1 and vertical jitter of.025 were applied for better visualisation.

carried all of them; One ate all of them). This is precisely the kind of response one would expect if the participants were deriving the inference that *none* of the characters acted on *all* of their objects. Most of the remaining *no*-response justifications (20% for adults, 29% for children) repeated everything that happened in the story (e.g. *None, two, two; Only two used some and one used all of them*). While such justifications do not clearly indicate that the associated judgments were motivated by a *None* inference, they are entirely consistent with them being so motivated.

The remaining *no*-response justifications in this condition (1% for adults, 12% for children) were coded as 'Two-NotAll'. These justifications seem to indicate that the associated responses were motivated by some non-target interpretation on which there was an expectation that all of the characters would act on all of their items. We will return to this in Section 4.4.

3.3 *Comparing the experiments*

Before we consider the results of Experiment 2 in more detail, let us contrast these results with the results from Experiment 1. Experiment 1 revealed that adults were significantly more likely than children to access inference-based interpretations of *Some* sentences. In contrast, Experiment 2 revealed that adults and children accessed inference-based interpretations of *EverySome* sentences at similar rates. In order to explore this contrast further, we compared the results of our two experiments using a mixed-effects logistic regression analysis. Specifically, we compared the Experiment 1 test condition to the 3All condition in Experiment 2. We selected this condition from Experiment 2 because the contexts in this condition were inconsistent with both of the relevant inferences; participant responses to this condition, therefore, provide a measure of the rate of inference-based interpretations to *EverySome* sentences. In this way, we could compare the inference-based interpretations that adults and children accessed for *Some* sentences versus *EverySome* sentences. Note that the participants in these two experiments were different, so Condition was a between-group factor. We fitted a mixed-effects logistic regression model using the LME4 package in R (Bates *et al.* (2015); R Core Team (2020)), including Group (Adults vs. Children), Experiment (Exp. 1 vs. Exp. 2_3ALL), and their interaction as fixed effects, random by-item and by-participant intercepts, and random by-item slopes for Group. Following Barr *et al.* (2013), in order to achieve model convergence and avoid singular fit, we: (i) re-coded the factors of Group and Experiment using deviation coding (also known as sum-coding), (ii) increased the maximum number of iterations in the estimation procedure, and (iii) removed the correlation parameter between random slopes and random intercepts for the random effect of Item. We then used a likelihood-ratio test to compare models with and without the fixed effects of Group and Experiment, and their interaction. This analysis revealed a significant effect of Group ($\chi^2(1)=23$, $p<0.001$), but no effect of Experiment ($\chi^2(1)=0.892$, $p=0.345$) or interaction between Group and Experiment ($\chi^2(1)=1.604$, $p=0.205$). As can be seen in Figure 5, the significant effect of Group was driven by adults accessing more inference-based interpretations overall than children.

3.4 *Discussion*

The aim of Experiment 2 was to investigate adults' and children's interpretations of *EverySome* sentences. Specifically, we were interested in the rates at which such interpretations would include the associated *NotEvery* or *None* inferences. The results of Experiment 2 suggest that adults and children derive inference-based interpretations of *EverySome* sentences at similar rates. However, the specific inference on which that interpretation is based appears to differ between adults and children, with children preferring interpretations involving the *None* inference, and adults preferring interpretations involving the *NotEvery* inference.

The adult results are consistent with previous work showing that when interpreting *EverySome* sentences, adults prefer interpretations including the *NotEvery* inference over those including the *None* inference (Geurts & Pouscoulous (2009); Clifton & Dube (2010); Chemla & Spector (2011); Potts *et al.* (2016)). As for children, this is the first time children's interpretations of *EverySome* sentences have been investigated. The finding that children derived inference-based interpretations of *EverySome* sentences at the same rate as adults contrasts with the results of previous studies investigating children's interpretation of *Some* sentences (e.g., Noveck (2001); Papafragou & Musolino (2003)), including our Experiment 1. Moreover, the finding that adults and children preferred different inference-based interpretations presents a new puzzle. Why should children differ from adults in the observed manner, and what factors lead children to ultimately settle on the adult interpretation of such sentences?

Interestingly, when we analysed the results of the two experiments together, we only found evidence of adults accessing more inference-based interpretations overall compared to children. The analysis revealed no evidence that *EverySome* sentences had any sort of facilitatory effect on children's access to such interpretations.

In the following section, we will explore these different puzzles further and propose some possible ways of accounting for them.

4 GENERAL DISCUSSION

4.1 *Main findings*

The goal of this paper was to investigate children's interpretations of *EverySome* sentences. We focused on these sentences because of the possibility, raised by the Alternatives-based approach, that children might access more inference-based interpretations of these sentences compared to *Some* sentences. We conducted this investigation by exploring adults' and children's interpretations of both *Some* and *EverySome* sentences.

Experiment 1 investigated adults' and children's interpretations of *Some* sentences, like (71), to determine the extent to which they accessed interpretations including the associated *OnlySome* inference in (72).

(71) The pig carried some of his rocks.
(72) The pig didn't carry all of his rocks.

The results of Experiment 1 replicated those of previous studies (Noveck (2001); Chierchia *et al.* (2001); Papafragou & Musolino (2003)), with children accessing fewer inference-based interpretations of *Some* sentences than adults.

Experiment 2 investigated adults' and children's interpretations of *EverySome* sentences, like (73), to measure the extent to which they derived the associated *NotEvery* inference in (74) and the *None* inference in (75).

(73) Every pig carried some of his rocks.
(74) Not every pig carried all of his rocks.
(75) No pig carried all of his rocks.

Experiment 2 revealed that adults and children derived inference-based interpretations of *EverySome* sentences at similar rates. However, the experiment also revealed that the specific inference on which these interpretations were based differed between adults and children, with adults preferring *NotEvery* inferences, and children preferring *None* inferences. The adult results are consistent with previous literature (Geurts & Pouscoulous (2009); Clifton & Dube (2010); Chemla & Spector (2011); Potts *et al.* (2016)), while the child results represent the first data on children's interpretation of *EverySome* sentences. The reason for the observed difference between adults' and children's interpretations of *EverySome* sentences is not immediately clear. We will return to this below.[17]

Finally, a comparison of the two experiments found that, overall, adults derived more inference-based interpretations than children, which is consistent with previous literature (Noveck (2001); Papafragou & Musolino (2003)). We discuss the implications of this pattern below.

4.2 *The Alternatives-based approach*

One of the reasons we investigated children's interpretations of *EverySome* sentences was to explore a possibility raised by the Alternatives-based approach, namely that children might

17 An anonymous reviewer notes that, while the difference is not statistically significant, children give *no*-responses more often in the 2SOME-1ALL condition than in the 3ALL condition. This is surprising because the interpretation that should lead to a *no*-response in the 2SOME-1ALL condition asymmetrically entails the interpretation that should lead to a *no*-response in the 3ALL condition. Therefore, we might expect that there should be at least as many *no*-responses in the 3ALL condition as there are in the 2SOME-1ALL condition. One possibility is that there was a subset of children that required all of the characters to behave in a uniform manner in order to give a *yes*-response. Such children would be expected to give a *no*-response in the 2SOME-1ALL condition, but not in the 3ALL condition, which would account for the observed pattern. We discuss this possibility further in Section 4.4.

more readily derive inferences from such sentences than has previously been found for *Some* sentences (Noveck (2001); Chierchia *et al.* (2001); Papafragou & Musolino (2003)). This possibility is driven by the hypothesis that children's failure to derive scalar inferences is the result of difficulties accessing the relevant alternative sentences.

According to the Alternatives-based approach, when children are presented with *Some* sentences in a context where the relevant alternatives are not made salient, they struggle to derive the *OnlySome* inference because they cannot generate the required alternatives (e.g., Noveck (2001)). The Alternatives-based approach also posits that, in contexts where children are assisted in accessing the relevant alternatives, they will more readily derive the associated inference. For example, when the alternative sentences are presented as sub-constituents of the original sentence, children will more readily derive the associated inference (e.g., Tieu *et al.* (2016)).

In the case of *EverySome* sentences, the relevant alternative sentences are generated by replacing scalar terms in the original sentence with the existential quantifier "some" or the universal quantifier "all". Therefore, the fact that the *EverySome* sentence explicitly presents each of these quantifiers in some form (i.e. the quantifiers "some" and "every" appear in the sentence) could conceivably facilitate the generation of the relevant alternatives, and by extension, the associated inference-based interpretations. This would lead us to expect that children might more readily access inference-based interpretations of *EverySome* sentences than *Some* sentences.

Experiment 1 revealed that children accessed inference-based interpretations of *Some* sentences at a lower rate than adults. Experiment 2 revealed that adults and children accessed inference-based interpretations of *EverySome* sentences at the same rate. However, when we included both experiments in the same statistical analysis, we only found that overall, children accessed fewer inference-based interpretations than adults; we did not observe a significant effect of Experiment or an interaction between Group and Experiment. If presenting the quantifier alternatives within the *EverySome* sentences facilitated children's access to the target inferences, and this drove the between-group similarity observed for *EverySome* sentences, one might have expected to observe a significant interaction between Group and Experiment, with children showing a difference between the two sentence types. We thus do not have any experimental evidence that the presence of the universal quantifier within the *EverySome* sentence facilitates children's generation of the associated alternatives and inferences. We should note that the findings don't necessarily present evidence *against* the Alternatives-based approach per se, which doesn't make explicit predictions about *EverySome* sentences specifically. Adopting the approach simply raises the possibility that *EverySome* sentences might facilitate inference derivation in a way that *Some* sentences do not.[18]

To summarize, the results of our investigation did not realize the possibility, inspired by certain 'lexicon-focused' versions of the Alternatives-based approach, that children would more readily derive scalar inferences from *EverySome* sentences than from *Some* sentences.

18 Note again that the version of the Alternatives-based approach presented in Skordos & Papafragou (2016) does not necessarily predict that children should differentiate *Some* and *EverySome* sentences in the way we have discussed. In fact, the approach in Skordos & Papafragou (2016) might even predict no difference in children's derivation of scalar inferences from these sentences, given our experiments were not designed to make them differ in terms of the *relevance* of their respective alternatives.

4.3 *Principles of interpretation*

As mentioned, we found a difference between adults' and children's preferred inference-based interpretations of *EverySome* sentences: adults accessed *NotEvery* inferences, while children tended to access *None* inferences. Any explanation of this observed difference still needs to be able to explain how, through the course of development, children eventually come to adopt an adult interpretation strategy for *EverySome* sentences. One possible avenue for developing such an explanation would be to invoke existing theories that predict differences in how adults and children interpret ambiguous sentences.

Sentences that can be associated with multiple possible meanings raise a special challenge for child language learners and adult language users alike; the ambiguity needs to be resolved in some way. There are a number of principles in the theoretical literature that are proposed to guide adults and children in this regard. One such principle is the *Subset Principle*, which is proposed to guide children to prefer stronger (i.e. subset) meanings (Berwick (1985); Crain *et al.* (1994); Crain & Thornton (1998); Notley *et al.* (2012); Moscati *et al.* (2016)). It is proposed that children do this for learnability reasons, as initially preferring stronger interpretations means that children can learn about the existence of any weaker interpretations through positive evidence. In the case of the inference-based interpretations of *EverySome* sentences, the strongest interpretation is the one with the *None* inference, as it entails the other relevant interpretations. The fact that children in our experiment preferred such interpretations is therefore consistent with the Subset Principle. Importantly, adults are not assumed to be guided by the Subset Principle when interpreting ambiguous sentences, as it is meant to be a developmental principle that explains how children arrive at the possible interpretations in the language they are being exposed to, something that adult speakers have already achieved. Therefore, the fact that adults' interpretations did not appear to be influenced by this principle is entirely expected.[19]

Unlike children, adults favoured interpretations that included the *NotEvery* inference. This behavior can be quite straightforwardly accounted for as being motivated by the *Principle of Charity*, a general principle that speakers are thought to employ by default when faced with an ambiguous sentence (Grice (1975)). The Principle of Charity leads hearers to prefer interpretations that make a sentence true in a given context. This principle could have encouraged our adult participants to prefer interpretations involving a *NotEvery* inference over those involving a *None* inference, as the former interpretation was true in more of our test conditions.

One might note that *Literal* interpretations are, in fact, an even weaker interpretation of *EverySome* sentences than those involving *NotEvery* inferences. Therefore, if all three target interpretations were under consideration, then the Principle of Charity should have encouraged adults to prefer *Literal* interpretations of our test sentences. If we assume, however, that the relevant participants preferred inference-based interpretations over literal

19 We should note that the Subset Principle is usually invoked to account for ambiguities that arise when a sentence is associated with multiple underlying structures, e.g., scopally ambiguous sentences for which one interpretation entails the other. One might think that such a principle should not apply to ambiguities that arise through the presence of scalar inferences. According to the Grammatical account of scalar inferences (Fox (2007); Chierchia (2013)), however, the different scalar inferences we have targeted can be modelled as involving different underlying syntactic structures (see Section 1.3.1). Appealing to the Subset Principle as an explanation for our findings could therefore tie in quite well with the Grammatical account of scalar inferences.

ones, then the Principle of Charity would only be predicted to influence the choice between the two inference-based interpretations.

One might also ask why children were not similarly affected by the Principle of Charity. One hypothesis is that when both the Subset Principle and Principle of Charity are at play in development, the Subset Principle wins out in guiding children's interpretations, as it allows them to learn the possible interpretations in the language they are being exposed to, a more important goal at this stage of development.

Finally, as noted by an anonymous reviewer, there are certain tensions between the Subset Principle and the 'Pragmatic Tolerance Hypothesis' proposed by Katsos & Bishop (2011). On the one hand, the Subset Principle predicts that children will be more restrictive in terms of the interpretations they accept for a given sentence. On the other hand, Katsos & Bishop (2011) suggest that children are less restrictive in a sense, in that they are relatively more tolerant of pragmatic violations than adults are. One possible way of reconciling these two views is to highlight that these two mechanisms are actually intended to capture different phenomena. That is, there is a distinction between identifying that a target sentence is a non-optimal way of describing a given context, and deriving an inference-based interpretation of a target sentence. More specifically, as outlined in Katsos & Bishop (2011), accessing an inference-based interpretation of (76) (i.e. (77)) requires at least two steps. First, the hearer must identify that there was a more informative way of describing the relevant context (i.e. (78)). And second, the hearer must interpret the speaker's choice not to say (78) as inferring its negation, thereby deriving (77). Katsos and Bishop's results could be interpreted as evidence that, while children can perform the first step as readily as adults (i.e. identifying that (76) is non-optimal), they often struggle with the second step (i.e. interpreting the use of (76) as implying (77)).

(76) The pig carried some of his rocks.
(77) The pig carried some but not all of his rocks.
(78) The pig carried all of his rocks.

In contrast, the Subset Principle could be seen as only applying in cases where children successfully complete both steps and so are in a position to access a genuine inference-based interpretation (i.e. (77)). In this way, Katsos and Bishop's Pragmatic Tolerance Hypothesis and the Subset Principle need not be in conflict with one another.

Future work could further investigate the relative role of these different interpretation principles in guiding adults' and children's access to inference-based interpretations. We will simply note here that an explanation along these lines could account for the observed differences between adults' and children's behavior in our study.

4.4 *Alternative explanations*

Before concluding, we would like to explore some alternative explanations for the results from Experiment 2. Specifically, we will explore the following four possibilities: i) that our unbalanced sessions artificially inflated certain kinds of responses, ii) that children's responses were influenced by a 'quantifier spreading' interpretation, iii) that children's responses were motivated by a desire for the characters' actions to be uniform, and iv) that *None* inferences are for some reason easier for children to derive.

First, an anonymous reviewer suggests that participants' responses may have been influenced by the fact that our experimental conditions were split across two different sessions, such that each session contained a different mix of items. For example, 3Some items, which were consistent with all of the targeted interpretations, were presented in the same session as 3All items, which were only consistent with the *Literal* meaning. This contrast could plausibly have inflated the rate of *no*-responses to 3All items, as these might have appeared less acceptable when contrasted with the 3Some items. However, as we outline in detail in Appendix B, when we consider the specific predictions of such an effect, they are not in line with the bulk of our data.

The second possibility we will consider is that children in Experiment 2 were accessing some version of what has been called a 'quantifier spreading' interpretation of our test sentences. Quantifier spreading refers to a particular non-adult-like interpretation of universal quantification where children judge a sentence like "Every pig carried a rock" as true if, in addition to every pig having carried a rock, it was also the case that every rock had been carried by a pig (e.g., Philip (1991); see Philip (2011) for a review). Applied to our *EverySome* sentences, a quantifier spreading interpretation would require that every pig carried at least one of his rocks, *and* that at least one of every pig's rocks was carried. Looking at our critical conditions, all of them except for the 3None condition were consistent with such a requirement. Therefore, if a child was accessing such an interpretation, they would be expected to give *yes*-responses in all but the 3None condition. However, this was not the case, with participants producing a substantial number of *no*-responses across both the 3All (i.e. 39%) and 2Some-1All (i.e. 56%) conditions. Such *no*-responses are not consistent with participants accessing a quantifier spreading interpretation.

Another possibility one might entertain is that children's *no*-responses in the 2Some-1All condition of Experiment 2 were motivated merely by a desire for the characters to behave uniformly, rather than resulting from the derivation of a *None* inference. The issue with this explanation is that it predicts that children should have accepted target sentences in the 3All condition because, in contrast to the 2Some-1All condition, all the characters behaved uniformly in this condition. However, children rejected test sentences in the 3All and 2Some-1All conditions at similar rates (i.e. we observed no statistically significant difference). The similar rates of *no*-responses in these two conditions are difficult to account for if children's behavior was motivated by a desire for the characters' actions to be uniform. In sum, while these three alternative possibilities were worth exploring, upon closer inspection, they seem to us to be less plausible than our proposal that participant responses were motivated by the targeted inference-based interpretations. As for the fourth possibility, one might wonder whether children preferred the *None* inference because it is for some reason easier for them to derive. Recall that there are two possible ways of arriving at a *None* inference, one of which involves deriving an embedded inference (see Section 1.3.1); it could be that children simply have a preference for embedded inferences, independently of a preference for strong meanings. We leave to future work a more detailed investigation of this possibility.

5 CONCLUSION

Many previous developmental studies of scalar implicatures have reported that children access fewer inference-based interpretations of *Some* sentences than adults (Noveck (2001); Papafragou & Musolino (2003); Guasti *et al.* (2005)). No previous research has investigated children's interpretations of *EverySome* sentences. However, the Alternatives-based approach (Barner & Bachrach (2010); Barner *et al.* (2011); Tieu *et al.* (2016); Singh *et al.*

(2016); Skordos & Papafragou (2016)) gives rise to the interesting possibility that children might more readily derive inferences from *EverySome* sentences than *Some* sentences, given that the former contain certain key lexical items that may facilitate children's access to the associated *NotEvery* and *None* inferences.

In this paper, we presented two experiments: Experiment 1 investigated adults' and children's interpretations of *Some* sentences and replicated previous findings, with children accessing fewer inference-based interpretations than adults; Experiment 2 investigated *EverySome* sentences and revealed that children accessed inference-based interpretations of such sentences at the same rate as adults. Experiment 2 also revealed that while children preferred interpretations involving a *None* inference, adults preferred those based on a *NotEvery* inference. The adult results are consistent with previous literature (Geurts & Pouscoulous (2009); Clifton & Dube (2010); Chemla & Spector (2011); Potts *et al.* (2016)) and can be explained by appealing to the Principle of Charity (Grice (1975)). The child results represent the first data on how children interpret such sentences. Analysing the experiments together revealed that, overall, adults accessed inference-based interpretations at a greater rate than children; there was no evidence that children accessed inference-based interpretations of *EverySome* sentences more easily than they did for *Some* sentences. In other words, the presence of the universal quantifier in the *EverySome* sentences did not appear to facilitate children's access to the associated alternatives and inferences.

We have suggested and explored some possible explanations for our data, including the proposal that adults and children are guided by different principles of interpretation. Future work might continue to explore the interpretation of sentences containing multiple scalar terms, as these appear to provide a promising avenue to further our understanding of children's acquisition of scalar inferences.

Author contributions

C.B., E.P., JR., L.T., and S.C. conceived the project; C.B. and E.P. designed the experiment in consultation with J.R., L.T., and S.C; C.B. and E.P. collected the data; C.B. analysed the data; C.B. prepared the initial manuscript; C.B., E.P., J.R., L.T., and S.C. revised the manuscript.

Data availability

The data for both experiments can be accessed at https://semanticsarchive.net/Archive/jM1N2JiN/.

Acknowledgements

We would like to thank all the families and adult participants who took part in the study. For helpful feedback and discussion, we would like to thank Emmanuel Chemla, Maria Teresa Guasti, Maribel Romero, Uli Sauerland, Nicole Gotzner, Philippe Schlenker, Benjamin Spector, Rosalind Thornton, Kazuko Yatsushiro, the audiences of SuB20, BUCLD45, and GALANA2021, and the members of the CCD Language Acquisition Group at Macquarie University. The research leading to these results was supported by the Australian Research Council Centre of Excellence in Cognition and its Disorders (CE110001021), and the Deutsche Forschungsgemeinschaft via the project '367130212' and the priority program 'SPP 1727'.

A Justifications for *no*-responses in Experiment 2

In Experiment 2, justifications were elicited from participants if they gave a *no*-response to a target sentence. Table A5 displays the proportion of *no*-responses to each of the conditions.

Table A5 Proportion of *no*-responses to each of the conditions in Experiment 2.

Condition	Number of *no*-responses	
	Adult	Child
3Some	0/72 (0%)	15/124 (12%)
2Some-1All	9/72 (12.5%)	69/124 (56%)
3All	40/72 (56%)	48/124 (39%)
3None	72/72 (100%)	118/124 (95%)

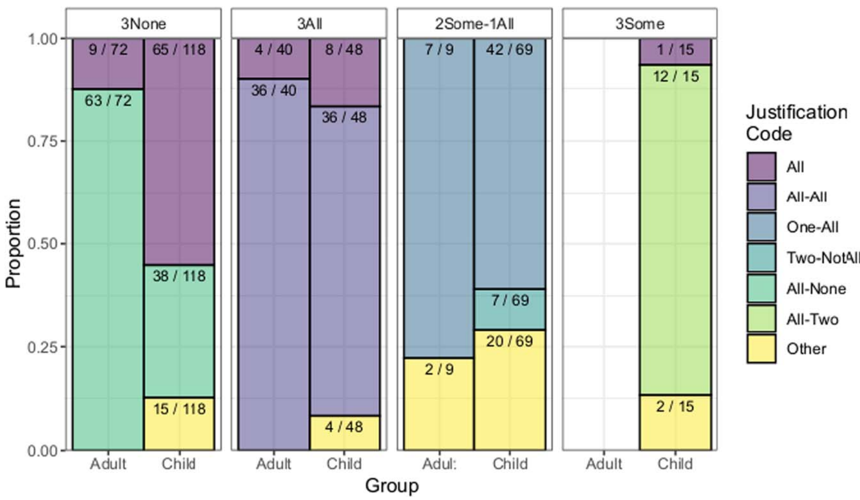


Figure A6 Participants' justifications for their *no*-responses. Proportions indicated within the bars provide the counts for each justification type.

We examined the content of the justifications and determined that they could be naturally divided into 7 categories. The label for these justification categories and four examples for each category are shown in Table A6. Figure A6 displays the distribution of these justification categories across the conditions of Experiment 2. Note that *no*-responses were not anticipated for the 3Some condition, as the contexts in this condition were consistent with all of the targeted interpretations. Nevertheless, 12% of child participant responses were rejections in this condition (see Table A5).

B Differences across experimental test sessions

An anonymous reviewer noted that splitting the conditions between two sessions in the way we did meant that there was a difference in the mix of items that were presented in each session. Such variation would not be expected to affect responses to the 3None or 3Some items, as these were consistent with none or all of the relevant interpretations, respectively. However, such variation could potentially have affected participants' responses to the 3All and 2Some-1All items, which were consistent with only some of the targeted interpretations. Let us consider more precisely how this might come about.

Table A6 Examples of the different justification types (participant IDs provided in parentheses).

Justification type	Examples of justifications
	They all burned their sticks. (ES.Child.10)
All	*They all polished their gemstones. (ES.Child.23)*
	The turtles all slept instead. (ES.Adult.13)
	Every rabbit used teabags to make tea. (ES.Adult.17)
	All of the cats throwed all of their glowsticks. (ES.Child.2)
All-All	*Because they burnt every single one. (ES.Child.9)*
	Every rabbit used all of their teabags (ES.Adult.5)
	Each lion burned all of their sticks. (ES.Adult.16)
	Because this one carried all of them. (ES.Child.1)
One-All	*One ate all of them. (ES.Child.13)*
	One dog ate all of his beans (not just 'some'). (ES.Adult.5)
	The third sheep lit all of his matches. (ES.Adult.12)
	Because these pigs didn't carry all of their rocks, like this one. (ES.Child.5)
Two-NotAll	*Because these two didn't use all of them. (ES.Child.18)*
	Two of them didn't. (ES.Child.24)
	These two didn't. (ES.Child.6)
	All of the rabbits didn't eat any of their pellets. (ES.Child.3)
All-None	*They didn't carry any. (ES.Child.17)*
	None of the three turtles rolled any of their marbles. (ES.Adult.2)
	No smurfs used any band-aids. (ES.Adult.15)
	They all had two flowers to eat. (ES.Child.14)
All-Two	*They did only two. (ES.Child.15)*
	Because they only eat two. (ES.Child.18)
	They all ate two of them. (ES.Child.24)
	None, two, two. (ES.Child.4)
Other	*Had a sleep (ES.Child.6)*
	Only two used some and one used all of them. (ES.Child.25)
	Some of the crabs leaved and just one didn't (ES.Child.31)

Focusing only on the critical conditions, (79) shows how the conditions with *EverySome* sentences were split between the two sessions. In Session A, 3Some items, which were consistent with all of the targeted interpretations, were presented alongside 3All items, which were only consistent with the *Literal* meaning. This contrast could plausibly have encouraged participants to give more *no*-responses to the 3All items, because they would appear less acceptable *compared to* the 3Some items. In Session B, 3None items, which were consistent with none of the targeted interpretations, were presented alongside 2Some-1All items, which were only consistent with the *Literal* meaning and the *NotEvery* inference. This contrast could have encouraged participants to give more *yes*-responses to 2Some-1All items, because they would seem more acceptable *in comparison to* the 3None items. In sum, the worry would be that by virtue of how the conditions were presented, participants might have been encouraged to give more *no*-responses to 3All items and fewer *no*-responses to 2Some-1All items. This could then lead to a *smaller* difference between the 3None and 3All conditions (because of inflated *no*-responses to the 3All condition),

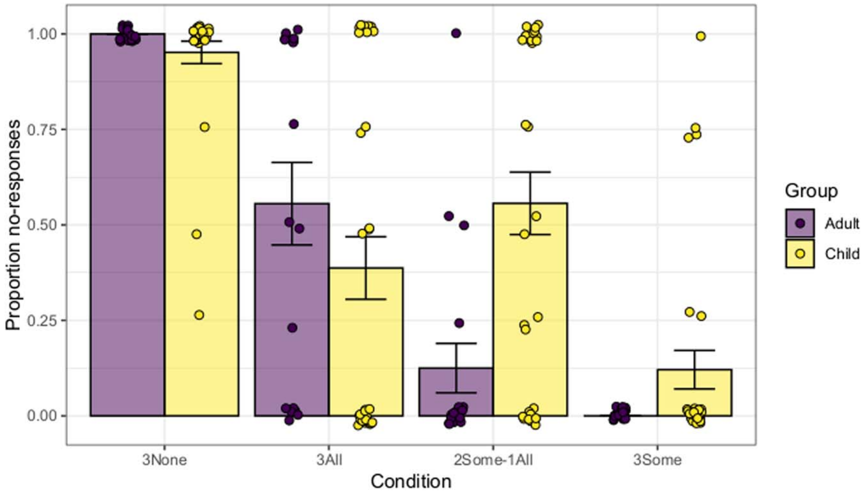


Figure B7 Mean proportion of *no*-responses across test conditions of Experiment 2. The vertical bars represent the standard error. Dots correspond to individual participants' mean rejection rates. A horizontal jitter of.1 and vertical jitter of.025 were applied for better visualization.

and to a *larger* difference between the 3All and 2Some-1All conditions (because of fewer *no*-responses to the 2Some-1All condition).

(79) **Session A:** 3Some, 3All
Session B: 2Some-1All, 3None

Let us now review the relevant results of Experiment 2 (see Figure B7) against the concern outlined above. First, for both groups we found a significant difference between the 3None and 3All conditions, with fewer *no*-responses in the 3All condition. The concern was that the difference between these conditions might have been artificially reduced. Therefore, even if there was such an effect, it evidently was not enough to obscure the difference between the two conditions.

Next, for children we found no significant difference between the 3All and 2Some-1All conditions. Yet on the concern outlined above, participants might have been led to produce *more no*-responses in the 3All condition and fewer *no*-responses in the 2Some-1All condition, thereby increasing the difference between these conditions. Even if there was an effect of a particular presentation of conditions, then, it evidently was not enough to create a significant difference between conditions.

Finally, for adults we found a significant difference between the 3All condition and the 2Some-1All condition. This result is indeed compatible with the concern outlined above, with participants potentially having been encouraged to produce fewer *no*-responses in the 2Some-1All condition. In this case, we cannot rule out the possibility that there was an effect of the particular presentation of conditions; however, the results we obtained are in fact a replication of previous findings reported in Chemla & Spector (2011), a study that did not involve splitting conditions across sessions. We are thus reassured that this result cannot be entirely due to the way that the conditions were presented across sessions.

In sum, while splitting the conditions across sessions is not ideal, even if this design was exerting some effect on participants' responses, it does not substantially alter the findings of Experiment 2.

References

Alonso-Ovalle, L. (2005), 'Distributing the disjuncts over the modal space'. In *Proceedings of the North East Linguistics Society (NELS)*, vol. 35. Graduate Linguistics Students Association (GLSA). Amherst.

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

Barner, D., N. Brooks & A. Bale (2011), 'Accessing the unsaid: the role of scalar alternatives in children's pragmatic inference'. *Cognition* **118**: 84–93.

Barr, D. J., R. Levy, C. Scheepers & H. J. Tily (2013), 'Random effects structure for confirmatory hypothesis testing: keep it maximal'. *Journal of Memory and Language* **68**: 255–78.

Bates, D., M. Mächler & B. Bolker (2015), 'Fitting linear mixed-effects models using lme4'. *Journal of Statistical Software* **67**: 1–48.

Benz, A. & N. Gotzner (2020), 'Embedded implicature: what can be left unsaid?' *Linguistics and Philosophy*. https://link.springer.com/article/10.1007/s10988-020-09310-x.

Berwick, R. (1985), *The Acquisition of Syntactic Knowledge*. Artificial Intelligence Series. MIT Press. Cambridge.

Bowler, M. (2014), 'Conjunction and disjunction in a language without "and"'. In *Semantics and Linguistic Theory*, vol. 24. 137–55.

Breheny, R., N. Klinedinst, J. Romoli & Y. Sudo (2018), 'The symmetry problem: current theories and prospects'. *Natural Language Semantics* **26**: 85–110.

Breheny, R., N. Klinedinst, J. Romoli & Y. Sudo (2017), 'Does the structural approach to alternatives give us just enough alternatives to solve the symmetry problem?' Unpublished manuscript. Available at https://www.ucl.ac.uk/~ucjtudo/pdf/alternatives.pdf.

Brown, R. & D. McNeill (1966), 'The "tip of the tongue" phenomenon'. *Journal of Verbal Learning and Verbal Behavior* **5**: 325–37.

Buccola, B., M. Križ & E. Chemla (2021), 'Conceptual alternatives: competition in language and beyond.' *Linguistics & Philosophy*, (https://link.springer.com/article/10.1007/s10988-021-09327-w#citeas).

Chemla, E. (2009), 'Similarity: towards a unified account of scalar implicatures, free choice permission and presupposition projection.' Forthcoming, (http://www.emmanuel.chemla.free.fr/Material/Chemla-SIandPres.pdf).

Chemla, E. & L. Bott (2014), 'Processing inferences at the semantics/pragmatics frontier: disjunctions and free choice'. *Cognition* **130**: 380–96.

Chemla, E. & R. Singh (2014), 'Remarks on the experimental turn in the study of scalar implicature, part 2'. *Language and Linguistics Compass* **8**: 387–99.

Chemla, E. & B. Spector (2011), 'Experimental evidence for embedded scalar implicatures'. *Journal of Semantics* **28**: 359–400.

Chierchia, G. (2006), 'Broaden your views: implicatures of domain widening and the "logicality" of language'. *Linguistic Inquiry* **37**: 535–90.

Chierchia, G. (2013), *Logic in Grammar: Polarity, Free Choice, and Intervention*. Oxford University Press. Oxford.

Chierchia, G., S. Crain, M. T. Guasti, Gualmini A. & L. Meroni (2001), 'The acquisition of disjunction: evidence for a grammatical view of scalar implicatures'. In *Proceedings of the 25th Boston University Conference on Language Development*. Cascadilla Press. Somerville, MA. 157–68.

Chierchia, G., D. Fox & B. Spector (2011), 'The grammatical view of scalar implicatures and the relationship between syntax and semantics'. In C. Maienborn, K. von Heusinger and P. Portner (eds.), *Semantics: An International Handbook of Natural Language Meaning*, vol. 1. Mouton de Gruyter. New York.

Clifton Jr., C. & C. Dube (2010), 'Embedded implicatures observed: a comment on Geurts and Pouscoulous (2009)'. *Semantics and Pragmatics* **3**: 1–13.

R Core Team (2020), *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Crain, S., W. Ni & L. Conway (1994), 'Learning, parsing and modularity'. In Jr. C. Clifton, L. Frazier and K. Rayner (eds.), *Perspectives on Sentence Processing*. Lawrence Erlbaum Associates. Hillsdale, NJ. 443–67.

Crain, S. & R. Thornton (1998), *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics*. Language, Speech, and Communication. MIT Press. Cambridge.

Cremers, A., F. Kane, L. Tieu, L. Kennedy, Y. Sudo, Folli R. & J. Romoli (2018), 'Testing theories of temporal inferences: evidence from child language'. *Glossa: A Journal of General Linguistics* **3**: 139.

Foppolo, F., M. T. Guasti & G. Chierchia (2012), 'Scalar implicatures in child language: give children a chance'. *Language Learning and Development* **8**: 365–94.

Fox, D. (2007), 'Free choice and the theory of scalar implicatures'. In U. Sauerland and P. Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*. Palgrave Macmillan UK. London. 71–120.

Fox, D. & R. Katzir (2011), 'On the characterization of alternatives'. *Natural Language Semantics* **19**: 87–107.

Franke, M., F. Schlotterbeck & P. Augurzky (2016), 'Embedded Scalars, preferred readings and prosody: an experimental revisit'. *Journal of Semantics* **34**: 153–99.

Geurts, B. & N. Pouscoulous (2009), 'Embedded implicatures?!?' *Semantics and Pragmatics* **2**: 1–34.

Gotzner, N. & J. Romoli (2018), 'The scalar inferences of strong scalar terms under negative quantifiers and constraints on the theory of alternatives'. *Journal of Semantics* **35**: 95–126.

Grice, H. P. (1975), 'Logic and conversation'. In P. Cole and L. M. J (eds.), *Syntax and Semantics*. Academic Press. New York. 41–58.

Grice, H. P. (1978), 'Further notes on logic and conversation'. In P. Cole (ed.), *Syntax and Semantics: Pragmatics*, vol. 9. Academic Press. New York. 113–27.

Gualmini, A., S. Crain, L. Meroni, Chierchia G. & M. T. Guasti (2001), 'At the semantics/pragmatics interface in child language'. In *Semantics and Linguistic Theory*, vol. 11. Linguistics Society of America. Washington DC. 231–47.

Guasti, M. T., G. Chierchia, S. Crain, F. Foppolo, Gualmini A. & L. Meroni (2005), 'Why children and adults sometimes (but not always) compute implicatures'. *Language and Cognitive Processes* **20**: 667–96.

Hochstein, L., A. Bale, D. Fox & D. Barner (2016), 'Ignorance and inference: do problems with gricean epistemic reasoning explain children's difficulty with scalar implicature?' *Journal of Semantics* **33**: 107–35.

Holm, S. (1979), 'A simple sequentially rejective multiple test procedure'. *Scandinavian Journal of Statistics* **6**: 65–70.

Horn, L. R. (1972), *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, University of California, Los Angeles.

Huang, Y. T. & J. Snedeker (2009), 'Semantic meaning and pragmatic interpretation in 5-year-olds: evidence from real-time spoken language comprehension'. *Developmental Psychology* **45**: 1723–39.

Huang, Y. T., E. Spelke & J. Snedeker (2013), 'What exactly do numbers mean?' *Language Learning and Development* **9**: 105–29.

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

Katzir, R. (2007), 'Structurally-defined alternatives'. *Linguistics and Philosophy* **30**: 669–90.

Klinedinst, N. W. (2007), *Plurality and Possibility*. Ph.D. thesis, University of California, Los Angeles.

Kratzer, A. & J. Shimoyama (2002), 'Indeterminate pronouns: the view from Japanese'. In *Proceedings of the Third Tokyo Conference on Psycholinguistics*. Hituzi Syobo. Tokyo. 1–25.

Lidz, J. & J. Musolino (2006), 'On the quantificational status of indefinites: the view from child language'. *Language Acquisition* **13**: 73–102.

Magri, G. (2011), 'Another argument for embedded scalar implicatures based on oddness in downward entailing environments'. *Semantics and Pragmatics* **4**: 1–51.

Moscati, V., J. Romoli, T. F. Demarie & S. Crain et al. (2016), 'Born in the usa: a comparison of modals and nominal quantifiers in child language'. *Natural Language Semantics* **24**: 79–115.

Musolino, J. (2004), 'The semantics and acquisition of number words: integrating linguistic and developmental perspectives'. *Cognition* **93**: 1–41.

Notley, A., P. Zhou, B. Jensen & S. Crain (2012), 'Children's interpretation of disjunction in the scope of 'before': a comparison of English and Mandarin'. *Journal of Child Language* **39**: 482–522.

Noveck, I. A. (2001), 'When children are more logical than adults: experimental investigations of scalar implicature'. *Cognition* **78**: 165–88.

Pagliarini, E., C. Bill, J. Romoli, Tieu L. & S. Crain (2018), 'On children's variable success with scalar inferences: insights from disjunction in the scope of a universal quantifier'. *Cognition* **178**: 178–92.

Papafragou, A. & J. Musolino (2003), 'Scalar implicatures: experiments at the semantics–pragmatics interface'. *Cognition* **86**: 253–82.

Philip, W. (1991), 'Quantification over events in early universal quantification'. In *Proceedings of the 16th Annual Boston University Conference on Language Development*. Cascadilla Press. Somerville.

Philip, W. (2011), 'Acquiring knowledge of universal quantification'. In J. de Villiers and T. Roeper (eds.), *Handbook of Generative Approaches to Language Acquisition*. Studies in Theoretical Psycholinguistics 41. Springer. Dordrecht. 351–94.

Potts, C., D. Lassiter, R. Levy & M. C. Frank (2016), 'Embedded implicatures as pragmatic inferences under compositional lexical uncertainty'. *Journal of Semantics* **33**: 755–802.

Reinhart, T. (2006), *Interface Strategies. Optimal and Costly Derivations*. MIT Press. Cambridge.

Romoli, J. (2012), *Soft but Strong. Neg-Raising, Soft Triggers, and Exhaustification*. Ph.D. thesis, Harvard University.

Sauerland, U., J. Anderssen & K. Yatsushiro (2005), 'The plural is semantically unmarked'. In S. Kepser and M. Reis (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*. Walter de Gruyter. Berlin, Germany.

Singh, R., K. Wexler, A. Astle-Rahim, Kamawar D. & D. Fox (2016), 'Children interpret disjunction as conjunction: consequences for theories of implicature and child development'. *Natural Language Semantics* **24**: 305–52.

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

Spector, B. (2007), 'Aspects of the pragmatics of plural morphology: on higher-order implicatures'. In U. Sauerland and P. Stateva (eds.), *Presupposition and Implicature in Compositional Semantics*. Palgrave Macmillan UK. London. 243–81.

Stiller, A. J., N. D. Goodman & M. C. Frank (2015), 'Ad-hoc implicature in preschool children'. *Language Learning and Development* **11**: 176–90.

van Tiel, B. & W. Schaeken (2017), 'Processing conversational implicatures: alternatives and counterfactual reasoning'. *Cognitive Science* **41**: 1119–54.

Tieu, L., C. Bill, J. Romoli & S. Crain (2020), 'Testing theories of plural meanings'. *Cognition* **205**: 104307.

Tieu, L., J. Romoli, P. Zhou & S. Crain (2016), 'Children's knowledge of free choice inferences and scalar implicatures'. *Journal of Semantics* **33**: 269–98.