When more is less: Children’s interpretation of sentences containing multiple scalar terms

Cory Bill1,*, Elena Pagliarini2, Jacopo Romoli1,3, Lyn Tieu1,4, and Stephen Crain1

1ARC Centre of Excellence in Cognition and its Disorders, Macquarie University
2Center for Brain and Cognition, Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra
3School of Communication, Ulster University
4Western Sydney University
*Corresponding author: Level 3, Australian Hearing Hub, 16 University Avenue, Macquarie University, NSW, Australia, 2109; corybill@gmail.com

Abstract

Previous research has revealed considerable variability in children’s performance when computing scalar inferences (Noveck 2001; Papafragou and Musolino 2003; Guasti et al. 2005, among others). One promising explanation for this variability attributes it to differences in children’s ability to generate the alternatives from which these inferences are derived (Reinhart 2006; Barner et al. 2011; Singh et al. 2016; Tieu et al. 2016). According to this ‘Alternatives-based’ approach, children’s difficulties are due to processing limitations or limitations in lexical knowledge. These limitations are only expected to affect the generation of alternative sentences that require lexical replacement. Therefore, children are predicted to readily generate alternative sentences (and associated scalar inferences) based on material from other sources, for example, those based on sub-constituents of the asserted sentence. This paper investigates children’s behaviour with one such sentence, where an existential expression is in the scope of a universal quantifier (e.g., Every pig carried some of his rocks). Such sentences have been associated with two scalar inferences: the inference that Not every pig carried all of his rocks, and the stronger inference that None of the pigs carried all of his rocks (Chemla and Spector 2011, among others). The Alternatives-based approach predicts that children will be successful in deriving at least one of these inferences. The results of our experiment are consistent with this prediction, with children deriving inference-based interpretations at adult-like rates. We also observe that adults and children differ in which of the two possible inferences is preferred. We discuss how this pattern is consistent with the hypothesis that children favour strong interpretations (Crain et al., 1994).
1 Introduction

The sentence in (1) is often interpreted along the lines of (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’. The traditional account of scalar inference derivation, as presented in Grice (1975), proposes that scalar inferences 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.

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 assertions like (5) (henceforth referred to as ‘EverySome’ sentences) are associated with ‘NotEvery’ inferences like (6), and ‘None’ inferences like (7) (Fox, 2007; Chemla, 2009; Chemla and 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 and Pouscoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; Potts et al., 2016; Gotzner and Romoli, 2018). In general, this work has found that while adults derive both the NotEvery and
None inferences to some extent, they derive the NotEvery inference at higher rates. To our knowledge, no previous work has examined whether children’s interpretations of sentences like (5) include either of these inferences.

However, a great deal of work has investigated children’s behaviour with ‘simpler’ scalar inferences like (2). The early work in this area consistently found that children struggled to compute these scalar inferences, especially when compared with adults (Noveck, 2001; Chierchia et al., 2001; Papafragou and Musolino, 2003; Guasti et al., 2005; Foppolo et al., 2012). However, this 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 meanings analysed as scalar inferences, it has been reported that there are a handful of scalar inferences that children compute quite readily (Papafragou and Musolino, 2003; Barner et al., 2011; Tieu et al., 2016; Hochstein et al., 2016; Pagliarini et al., 2018). Any explanation of children’s behaviour with scalar inferences needs to account for this variability.

One family of explanations, which we will call the ‘Alternatives-based approach’, is particularly successful in accounting for this variability (Chierchia et al., 2001; Gualmi et al., 2001; Reinhart, 2006; Barner and Bachrach, 2010; Barner et al., 2011; Tieu et al., 2016; Singh et al., 2016). This approach proposes that children’s difficulties with scalar inferences are a result of certain limitations that affect their ability to generate alternative sentences. Notably, the limitations proposed by this approach are only expected to apply to alternative sentences generated through lexical replacement. Therefore, this approach predicts that children will have an easier time generating alternative sentences (and their associated scalar inferences) through sources other than the lexicon. Notably, many of the inferences that children have been found to derive more readily involve alternatives that correspond to sub-constituents of the assertion.

The observed variation in children’s performance with scalar inferences and the Alternatives-based approach to this variation intersect quite well with recent theoretical work on the different sources of alternative sentence generation. The question of how a sentence like (4) comes to be identified as an alternative sentence to an assertion like (1) is an issue that has been the focus of ongoing debate in the theoretical literature. The traditional perspective, originally posited by Horn (1972), is that certain lexical items exist on abstract ‘lexical scales’, ordered by informational strength (e.g., <some, all>, <or, and>, <might, must>). Replacing one item with one of its scale-mates is what creates the relevant alternative for inference computation. Recently, more general alternative sentence-generating algorithms have been proposed (Katzir, 2007; Fox and Katzir, 2011; Trinh and Haida, 2015; Breheny et al., 2016). The details of these accounts are not relevant for our purposes, but there is a distinction proposed in this literature, in particular by Katzir (2007) and Fox and Katzir (2011), between two sources of alternative sentences, which aligns remarkably well with children’s variable performance deriving scalar inferences. Specifically, the proposal is that alternative sentences can be generated by replacing parts of the asserted sentence with expressions in the lexicon, or they can be derived by simply substituting constituents of the sentence with other sub-constituents.
of the same sentence. Considering children’s behaviour in light of this proposed distinction, the data would suggest that children are able to generate alternatives from sub-constituents of the assertion, but not from the lexicon.

As we will discuss in further detail, the EverySome sentence in (5) contains within it (i.e. as sub-constituents) the linguistic material required to compose the alternative sentences associated with the two identified scalar inferences (i.e. the NotEvery inference in (6) and the None inference in (7)). As a result, the Alternatives-based approach predicts that children will successfully derive at least one of these inferences, when presented with an EverySome sentence like (5). It is worth noting, however, that the Alternatives-based approach does not make any particular predictions about which of the two inferences children might derive or prefer. Therefore, in addition to testing the predictions of the Alternatives-based approach, investigating children’s interpretations of EverySome sentences will also provide an opportunity to investigate whether children’s interpretive preferences align with those of adults.

As will become clear, our experimental investigation of adults’ and children’s interpretation of EverySome sentences reveals that adults and children do indeed access an inference-based interpretations at a similar rate. However, we also find that the two groups do not align in their interpretive preferences: while children tend to derive the None inference, adults tend to derive the NotEvery inference. We will discuss how this striking difference might connect with theories that suggest children initially prefer stronger meanings when faced with ambiguous sentences (Berwick, 1985; Crain et al., 1994; Gualmini and Schwarz, 2009).

The rest of this paper is structured as follows. First, we review the previous developmental literature on scalar inferences, including recent proposals that attempt to capture children’s behaviour. Next, we outline the predictions of the Alternatives-based approach as they relate to EverySome sentences. We then present the details and results of our experiment, designed to investigate children’s interpretations of EverySome sentences. We conclude by discussing how these results bear on the Alternatives-based approach, and present some possible explanations for the difference we find between adults’ and children’s behaviour.

1.1 Children’s scalar inferences

Over the past few decades, a great deal of research has focused on children’s ability to compute scalar inferences. The first wave of such investigations consistently produced results suggesting that children struggled to derive scalar inferences, as compared to adults (Noveck 2001; Papafragou and Musolino 2003; Huang and Snedeker 2009; Foppolo et al. 2012, among others). For example, Noveck (2001) used a ‘reasoning scenario’ to investigate children’s behaviour with a number of different scalar inferences. For one of these inferences, participants were presented with sentences like (8) and asked whether they agreed

---

1 A third source of alternatives for the approach above corresponds to alternatives made salient in the context. Moreover, alternatives arising from a combination of these different sources is possible.
with them or not. Based on world knowledge, the associated OnlySome scalar 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 it 11% (8 y/o)-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. The bulk of results produced by subsequent experimental studies over the following decade produced similar results (Chierchia et al., 2001; Papafragou and 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 these investigations continued, and particularly as studies started to include a wider range of scalar inferences, a different behaviourial pattern began to emerge. Namely, it was found that there was a handful of inferences that children seemed to derive readily (Papafragou and Musolino, 2003; Barner and 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 interpretations of sentences like (10), to determine whether they included free choice inferences like (11), an inference that is commonly analysed as a scalar inference (Kratzer and Shimoyama 2002; Alonso-Ovalle 2005; Fox 2007; Klinedinst 2007; Chemla 2009; though see Zimmermann 2000; Geurts 2005 and Barker 2010 for alternative analyses).

(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. used a Truth Value Judgment Task (Crain and Thornton, 1998) and conducted their experiment with 22 Mandarin-speaking 3- to 4-year-old children. Tieu et al. 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 may only 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 a more traditional scalar inference (the exclusivity inference from disjunctive sentences) at a typically low rate (18%), they derived free choice inferences, like (11), at a much higher rate (91%).
Similar results were produced in a study by Hochstein et al. (2016), which investigated children’s interpretations of sentences like (12), to see if they would include associated ignorance inferences in (13). Specifically, Hochstein et al. found evidence of 5-year-old children deriving the ignorance inference at a much higher rate (≈76%) than the exclusivity inference (≈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.

Moreover, a study conducted by Stiller et al. (2015) tested whether children would derive ad-hoc inferences from sentences like (14). Stiller et al. presented sentences like (14) to 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 resulted in 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%.

(14) My friend has glasses.
(15) My friend does not have a hat.

Recently, 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).

(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 at the same rate as adults, approximately 55% of the time.

A series of studies have also reported evidence of children readily deriving an exactly-n interpretation of numerals (i.e. “one” as meaning exactly one) (Papafragou and Musolino, 2003; Barner and Bachrach, 2010; Huang et al., 2013), a meaning that is also proposed to be computed as a scalar inference.

Finally, a study by Barner et al. (2011) found evidence of 4-year-old children accessing exhaustive interpretations of certain sentences containing ‘only’, an interpretation that is thought to be accessed through a similar process to scalar inferences.

In sum, while earlier studies found that children struggled to derive the target scalar inferences, more recently, there has been a small but growing number of studies that have found that children 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 variable success in computing scalar inferences. These explanations can be categorised into two different groups based on the nature of the limitation they attribute to children. We will consider each of these in turn.

1.2.1 The Pragmatics-based approach

The first group of explanations attributes children’s performance on scalar inferences to limitations in their pragmatic knowledge (Noveck, 2001; Katsos and Bishop, 2011; Skordos and Papafragou, 2016). For example, Katsos and Bishop (2011) propose that children are more pragmatically tolerant than adults. According to this hypothesis, children can compute scalar inferences, but they are less likely than adults to reject sentences whose inferences are false. In support of this proposal, Katsos et al. present results from an experiment in which children were asked to judge the acceptability of a target sentence using a 3-point scale, rather than the traditional binary, true/false response options. Children in this experiment tended to use the intermediate judgment option when judging target sentences whose literal content was true, but whose scalar inference was false. Katsos et al. thus suggest that children’s previously documented poor performance in scalar inference computation was a result of the response measures used, rather than the result of children failing to have computed the relevant inferences.

Similarly, Skordos and Papafragou (2016) propose that children have a limited understanding of which alternative sentences are relevant in a given context (see Noveck 2001 for similar ideas). Skordos and Papafragou support this proposal with experimental results showing that when the relevance of alternatives is made more salient, children more readily derive scalar inferences.

While these Pragmatics-based explanations can account for why children seem to struggle with certain scalar inferences, these accounts face a challenge in the observation that there exist cases where children readily derive scalar inferences. This is because the limitations attributed to children by this approach should be expected to affect children’s derivation of scalar inferences consistently, rather than affecting only certain inferences and not others. As a result, this approach fails to explain why, for example, Tieu et al. (2016) found children deriving free choice inferences significantly more than exclusivity inferences, despite the contexts being equivalent in the relevant respects.

One option for such accounts would be to adopt an alternative, non-scalar inference analysis for the inferences that children readily derive. In fact, such a strategy is entertained in the paper by Papafragou and Musolino (2003) to explain the high rate at which children computed an exactly-n interpretation of numerals. However, adopting such a strategy means abandoning the gains in parsimony achieved by explaining the derivation of so many, seemingly disparate, inferences through the one scalar inference 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. The group of explanations we will now present avoid such difficulties, as they are able to explain children’s variable behaviour while retaining a unified analysis of scalar inference phenomena.

1.2.2 The Alternatives-based approach

The Alternatives-based approach attributes children’s variable performance on scalar inferences to limitations that affect children’s ability to generate the alternative sentences involved (Chierchia et al., 2001; Gualmini et al., 2001; Reinhart, 2006; Barner and Bachrach, 2010; Barner et al., 2011; Tieu et al., 2016; Singh et al., 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 proposes that children’s low rate of scalar inference derivation is a result of limitations in their processing capacity, which could for example impact on their ability to retrieve scalar alternatives (Chierchia et al., 2001; Gualmini et al., 2001; Reinhart, 2006; Tieu et al., 2016). This proposal contends that children do not possess the processing resources (e.g., working memory capacity) necessary to carry out the generation of alternative sentences, or to compare alternative sentences to the asserted sentence. Both of these steps are necessary to the derivation of scalar inferences.

Another variant of the Alternatives-based approach contends that children’s behaviour is the result of limitations in their knowledge of abstract lexical scales which, on some accounts (e.g., Horn 1972), is necessary to successfully generate many of the alternative sentences from which scalar inferences are derived (Barner and Bachrach, 2010; Barner et al., 2011). On this variant of the approach, children have not fully learned the lexical scales required to compute scalar inferences.

Both of these proposals lead us to expect that children will experience difficulties deriving many scalar inferences. However, if we assume that in all other relevant respects children’s abilities are adult-like, then at least some versions of the Alternatives-based approach also predict that children will, in certain cases, derive scalar inferences successfully. Consider the proposal that children’s difficulties are a result of limitations in their lexical-scale knowledge. This proposal should predict that children’s difficulties will be restricted to scalar inferences that require this knowledge in order to be derived. On the other hand, there is no reason on this approach why children should not be more successful in deriving scalar inferences that do not require this lexical knowledge, for example, if the content from which the alternative sentences are composed is presented as a sub-constituent of the assertion.

As for proposals appealing to processing limitations (e.g., Reinhart 2006), the expectations are not so straightforward. Taken at face value, such proposals might appear to predict that children’s difficulties deriving scalar inferences would be uniform, irrespective of how the alternatives are generated. However,
if we assume that generating alternative sentences through accessing the lexicon is more demanding than generating them through other sources, then these proposals would also lead us to expect some variation in children’s performance with scalar inferences. In fact, there are results from the adult sentence processing literature suggesting that the cost of processing scalar inferences varies precisely along these lines. Specifically, Chemla and Bott (2014) present evidence that adults are faster at computing free choice inferences, which involve sub-constituent alternatives, than at computing OnlySome inferences, which involve lexical alternatives.

In the rest of this paper, when we refer to the ‘Alternatives-based approach’, we are referring to the versions of it that predict children will more easily derive scalar inferences involving alternative sentences that are not generated through lexical replacement of alternatives.

In support of this, the studies that have found evidence of children successfully deriving scalar inferences have generally involved sentences where the alternative sentences are generated from sub-constituents of the assertion. For example, free choice inferences like (20), which children derive readily from assertions like (19) (Tieu et al. 2016), are derived from such alternative sentences. 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 assertion. Similarly, it has been found that children readily derive ignorance inferences (Hochstein et al., 2016) and distributive inferences (Pagliarini et al., 2018), which also both involve alternatives that can be composed from content presented in the assertion.

(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.

The ideas presented by the Alternatives-based approach align remarkably well with work in the theoretical literature by Katzir (2007) and Fox and Katzir (2011) on the distinct sources of alternatives. 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.

In sum, the Alternatives-based approach is able to explain both children’s typical difficulties with scalar inference computation, as well as many of the cases in which they succeed. Moreover, the Alternatives-based approach makes the following prediction regarding the influence of the linguistic context on children’s behaviour with scalar inferences. If the linguistic material from which a scalar inference’s alternatives are composed is presented within the asserted sentence,
then children will successfully derive that scalar inference. This prediction leads to some interesting and counter-intuitive expectations, which we turn to now.

1.3 Testing the Alternatives-based approach

As mentioned, previous studies such as Noveck (2001) revealed that when presented with sentences like (23), children tend not to derive the associated OnlySome inference in (24). According to the Alternatives-based approach, this is because children experience difficulties generating the alternative sentence in (25), since it requires retrieving the relevant scale-mate ‘all’ from the lexicon.

(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.

Sentences like (23) can be placed in the scope of a universal quantifier as in (26), to create what we will call an EverySome sentence. Such EverySome sentences 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 like (26) (Geurts and Pouscoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; Potts et al., 2016). This work has revealed that adults derive both of these inferences to some extent, with an apparent preference for the NotEvery inference over the None inference.

Given children’s difficulties deriving OnlySome inferences from the simpler sentence in (23), it is perhaps unsurprising that no previous work has investigated children’s ability to derive the inferences associated with EverySome sentences. However, although these sentences are structurally more complex than sentences like (23), in that the scalar term is embedded under a universal quantifier, the Alternatives-based approach leads us to expect that the associated inferences in (27) and (28) may in fact be easier for children than the OnlySome inference.

In Appendix A, we provide details of how the OnlySome inference and the inferences in (27) and (28) are derived. What is crucial for our purposes here is that while the alternative sentence required to derive the OnlySome inference involves the lexical replacement of ‘some’ with ‘all’, the two inferences in (27) and (28) can be generated from content retrievable from the asserted sentence in (26). The NotEvery inference in (27) is derived through the negation of the alternative sentence in (29), which is generated by replacing the existential quantifier ‘some’ in the asserted sentence with the universal quantifier ‘all’.
Every pig carried all of his rocks.

Since, in this case, the universal quantifier is provided explicitly in the asserted sentence (as the subject quantifier), the Alternatives-based approach predicts that children will be facilitated in generating the alternative in (29), and by extension, the scalar inference in (27). Generating (28), on the other hand, involves exhaustifying at the embedded level, with ‘some of his rocks’ being compared to the alternative ‘all of his rocks’; Appendix A.3 provides details of the derivation, but what is relevant for our purposes is that here too, the universal quantifier is retrievable from the subject quantifier of the asserted sentence.

The prediction of the Alternatives-based approach is therefore that while children might struggle to compute the OnlySome inference from simple ‘some’ sentences, they should actually be successful in generating one of the inferences associated with the more complex EverySome sentences. This is because the alternatives required for the NotEvery and None inferences are retrievable from the asserted sentence. To put it another way, using Fox and Katzir (2011)’s model of alternatives, children are expected to more easily derive the NotEvery and None inferences because they involve alternatives generated from subconstituents of the assertion, as opposed to the OnlySome inference, which involves alternatives generated from the lexicon.

Note that there is a key difference between EverySome sentences and previous cases where children have been found to successfully derive inferences. The sentences investigated in previous studies, such as free choice inferences and ignorance inferences, contain the relevant linguistic material as part of a larger constituent, for example, one of the disjuncts of the asserted sentence (Hochstein et al., 2016; Tieu et al., 2016; Singh et al., 2016; Pagliarini et al., 2018). In contrast, while the EverySome sentence does provide the critical linguistic material (i.e. the universal quantifier), a greater level of reconstruction is required to formulate the alternative sentences. To put it another way, it needs to be assumed that the content presented in the assertion can be ‘recycled’ and used in the generation of alternative sentences. For this reason, the current paper is an extension of this previous work, as it investigates whether only presenting this key linguistic material is sufficient to enable children to overcome their limitations.

1.4 Children’s inferential preferences

As we have just outlined, the Alternatives-based Approach predicts that children will be successful in accessing one of the two inference-based interpretations of EverySome sentences. However, it does not predict which of these inferences might be preferred. Previous work investigating adults’ interpretations of EverySome sentences found that adults preferred interpretations involving the NotEvery inference over those involving the None inference (Geurts and Pouscoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; van Tiel,

2Note that we are assuming that at the appropriate level of representation for constructing scalar alternatives, ‘every’ and ‘all’ are interchangeable.
Children’s interpretations of these sentences have not, to our knowledge, been previously investigated. Therefore, in addition to testing the predictions of the Alternatives-based approach, our experiment will also explore whether children’s interpretive preferences align with those of adults.

2 Experiment

The aim of our experiment was to investigate children’s derivation of the NotEvery inference and the None inference, associated with EverySome sentences. We designed the experiment to identify and compare adults’ and children’s interpretations of EverySome sentences. First, we wanted to know whether, as the Alternatives-based approach predicts, children would access an inference-based interpretation at the same rate as adults. Second, we wanted to investigate whether children would follow adults in preferring interpretations involving the NotEvery inference over those involving the None inference.

2.1 Method

2.1.1 Participants

Eighteen monolingual English-speaking undergraduate students (14 female) and 31 monolingual English-speaking children (4;00-5;10, M = 4;05, 20 female) participated in the experiment. The adults took part in the experiment for course credit, or for a payment of 15AUD. All participants were recruited and tested in Sydney. 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 20AUD for travel expenses.

2.1.2 Procedure

The experiment took the form of a Truth Value Judgment Task (Crain and Thornton, 1998). This task involves two experimenters. One experimenter presents a story, and the other operates a puppet who watches the stories along with the participant. Once a story is finished, 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 asks the participant to provide a justification (i.e. “Why

---

3As we outline in Section 2.1.4, 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.

4Note that for adult participants the puppet’s sentences were pre-recorded and presented to participants at the appropriate times via a laptop. Child participants were given the traditional live presentation of the sentences via a second experimenter.
do you think <Puppet’s name> is wrong?"/“What really happened?”). Justifications were not requested for yes-responses, as such requests were deemed to be infelicitous and could potentially confuse child participants (Crain and Thornton, 1998). Adults saw the same materials; however, they were asked to write down their judgments and justifications, and to include 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 write out a justification for a ‘no-judgment’.5 Children were tested individually, either in the lab or in a quiet room at their daycare. Adults were tested either individually or 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.

2.1.3 Materials
The experiment included four test conditions and two control conditions.

Test conditions
There were four items for each of the test conditions. Each of these items 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 took 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. Once all the characters had gone through this process, the experimenter asked the puppet what had happened in the story, to which the puppet responded with the test sentence.

There were three possible interpretations of our test sentences that participants might access. We refer to these as the Literal, NotEvery, and None interpretations. In the case of a test sentence like (30), the Literal interpretation can be paraphrased as in (31), which corresponds to the basic truth conditions of the sentence, without any inferences. The corresponding NotEvery interpretation can be paraphrased as in (32), and corresponds to the Literal interpretation enriched with the NotEvery inference. Finally, the relevant None interpretation can be paraphrased as in (33), and corresponds to the Literal interpretation enriched with the None inference.

(30) Every pig carried some of his rocks.

(31) **Literal interpretation:** Every pig carried at least one of his rocks.

(32) **NotEvery interpretation:** Every pig carried at least one of his rocks & Not every pig carried all of his rocks.

5We were less concerned that children would be influenced in the same way, on the assumption that communicating justifications verbally is less effortful (and generally takes less time) than writing them down.


<table>
<thead>
<tr>
<th>Test Conditions</th>
<th>Context</th>
<th>Target Interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>2/4, 2/4, 2/4</td>
<td>Literal, NotEvery, None</td>
</tr>
<tr>
<td>ONEINERENCE</td>
<td>2/4, 2/4, 4/4</td>
<td>Literal, NotEvery, None</td>
</tr>
<tr>
<td>NOINERENCE</td>
<td>4/4, 4/4, 4/4</td>
<td>Literal, NotEvery, None</td>
</tr>
<tr>
<td>FALSE</td>
<td>0/4, 0/4, 0/4</td>
<td>Literal, NotEvery, None</td>
</tr>
</tbody>
</table>

Table 1: Contexts and target interpretations made true in each test condition.

(33) **None interpretation:** Every pig carried at least one of his rocks & None of the pigs carried all of his rocks.

These target interpretations stand in certain entailment relations, as can be seen in Table 1. In the FALSE condition none of the characters acted on any of their objects. Therefore, the FALSE condition was inconsistent with all target interpretations. In contrast, in the TRUE condition, each character acted on some but not all of their objects. Therefore, the TRUE condition was consistent with all three of the target interpretations. In the NOINERENCE condition, every character acted on all of his objects, making this condition consistent with the Literal interpretation, but inconsistent with both the NotEvery interpretation and the None interpretation. Finally, in the ONEINERENCE condition, two characters acted on two of their four objects, and one character acted on all four of their objects. Therefore, the ONEINERENCE condition was consistent with the Literal interpretation and the NotEvery interpretation, but was inconsistent with the None interpretation.

By comparing participants' responses across the four test conditions, we could determine the different interpretations participants were accessing in the following way. A difference in a group's target sentence rejection rates between the FALSE and the NOINERENCE conditions was evidence that participants were accessing a Literal interpretation. A difference in a group's target sentence rejection rates between the NOINERENCE and ONEINERENCE conditions was evidence that participants were accessing a NotEvery interpretation. Finally, a difference in a group's target sentence rejection rates between the ONEINERENCE and TRUE conditions was evidence that participants were accessing a None interpretation.

We were also interested in whether adults and children would produce similar interpretation preferences for these sentences. We could directly investigate this by testing for any differences between the groups' responses within each of our test conditions.

To illustrate, one of the items from the ONEINERENCE condition is presented in (34). This story would be associated with the test sentence in (35). The final scene of this story is shown in Figure 1.

(34) **Example ONEINERENCE condition item.**

*This is a story about three pigs [see Figure 1a]. These pigs each have rocks that they can carry if they want to. Let’s see what they do:*
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? Yes I will, as I am tired.

Experiment: Ok <Puppet’s name>, what happened in that story?

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 (36) either in an Every_TRUE context that made the sentence true, or in an Every_FALSE context that made the sentence false, as outlined in Table 2.

(36) **Sentence:** Every pig carried rocks.

In addition to the test and control conditions, participants also received five filler items. The filler items were designed so that they could each be paired with one of two possible sentences. One was designed to elicit a no-response, while the other was designed to elicit a yes-response. The sentence that was ultimately used was chosen based on a participant’s responses to previous trials, so that we could reduce the occurrence of 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.

### 2.1.4 Design

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 (37).

(37) a. **Session A:** TRUE, NOInference, Every_TRUE, Every_FALSE  
b. **Session B:** ONEInference, FALSE

The ordering of the items within each session 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 each session was 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.

### 2.2 Results

In order to be included in the data analysis, participants needed to meet two inclusion criteria. First, participants needed to correctly answer at least 4 of

<table>
<thead>
<tr>
<th>Control Conditions</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every_TRUE</td>
<td>2/4, 2/4, 4/4</td>
</tr>
<tr>
<td>Every_FALSE</td>
<td>2/4, 2/4, 0/4</td>
</tr>
</tbody>
</table>

Table 2: Contexts that made the Every control sentences true and false, respectively.
Figure 2: Mean proportion of test sentence acceptances across our test conditions. The vertical bars represent the standard error. Each dot represents one participant’s mean acceptance rate for that condition. A horizontal jitter of .1 and vertical jitter of .025 were applied for better visualisation.

the 5 filler items. This was to ensure that participants understood the task, and were paying attention to the stories and target sentences. All of the adult and child participants met this criterion. Second, participants were required to correctly answer at least 3 of the 4 items in the EVERY_TRUE and EVERY_FALSE control conditions. This was to ensure that participants understood the basic meaning of the universal quantifier ‘every’. All of the adult participants met this criterion, however, 10 of the children did not, and so were excluded from the final data-set. In all, 18 adults and 21 children were included in the data analysis.

Participant responses came in the form of judgments and justifications. We will consider each of these in turn.

### 2.2.1 Judgments

The judgment results are presented in Figure 2. 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.

These results lead us to the following conclusions regarding how participants...
Table 3: Results of Wilcoxon signed-rank tests comparing differences between conditions within each group. *Significant at $\alpha = .05$, based on the Holm-Bonferroni correction procedure.

<table>
<thead>
<tr>
<th>Group</th>
<th>Comparison</th>
<th>Test statistic</th>
<th>Effect size</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>FALSE vs. NOINF.</td>
<td>$Z = -3.6$</td>
<td>$r = -.56$</td>
<td>$p = .003^*$</td>
</tr>
<tr>
<td></td>
<td>NoINF. vs. ONEINF.</td>
<td>$Z = -1.98$</td>
<td>$r = -.31$</td>
<td>$p = .048$</td>
</tr>
<tr>
<td></td>
<td>ONEINF. vs. TRUE</td>
<td>$Z = -3.21$</td>
<td>$r = -0.5$</td>
<td>$p = .001^*$</td>
</tr>
<tr>
<td>Adult</td>
<td>FALSE vs. NOINF.</td>
<td>$Z = -2.82$</td>
<td>$r = -0.47$</td>
<td>$p = .005^*$</td>
</tr>
<tr>
<td></td>
<td>NoINF. vs. ONEINF.</td>
<td>$Z = -2.92$</td>
<td>$r = -0.49$</td>
<td>$p = .004^*$</td>
</tr>
<tr>
<td></td>
<td>ONEINF. vs. TRUE</td>
<td>$Z = -1.66$</td>
<td>$r = -0.28$</td>
<td>$p = .098$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Condition</th>
<th>Comparison</th>
<th>Test statistic</th>
<th>Effect size</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FALSE</td>
<td>Child vs. Adult</td>
<td>$W = 171$</td>
<td>$r = -0.21$</td>
<td>$p = .197$</td>
</tr>
<tr>
<td>NoINF.</td>
<td>Child vs. Adult</td>
<td>$W = 139.5$</td>
<td>$r = -0.24$</td>
<td>$p = .137$</td>
</tr>
<tr>
<td>ONEINF.</td>
<td>Child vs. Adult</td>
<td>$W = 277$</td>
<td>$r = -0.44$</td>
<td>$p = .006^*$</td>
</tr>
<tr>
<td>TRUE</td>
<td>Child vs. Adult</td>
<td>$W = 207$</td>
<td>$r = -0.21$</td>
<td>$p = .197$</td>
</tr>
</tbody>
</table>

engaged with the target interpretations in (38)-(40).

(38) **Literal interpretation**: Every pig carried at least one of his rocks.
(39) **NotEvery interpretation**: Every pig carried at least one of his rocks & Not every pig carried all of his rocks.
(40) **None interpretation**: Every pig carried at least one of his rocks & None of the pigs carried all of his rocks.

We observed that both children and adults displayed a difference in target sentence acceptances between the FALSE and NOINFEERENCE conditions, providing evidence that both groups accessed the Literal interpretation in (38). Adults displayed a difference between the NOINFEERENCE and ONEINFEERENCE conditions, suggesting they accessed the NotEvery interpretation in (39); children did not display this difference, and thus we have no evidence that they accessed the NotEvery interpretation. Children did, however, display a difference between the ONEINFEERENCE and TRUE conditions, providing evidence that they accessed the None interpretation in (40). Adults, on the other hand, did not display this difference; thus we do not have any evidence that adults accessed the None interpretation.

Comparing target sentence acceptances between adults and children across each condition (Table 4), we only found a difference between adults and children in the ONEINFEERENCE condition.
Table 5: Expected responses on each target interpretation.

<table>
<thead>
<tr>
<th></th>
<th><strong>NoInference condition</strong></th>
<th><strong>OneInference condition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Literal</td>
<td>accept</td>
<td>accept</td>
</tr>
<tr>
<td>NotEvery</td>
<td>reject</td>
<td>accept</td>
</tr>
<tr>
<td>None</td>
<td>reject</td>
<td>reject</td>
</tr>
</tbody>
</table>

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

To explore the data further, we turned to the response patterns of individual participants, to see if we could identify which of the target interpretations each participant had accessed. To do this, we first identified the type of response each participant tended to produce in the NoInference and OneInference conditions. For example, if a participant accepted at least 3 of the 4 items in the NoInference condition, then they were categorised as ‘accepting’ the test sentences in that condition. Following this, we identified how many participants displayed the combination of responses expected for each target interpretation, as laid out in Table 5. For example, if a participant was categorised as accepting test sentences in both the NoInference and the OneInference conditions, they were categorised as having accessed a Literal interpretation. Twelve participants (8 children and 4 adults) provided mixed responses in at least one of these test conditions, and so they were not included in any of the interpretation categories.

For the 27 participants (13 children and 14 adults) who did provide consistent responses, Figure 3 presents the results of this categorisation. As Figure 3 shows, both groups were similar in the rates at which they accessed a Literal interpretation of our target sentences. However, they differed in regards to the inference-based interpretation they accessed, with children only accessing a None interpretation, and adults tending to access a NotEvery interpretation.

The individual response pattern data provide further corroborating evidence that children primarily accessed a None interpretation, while adults primarily accessed a NotEvery interpretation.

2.2.2 Justifications

We also elicited and recorded justifications for all of our participants’ no-judgments. A number of considerations affected our analysis of this data. First, we focused
Literal interpretation: Every pig carried at least one of his rocks.

NotEvery interpretation: Every pig carried at least one of his rocks & Not every pig carried all of his rocks.

None interpretation: Every pig carried at least one of his rocks & None of the pigs carried all of his rocks.

First we will consider the justifications children provided for their rejections of test sentences in the OneInference condition. There were 44 justifications of this kind. These justifications were coded into one of three categories. The bulk of children’s justifications (36/44; 82%) focused on the fact that one of the characters had acted on all of their objects (e.g., No, because this dog ate all of them; No, that sheep carried all of the rocks; No, because this one used all of them). Therefore, these justifications support the suggestion that the vast majority of associated judgments are a result of participants accessing the None interpretation in (43). That is, these justifications focus on the aspect of the context that was not consistent with the None interpretation.
The next most frequent type of justification (7/44; 16%) focused on the fact that two of the characters had not acted on all of their objects (e.g., No, because these two didn’t carry all of their rocks, like this one; No, these two didn’t; No, because these two didn’t use all of them). These justifications seem to indicate that the associated justifications were motivated by some alternative interpretation on which there was an expectation that all of the characters would act on all of their items.

Finally, one of the justifications (2%) merely repeated what had happened in the story, and so did not provide any insight into the specific interpretation motivating the associated judgment.

Let us turn next to the justifications that adults provided for their rejections of the test sentences in the NoInference condition. There were 40 justifications of this kind. They fell into two categories. The vast majority of justifications (36/40; 90%) focused on the fact that all of the characters had acted on all of their objects (e.g., No, each of the unicorns polished all of their gemstones; No, every cat threw all of his glowsticks; No, the rabbits used all of their tea-bags to make their pots of tea). The focus these justifications put on the fact that all of the characters acted on all of their objects provides support for the conclusion that the associated judgments were motivated by a NotEvery interpretation. That is, these justifications focused on the aspect of the context that was inconsistent with the NotEvery interpretation in (42). The remaining justifications (4/40; 10%) merely repeated the basic elements of the story, for example mentioning that all of the characters had acted on their objects, without explicitly identifying that they had acted on 'all' of them (e.g., No, all the cats threw glowsticks; No, every rabbit used teabags to make tea; No, each unicorn polished her stones). These justifications did not provide any clear insights into the interpretation participants were accessing.

In sum, a coding analysis of the relevant justifications for each participant group supports the conclusion that the relevant judgments were largely based on participants having accessed the associated inference-based interpretation. For adults, this was the NotEvery interpretation, and for children, it was the None interpretation.

3 Discussion

The aim of our experiment was to investigate children’s interpretations of EverySome sentences. We focused on these sentences because, according to the Alternatives-based approach, they have attributes that are expected to assist children in deriving the associated NotEvery and None inferences. In particular, the relevant alternatives can be generated by retrieving sub-constituents of EverySome sentences, rather than from the lexicon. The Alternatives-based approach contends that children are capable of generating such alternative sentences, and their associated inferences. If children were indeed found to derive these inferences, we were also interested in whether they would mirror adults in preferring interpretations involving the NotEvery inference over those involving...
the None inference. We investigated both these questions by measuring the rates at which children and adults derived NotEvery inferences like (45) and None inferences like (46), from EverySome sentences like (44).

(44) Every pig carried some of his rocks.
(45) Not every pig carries all of his rocks.
(46) No pig carried all of his rocks.

The results of our experiment revealed that children derived an inference-based interpretation of the target EverySome sentences at the same rate as adults, consistent with the predictions of the Alternatives-based approach. Moreover, we observed that children’s inference-based interpretations primarily involved the None inference, whereas the data from our adult participants replicated previous work in displaying a preference for the NotEvery inference (Geurts and Pousoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; Potts et al., 2016). Finally, we would note that this is the first work to shed light on the inferences that children draw from sentences containing multiple quantifiers (i.e. sentences containing quantified noun phrases in both subject and object position). We will now consider the implications of these results in more detail.

3.1 Children’s inference-based interpretations

The finding that children accessed inference-based interpretations of our test sentences at adult-like rates is intriguing when considered in light of the previous literature in this area, in particular, against the previous robust finding that when presented with sentences like (47), children tend not to derive the associated OnlySome inference in (48) (Noveck 2001; Papafragou and Musolino 2003; Guasti et al. 2005; Bill et al. 2016, among others). This is particularly striking, since sentences like (47) are in some sense simpler versions of our test sentences.

(47) The pig carried some of his rocks.
(48) The pig didn’t carry all of his rocks.

The contrast between the previous findings and the present results provides further support for the Alternatives-based approach, according to which children’s variable success on scalar inferences is subject to limitations in their generation of alternative sentences through lexical access. When the alternative sentences are generated from a source children can access (i.e. sub-constituents of the assertion), then, even if the sentence is more complex, children will nevertheless successfully compute the associated inference(s).

Moreover, the fact that children’s behaviour varies in this way lines up well with recent proposals that distinguish between these different sources of alternative sentences (Katzir, 2007; Fox and Katzir, 2011). Specifically, it provides support for the idea that the *lexicon* and *sub-constituents of the assertion* are
genuinely distinct sources of alternative sentence generation, given that children appear only able to generate alternative sentences from the latter source.

In sum, the findings of our experiment provide further support for the Alternatives-based approach as an explanation of children’s variable success on scalar inferences. Moreover, they are consistent with theories positing multiple distinct sources of alternative sentence generation.

3.2 Principles of interpretation

The second aim of this study was to investigate whether children’s pattern of inference-based interpretations would mirror that of adults. Previous research on adults’ interpretations of EverySome sentences revealed that adults prefer the NotEvery interpretation over the None interpretation (Geurts and Pouscoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; Potts et al., 2016). Our experiment replicated this finding. The response patterns of our child participants, however, did not mirror that of adults. Instead, children displayed an overwhelming preference for the None interpretation. This aspect of the results is surprising and it is not immediately obvious why we should expect to observe this difference.

Any attempt at explaining this difference between adults and children needs to be able to account for the fact that, 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 about how adults and children interpret ambiguous sentences.

When presented with sentences that can be associated with multiple meanings, language users need to decide which meaning they should adopt first. There are a number of principles proposed in the theoretical literature that are said 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 and Thornton, 1998). 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 None interpretation, as it entails the other relevant interpretations. The fact that children in our experiment preferred this interpretation is therefore consistent with the Subset Principle. Importantly, adults are not assumed to be guided by the Subset Principle when interpreting ambiguous sentences. The fact that they did not align with children in preferring the stronger None interpretation is not a problem for this principle.\footnote{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. One might therefore 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 are}
Unlike children, adults favored the NotEvery interpretation over the None interpretation. This behavior can be quite straightforwardly accounted for as being motivated by the Principle of Charity, a general principle that co-operative speakers are thought to employ 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 NotEvery interpretations over None interpretations, as the former were true in more of our test conditions. One might 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. It seems plausible that the Principle of Charity becomes relevant in guiding children’s interpretations of ambiguous sentences only once they have acquired knowledge of what the possible interpretations are.

Future work could further investigate the relative role of principles like the Subset Principle and the Principle of Charity in guiding people’s access to inference-based interpretations. We will simply note here that an explanation along the lines of the Subset Principle could conceivably capture the difference we observed between adults’ and children’s inference-based interpretations.

3.3 An alternative explanation for our results

The following argument could be advanced to suggest that, rather than being a result of having derived the None inference, children’s behavior was motivated merely by a desire for the characters in our stories to behave uniformly. Such an argument would take the following form.

The OneInference condition was the only condition where the characters’ behavior was not uniform. That is, in the OneInference condition, while two of the characters acted on 2 of their 4 objects, one character acted on 4 of their 4 objects. In contrast, in every other test condition (i.e. False, NoInference, True) each character acted upon the same number of objects. If, perhaps due to the presence of a universal quantifier, children wanted the characters’ actions to be uniform, then this would also have generated our finding that children rejected the target sentences in the OneInference condition. The pattern of results we associated with derivation of the None inference would then actually have been a result of children requiring the characters’ actions in fact associated with different underlying syntactic structures for the asserted sentence (see Appendices A.2 and A.3 for more information on this point). Appealing to the Subset Principle as an explanation for our findings could therefore tie in quite well with the Grammatical account of scalar inferences.

7 One might also note that the Literal interpretation is, in fact, an even weaker interpretation of EverySome sentences than the NotEvery interpretation. Therefore, if all three target interpretations were under consideration, then the Principle of Charity should have encouraged adults to prefer the Literal interpretation of our test sentences. If we assume, however, that our participants preferred inference-based interpretations over literal ones, then the charity principle would only influence the choice between the two inference-based interpretations.
to be uniform. And in fact, children's justifications for rejecting test sentences in the OneInference condition provides evidence that some participants may have been motivated by such an interpretation. Specifically, the 16% of these justifications focusing on the fact that two of the characters had not acted on all of their objects would appear to be motivated by such an interpretation.

However, we do not think this explanation can account for the bulk of children's behaviour in the OneInference condition for three reasons. First, as already mentioned, in contrast to the OneInference condition, the NoInference condition did present stories where the characters' behaviour was uniform, with each character acting upon the same number of objects (4 out of 4). Therefore, if uniformity considerations were the only reason children rejected the OneInference condition, children should not have also rejected the target sentences in the NoInference condition. Despite this, children rejected test sentences in the NoInference and OneInference conditions at similar rates.

Second, in order to be included in the data analysis, participants were required to correctly respond to at least 3 out of the 4 Every control items. The two Every_True controls presented exactly the same non-uniform context as in the OneInference condition, but were described with a target sentence like (49), which included a universal quantifier in the same position as our test sentences. Therefore, the participants that were included in our analysis had accepted at least one target sentence like (49) as a description of a non-uniform context. If children required characters to be uniform in their behaviour, we would not expect them to have accepted the target sentences in this condition.

(49) Every pig carried rocks.

Finally, the vast majority (82%) of children's justifications for rejecting test sentences in the OneInference condition indicated that they were accessing a None interpretation. That is, their justifications focused on the fact that one of the characters had acted on all of their objects. Therefore, the bulk of relevant justifications are not consistent with this alternative explanation.

Given these considerations, it appears that the most plausible explanation for the bulk of children's behaviour is that children derived a None inference, and not that they accessed some non-adult-like interpretation of the universal quantifier.

3.4 Conclusion

Previous research has reported considerable variation in children's ability to compute scalar inferences. The Alternatives-based approach attributes this variation to certain limitations in how children engage with the alternative sentences involved (Chierchia et al., 2001; Gualmini et al., 2001; Reinhart, 2006; Barner and Bachrach, 2010; Barner et al., 2011; Tieu et al., 2016). However, these limitations are only expected to affect inferences that require access to the lexicon in order to perform lexical replacement of alternatives. Children are expected to be able to derive scalar inferences in cases where the required alternatives correspond to sub-constituents of the asserted sentence. One kind of sentence that
exhibits this property is the EverySome sentence. Two scalar inferences have been associated with such sentences — the NotEvery inference and the None inference (Sauerland, 2004; Chierchia, 2004; Fox, 2007; Chemla, 2009; Geurts and van Tiel, 2013). For both of these inferences, the linguistic material from which the relevant alternative sentences are composed is contained in the asserted EverySome sentence. Therefore, the Alternatives-based approach predicts that children will be successful in accessing an inference-based interpretation of EverySome sentences at an adult-like rate. Our results were consistent with this prediction, providing further support for the Alternatives-based approach as an account of children’s variable success on scalar inferences.

The Alternatives-based approach does not, however, make any predictions about which inference-based interpretation should be preferred by children. Previous literature has found that adults prefer NotEvery inferences over None inferences (Geurts and Pouscoulous, 2009; Clifton Jr and Dube, 2010; Chemla and Spector, 2011; Potts et al., 2016). Our experiment replicated this pattern, with adults preferring NotEvery interpretations over None interpretations. The current study is the first to investigate children’s interpretations of these sentences. We found that, in contrast to adults, children preferred None inferences over NotEvery inferences. We discussed one possible explanation for this difference, according to which the two groups are guided by different interpretation principles. Specifically, adults are guided by the Principle of Charity while children are guided by the Subset Principle. This explanation makes certain predictions regarding how these groups will interpret other kinds of sentences that may be associated with multiple scalar inferences. In particular, assuming children are able to access the required alternatives to generate the relevant inferences, this explanation would predict that children should prefer stronger inferences to weaker ones, while adults might be led by Charity to prefer weaker ones. We leave the investigation of this prediction to future research.

In closing, the present novel investigation of children’s interpretation of sentences containing multiple scalar quantifiers reveals that children are indeed capable of deriving inference-based interpretations when the required alternatives are provided within the asserted sentences themselves.

4 Author’s address

[Removed for review.]

5 Acknowledgements

[Removed for review.]

6 Data availability

Data and R script available at https://semanticsarchive.net/Archive/jM1N2jiN/
A Deriving scalar inferences

In this Appendix, we show how the simple OnlySome inference can be derived. Following this, we outline how the same process is used to account for the derivation of the NotEvery and None inferences. As mentioned, there is ongoing debate regarding the exact mechanism underlying scalar inferences. The main contribution of our study does not hinge on assuming any one particular account, but for the purposes of illustration, we will adopt the ‘Grammatical account’ of scalar inferences (Chierchia, 2006; Fox, 2007; Chierchia et al., 2011). We will first sketch the main ingredients of this account and then show how it can derive the three inferences in turn.

The gist of the Grammatical account is that scalar inferences arise due to the presence of a covert exhaustivity operator $exh$, akin to ‘only,’ which quantifies over a sentence and its relevant alternatives. Intuitively, what $exh$ generally does is strengthen the sentence as much as possible, while avoiding contradictions and arbitrary choices among the sentence’s alternatives. The definition of $exh$ is provided in (50), and the definition of the alternatives that can be negated, the ‘excludable’ alternatives, is provided in (51). In essence, $exh$ takes a sentence and a set of alternatives as its input and outputs the conjunction of that sentence and the negation of a subset of its alternatives. An alternative is excludable if its negation doesn’t contradict the literal meaning of the asserted sentence, and doesn’t force us to accept any other alternative in the set.\(^8\)

\[
\begin{align*}
(50) & \quad [exh \phi](w) = [\phi](w) \land \forall \psi \in EXCL(\phi, ALT(\phi)][\neg[\psi](w)] \\
(51) & \quad EXCL(\phi, X) = \\
& \quad \{\psi \in X : [\phi] \not\subseteq [\psi] \land \neg \exists \chi \in X \land (\exists \psi \in X \land [\phi] \land \neg[\psi]) \subseteq [\chi]\}
\end{align*}
\]

In the following, we turn to show how the ingredients above can derive the OnlySome inference, and, subsequently, the NotEvery and None inferences.

A.1 Deriving the OnlySome inference

Let us consider the sentence in (52), schematised as ‘SOME’. The OnlySome inference in (53) is derived by assuming that $exh$ is now underlyingly present in the sentence above as in (54) and that it quantifies over the alternatives in (55).\(^9\)

\[
\begin{align*}
(52) & \quad \text{The pig carried some of his rocks.} \\
(53) & \quad \text{The pig didn’t carry all of his rocks.} \\
(54) & \quad exh[SOME] = SOME \land \neg ALL
\end{align*}
\]

\(^8\)This is what Fox (2007) refers to as ‘innocent exclusion’. Fox (2007) further refines the version above, but it will suffice for our purposes; see Fox (2007) for discussion.

\(^9\)A key question for the Grammatical account and for approaches to scalar inferences in general, is how to determine the alternatives that $exh$ quantifies over. This is a very controversial topic in the literature, which would take us beyond the scope of this paper. We will simply assume the alternatives indicated here; see Breheny et al. (2016) and references therein for discussion.
The **all** alternative is excludable in this case, and the meaning obtained is the desired **OnlySome** inference \( \neg(\text{ALL}) \) in (53).

We turn now to how the **NotEvery** inference can be derived via the same mechanism.

### A.2 Deriving the **NotEvery** inference

As discussed, sentences like (56), schematised as ‘**EVERY**(SOME)’ have been associated with **NotEvery** inferences like (57).

(56) Every pig carried some of his rocks.

(57) Not every pig carried all of his rocks.

The grammatical approach can account for this inference in the same way as above by assuming that the sentence in (56) is parsed with a global silent exh, and that the alternatives it quantifies over are those in (59):

(58) \( \text{exh[EVERY(SOME)] = EVERY(SOME) \land \neg EVERY(ALL)} \)

(59) \( \text{Alt = \{ EVERY(SOME), EVERY(ALL) \}} \)

**EVERY**(ALL) is excludable and its negation corresponds to the desired inference in (57).

### A.3 Deriving the **None** inference

Within the grammatical account, there are two main ways of deriving the **None** inference in (60) from (56).

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

The first way is based on assuming that the set of alternatives over which exh quantifies is larger than that above and also includes the alternatives created by replacing the quantifier **every** with its scale-mate **some** (Chemla and Spector, 2011; Magri, 2011; Romoli, 2012; Gotzner and Romoli, 2018). The computation is as in (61) and the set of alternatives is that in (62).

(61) \( \text{exh[EVERY(SOME)] = EVERY(SOME) \land \neg EVERY(ALL)} \)

(62) \( \text{Alt = \{ EVERY(SOME), EVERY(ALL), SOME(SOME), SOME(ALL) \}} \)
The crucial ingredient to note is that the alternative some(all) is excludable and its negation, ¬some(all) corresponds to the inference in (60).

The second way to derive (60) 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 (63) have two main sites at which exh can appear: at the matrix level, as in (64), or embedded under the universal quantifier, as in (65).

(63) Every pig carried some of his rocks.
(64) exh[Every pig carried some of his rocks.]
(65) Every pig λx[exh[x carried some of his rocks.]]

Applying exh at the matrix level results in the computation of the NotEvery inference in (57), as we just outlined (see Section A.2). Embedding exh under the universal quantifier, on the other hand, results in the derivation of the None inference in (60). To illustrate, consider the alternative over which the embedded exh operates:

(66) exh[x SOME] = x SOME ∧ ¬x ALL
(67) Alt = \{ x SOME, x ALL \}

After abstracting over the variable, the predicate that combines with the topmost quantifier every is that in (68), resulting in the interpretation entailing the None inference in (69), i.e. every pig carried some and not all of his rocks entails that no pig carried all of his rocks.

(68) λx[x SOME ∧ ¬x ALL]
(69) EVERY(λx[x SOME ∧ ¬x ALL])

In sum, there are two ways to derive the None inference in the grammatical account, one involving an embedded exh and the other expanding the set of alternatives that are involved. Both of these options touch upon a variety of controversial issues regarding local vs. global computation of inferences and the combination of alternatives, a full discussion of which would take us beyond the scope of this paper.

References


